



Serine/Threonine protein kinases from bacteria, archaea and eukarya share a common evolutionary origin deeply rooted in the tree of life

Stancik, Ivan Andreas; Šestak, Martin Sebastijan; Ji, Boyang; Axelson-Fisk, Marina; Franjevic, Damjan; Jers, Carsten; Domazet-Lošo, Tomislav; Mijakovic, Ivan

Published in:
Journal of Molecular Biology

Link to article, DOI:
[10.1016/j.jmb.2017.11.004](https://doi.org/10.1016/j.jmb.2017.11.004)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Stancik, I. A., Šestak, M. S., Ji, B., Axelson-Fisk, M., Franjevic, D., Jers, C., ... Mijakovic, I. (2018). Serine/Threonine protein kinases from bacteria, archaea and eukarya share a common evolutionary origin deeply rooted in the tree of life. *Journal of Molecular Biology*, 430(1), 27-32. DOI: 10.1016/j.jmb.2017.11.004

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

Serine/Threonine protein kinases from bacteria, archaea and eukarya share a common evolutionary origin deeply rooted in the tree of life



Ivan Andreas Stancik, Martin Sebastijan Šestak, Boyang Ji, Marina Axelson-Fisk, Damjan Franjevic, Carsten Jers, Tomislav Domazet-Lošo, Ivan Mijakovic

PII: S0022-2836(17)30542-9
DOI: doi:[10.1016/j.jmb.2017.11.004](https://doi.org/10.1016/j.jmb.2017.11.004)
Reference: YJMBI 65549

To appear in: *Journal of Molecular Biology*

Received date: 2 July 2017
Revised date: 4 November 2017
Accepted date: 5 November 2017

Please cite this article as: Stancik, I.A., Šestak, M.S., Ji, B., Axelson-Fisk, M., Franjevic, D., Jers, C., Domazet-Lošo, T. & Mijakovic, I., Serine/Threonine protein kinases from bacteria, archaea and eukarya share a common evolutionary origin deeply rooted in the tree of life, *Journal of Molecular Biology* (2017), doi:[10.1016/j.jmb.2017.11.004](https://doi.org/10.1016/j.jmb.2017.11.004)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Serine/Threonine protein kinases from bacteria, archaea and eukarya share a common evolutionary origin deeply rooted in the tree of life

Ivan Andreas Stancik^{1,2*}, Martin Sebastijan Šestak^{3,*}, Boyang Ji¹, Marina Axelson-Fisk⁴, Damjan Franjevic⁵, Carsten Jers², Tomislav Domazet-Lošo^{3,6}, Ivan Mijakovic^{1,2,#}

¹Department of Biology and Biological Engineering, Chalmers University of Technology, Kemivägen 10, 41296 Gothenburg, Sweden

²Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, 2800 Lyngby, Denmark

³Laboratory of Evolutionary Genetics, Ruđer Bošković Institute, Bijenička cesta 54, HR-10002 Zagreb, Croatia

⁴Department of Mathematical Sciences, Chalmers University of Technology, Kemivägen 10, 41296 Gothenburg, Sweden

⁵Department of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia

⁶Catholic University of Croatia, Ilica 242, HR-10000 Zagreb, Croatia

*These authors contributed equally.

#Corresponding author:

Professor Ivan Mijakovic, Phone: +46 70 982 84 46, E-mail: ivan.mijakovic@chalmers.se

Keywords: bacterial protein kinase; Hanks-type kinase; eukaryotic-type kinase; eSTK; phylostratigraphy

Abstract

The main family of serine/threonine/tyrosine protein kinases present in eukarya was defined and described by Hanks *et al.* in 1988. It was initially believed that these kinases do not exist in bacteria, but extensive genome sequencing revealed their existence in many bacteria. For historical reasons, the term “eukaryotic-type kinases” propagated in the literature to describe bacterial members of this protein family. Here, we argue that this term should be abandoned as a misnomer, and we provide several lines of evidence to support this claim. Our comprehensive phylostratigraphic analysis suggests that Hanks-type kinases present in eukarya, bacteria and archaea all share a common evolutionary origin in the lineage leading to the last universal common ancestor (LUCA). We found no evidence to suggest substantial horizontal transfer of genes encoding Hanks-type kinases from eukarya to bacteria. Moreover, our systematic structural comparison suggests that bacterial Hanks-type kinases resemble their eukaryal counterparts very closely, while their structures appear to be dissimilar from other kinase families of bacterial origin. This indicates that a convergent evolution scenario, by which bacterial kinases could have evolved a kinase domain similar to that of eukaryal Hanks-type kinases, is not very likely. Overall, our results strongly support a monophyletic origin of all Hanks-type kinases, and we therefore propose that this term should be adopted as a universal name for this protein family.

In a landmark paper from 1988, Hanks *et al.* performed a multiple sequence alignment of known Ser/Thr/Tyr protein kinases catalytic domains (1). A common structural feature of these proteins is a 250-300 amino acid catalytic domain, containing 11 conserved sequence motifs that constitute the active site. Initially, protein kinases of this type were identified only in eukarya. When *Escherichia coli* genome was sequenced in 1997 (2), no genes encoding the canonical Hanks-type kinases were found. It was previously known that *E. coli* possesses a different type of Ser/Thr kinase, the bifunctional isocitrate dehydrogenase kinase/phosphatase (3). However, genome sequencing of other bacteria, such as *Myxococcus xanthus*, *Mycobacterium tuberculosis*, *Bacillus subtilis*, *Helicobacter pylori* and *Haemophilus influenzae*, provided evidence for existence of Hanks-type kinases, which was subsequently confirmed by biochemical characterization (4). Today, Hanks-type kinases are known to control many important processes in bacterial physiology via protein substrate phosphorylation (4). To provide just a few examples, in *M. tuberculosis*, kinases PknG and PknH contribute to virulence and survival during the infection process (5,6). Similarly, in *Yersinia pseudotuberculosis*, a secreted kinase YpkA is required for virulence (7). In *S. pneumoniae*, the kinase StkP controls cellular cytokinesis and morphogenesis (8,9). In *B. subtilis*, the kinase PrkC controls spore germination (10), while the kinase YabT controls DNA proofreading and forespore metabolic quiescence during spore formation (11-13). In *Myxococcus xanthus*, development of fruiting bodies and spores is controlled by a key transcriptional regulator MrpC, which is regulated via phosphorylation by the kinase Pkn14 (14). Interestingly, MrpC is also controlled by a two component system

MrpA/MrpB. It has recently been highlighted that bacterial Hanks-type kinases can phosphorylate transcription regulators and histidine kinases of the two component systems, and thus contribute to signal integration in these otherwise linear signal transduction pathways (15).

Regarding the evolutionary origin of bacterial Hanks-type Ser/Thr kinases, in 1998, Leonard *et al.* argued that these putative bacterial kinases constitute novel families of the “eukaryotic” protein kinase superfamily (16). The authors proposed an evolutionary scenario supporting a monophyletic origin of this entire kinase superfamily, which appeared before the separation of the three domains of life. However, despite the evidence of a monophyletic origin, before the emergence of eukarya, the term “eukaryotic-type kinases” unfortunately propagated in the literature. We have in the past argued that the term “eukaryotic-type kinase” is a misnomer when describing Hanks-type kinases in bacteria (17,18). Our argument was based on the fact that these kinases are present in bacteria and archaea, and that there is no evidence supporting the notion that they originated in eukarya. Nevertheless, the term “eukaryotic-type kinase” remains in use until this day, especially among microbiologists (19-24). In order to further substantiate our claim that Hanks-type kinases are an ancient protein family, shared by archaea, bacteria and eukarya, we here attempted to provide a new line of evidence, independent from that of Leonard *et al.* (16). For this, we decided to take a phylostratigraphic approach (25), a method for tracing the origin of evolutionary novelties based on the similarity searches of a well populated protein sequence database. The phylostratigraphic approach has been successfully applied to trace the evolutionary origin of cancer-related genes (26), for precise evolutionary stratification of developmental phenomena (27), and in studies of evolutionary emergence of *de novo* genes from non-genic regions (28). Theoretical background, advantages and limits of this method have been recently covered by Domazet-Lošo *et al.* (29).

In analogy to the study of Hanks *et al.* (1), where the starting point was the ensemble of characterized kinases from eukarya, our point of departure was the set of experimentally characterized Hanks-type kinases from bacteria. We focused on the 30 well characterized bacterial Hanks-type kinases, as listed by Pereira *et al.* (4), with the addition of two recently characterized *B. subtilis* kinases YabT (11) and PrkD (30). Complete sequences of these 32 kinases were obtained from the National Center for Biotechnology Information (NCBI) (Supplementary Table 1) and subjected to the analytical pipeline presented in the Supplementary Fig. 1. Briefly, the kinase catalytic domain sequences were aligned and used to construct a hidden Markov model consensus profile. The consensus sequence profile was used to search UniProtKB/Swiss-Prot, the manually annotated section of the UniProt databases, under a stringent cut-off (E-value of e^{-5}). The query returned 1457 Ser/Thr protein kinase sequences (Supplementary Table 2), from a total of 164 genomes, covering species from all three domains of life: eukarya, bacteria and archaea (Supplementary Table 3). Fig. 1A presents a multiple sequence alignment of a subset of catalytic domains from all three domains of life, showing that the Hanks motifs (1) are well conserved in these kinases from archaea, bacteria and eukarya. Sequences of bacterial HipA kinases, distantly related to the Hanks-type kinases (31), were included to highlight the core signature motifs of Hanks-type kinases.

Next, to determine their phylogenetic origin, the sequences of the 32 experimentally characterized bacterial Ser/Thr kinases (Supplementary Table 1) and 1457 of their closest

homologues from UniProtKB/Swiss-Prot (Supplementary Table 2) were subjected to a phylostratigraphic analysis, as described previously (26-28). We first built a custom protein database by combining complete genomes from NCBI, Ensembl and Joint Genome Institute (JGI). In total, 113 834 351 protein sequences from 25 223 genomes were collected. We searched the protein kinase sequences against this custom protein database by using BLASTP at E-value cut-off of e^{-3} (26). Using the obtained BLAST output, we determined the phylogenetic origin of the kinase genes on a premade consensus phylogenetic tree. Our phylostratigraphic analysis placed all of the 32 experimentally characterized bacterial Ser/Thr kinases unambiguously in the phylostratum 1. This is the phylogenetic level corresponding to the lineage leading to the last universal common ancestor (LUCA); i.e. prior to separation of the three domains of life (Supplementary Table 1). In the consensus phylogenetic tree constructed for analyzing bacterial kinases, this phylostratum is represented by reference genomes of 256 archaeal and 1458 eukaryal species. Essentially all 32 bacterial kinases had significant hits within archaeal (97%) and eukaryal (100%) sequences, and on average each kinase had significant hits to 32% archaeal and 92% eukaryal species in the database. A very similar outcome was obtained with the 1457 kinases extracted by the UniProt search. 1453 kinases from this dataset (99,7%) were placed in the phylostratum 1. The remaining four kinases were found in the phylostratum 2 (the lineage leading to the origin of eukarya), and they are: Bub1 from *Schizosaccharomyces pombe* and 7tmk2, Sky1 and Scy2 from *Dictyostelium discoideum*. To illustrate these results, phylostratigraphic maps are shown for three organisms harboring a large number of kinase genes: *Homo sapiens*, *Arabidopsis thaliana* and *Mycobacterium tuberculosis* (Fig. 1B). All of their kinases map consistently to the phylostratum 1. As previously proposed by Leonard *et al.* (16), our analysis clearly confirmed that this protein superfamily originated in LUCA, and is comparably spread in bacteria, eukarya and archaea.

It is known that some genes of eukaryal origin can end up in genomes of bacterial pathogens and commensals by horizontal gene transfer (32-34). It is therefore possible that some Hanks-type kinase genes have been transferred, for example, from humans to their bacterial commensals. However, our examination of available extensive metagenome datasets of human gut microbiota (35) did not reveal any particular enrichment of Hanks-type kinases in human commensals. Moreover, we performed a comprehensive analysis of 967 publicly available bacterial genomes encoding Hanks-type kinase homologues (1475 putative kinase genes), to assess whether they have any significant association to the sites of integron-mediated gene transfer (36). Only 0.1% of the identified putative kinase genes could be associated to these gene transfer sites. Finally, we asked whether phylogenetic trees based on Hanks-type kinase sequences would differ significantly from standard phylogenetic trees based on small subunit ribosomal RNAs. Extensive gene transfer would be expected to alter the topology of the kinase-based trees. We generated corresponding phylogenetic trees from our entire kinase dataset, and the topology of the kinase-based trees matched closely with the topology of rRNA-based trees. A representative tree comparison is shown in Supplementary Fig. 2. All these findings support the notion that Hanks-type kinase gene transfer most likely did not occur on a large scale.

Finally, we asked whether the present-day Hanks-type kinase superfamily could have appeared through convergent evolution. Is it possible that bacterial Hanks-type kinases arose

from some ancestral bacterial kinase, and through the process of convergent evolution obtained a catalytic site similar to their counterpart protein kinases in archaea and eukarya? In order to examine that hypothesis, we compared the available tertiary structures of bacterial Hanks-type kinases to some of their bacterial and eukaryal counterparts, aiming to reconstruct their phylogenetic relationships based on structural alignments (37). For bacterial kinases, we used 13 available structures of P-loop kinases, enzymes which phosphorylate nucleosides, nucleotides, sugars, and coenzyme precursors (38), and 2 bacterial protein-tyrosine kinases (BY-kinases) (39). For kinases of eukaryal origin, we selected 31 structures of eukaryal protein kinases from the study by Scheeff & Bourne (37). These were compared to the 26 available structures of characterized bacterial Hanks-type kinases from our dataset. We performed an all against all structure comparison using the Dali server (40,41). The results, presented as a structural similarity matrix in Fig. 2, clearly indicate that bacterial Hanks-type kinases cluster with their eukaryal counterparts, and their structures are quite distinct from bacterial P-loop kinases and BY-kinases. The bacterial P-loop and BY-kinases, for which a common evolutionary origin has been suggested (39), cluster together in our analysis of tertiary structures. The results of the same comparison using the Dali server are also provided as a dendrogram and a correspondence analysis (Supplementary Fig. 3). These results clearly support a monophyletic origin of Hanks-type kinases, and are not consistent with the hypothesis of convergent evolution, whereby bacterial Hanks-type kinases could have evolved from an ancestor of P-loop and BY-kinases.

Based on the presented evidence, we conclude that the Hanks-type kinases are an ancient and ubiquitous protein family with a monophyletic origin, in no way linked specifically to eukarya. Therefore, we propose that the term Hanks-type kinases should be adopted to replace the misleading term “eukaryotic-type kinases”, and all derivatives thereof.

Acknowledgments

This work was supported by grants from the Chalmers University of Technology, Vetenskapsrådet and the Novo Nordisk Foundation to IM.

References

1. Hanks SK, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*, 241: 42-52.
2. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, 277: 1453-1462.
3. Nimmo GA, Borthwick AC, Holms WH, Nimmo HG (1984) Partial purification and properties of isocitrate dehydrogenase kinase/phosphatase from *Escherichia coli* ML308. *Eur J Biochem* 141: 401-408.

4. Pereira SF, Goss L, Dworkin J (2011) Eukaryote-like serine/threonine kinases and phosphatases in bacteria. *Microbiol Mol Biol Rev* 75: 192-212.
5. Cowley S, Ko M, Pick N, Chow R, Downing KJ, Gordhan BG, Betts JC, Mizrahi V, Smith DA, Stokes RW, Av-Gay Y (2004) The *Mycobacterium tuberculosis* protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth *in vivo*. *Mol Microbiol* 52: 1691-1702.
6. Papavinasasundaram KG, Chan B, Chung JH, Colston MJ, Davis EO, Av-Gay Y (2005) Deletion of the *Mycobacterium tuberculosis pknH* gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. *J Bacteriol* 187: 5751-5760.
7. Galyov EE, Hakansson S, Forsberg A, Wolf-Watz H (1993) A secreted protein kinase of *Yersinia pseudotuberculosis* is an indispensable virulence determinant. *Nature* 361: 730-732.
8. Fleurie A, Manuse S, Zhao C, Campo N, Cluzel C, Lavergne JP, Freton C, Combet C, Guiral S, Soufi B, Macek B, Kuru E, VanNieuwenhze MS, Brun YV, Di Guilmi AM, Claverys JP, Galinier A, Grangeasse C (2014) Interplay of the serine/threonine-kinase StkP and the paralogs DivIVA and GpsB in pneumococcal cell elongation and division. *PLoS Genet* 10: e1004275.
9. Fleurie A, Lesterlin C, Manuse S, Zhao C, Cluzel C, Lavergne JP, Franz-Wachtel M, Macek B, Combet C, Kuru E, VanNieuwenhze MS, Brun YV, Sherratt D, Grangeasse C (2014) MapZ marks the division sites and positions FtsZ rings in *Streptococcus pneumoniae*. *Nature* 516: 259-262.
10. Shah IM, Laaberki MH, Popham DL, Dworkin J (2008) A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell* 135: 486-496.
11. Bidnenko V, Shi L, Kobir A, Ventroux M, Pigeonneau N, Henry C, Trubuil A, Noirot-Gros MF, Mijakovic I (2013) *Bacillus subtilis* serine/threonine protein kinase YabT is involved in spore development via phosphorylation of a bacterial recombinase. *Mol Microbiol* 88: 921-935.
12. Derouiche A, Petranovic D, Macek B, Mijakovic I (2016) *Bacillus subtilis* single-stranded DNA-binding protein SsbA is phosphorylated at threonine 38 by the serine/threonine kinase YabT. *Period Biol* 118: 399-404.
13. Pereira SF, Gonzalez RL, Dworkin J (2015) Protein synthesis during cellular quiescence is inhibited by phosphorylation of a translational elongation factor. *Proc Natl Acad Sci USA* 112: E3274-3281.
14. Nariya H, Inouye S (2006) A protein Ser/Thr kinase cascade negatively regulates the DNA-binding activity of MrpC, a smaller form of which may be necessary for the *Myxococcus xanthus* development. *Mol Microbiol* 60: 1205-1217.
15. Kalantari A, Derouiche A, Shi L, Mijakovic I (2015) Serine/threonine/tyrosine phosphorylation regulates DNA binding of bacterial transcriptional regulators. *Microbiology* 161: 1720-1729.
16. Leonard CJ, Aravind L, Koonin EV (1998) Novel families of putative protein kinases in Bacteria and Archaea: evolution of the "eukaryotic" protein kinase superfamily. *Genome Res*, 8: 1038-1047.

17. Mijakovic I, Macek B (2012) Impact of phosphoproteomics on studies of bacterial physiology. *FEMS Microbiol Rev* 36: 877-892.
18. Cousin C, Derouiche A, Shi L, Pagot Y, Poncet S, Mijakovic I (2013) Protein-serine/threonine/tyrosine kinases in bacterial signaling and regulation. *FEMS Microbiol Lett* 346: 11-19.
19. Wright DP, Ulijasz AT (2014) Regulation of transcription by eukaryotic-like serine-threonine kinases and phosphatases in Gram-positive bacterial pathogens. *Virulence* 5: 863-885.
20. Libby EA, Goss LA, Dworkin J (2015) The eukaryotic-Like Ser/Thr Kinase PrkC regulates the essential WalRK two-component system in *Bacillus subtilis*. *PLoS Genet* 11: e1005275.
21. Manuse S, Fleurie A, Zucchini L, Lesterlin C, Grangeasse C (2016) Role of eukaryotic-like serine/threonine kinases in bacterial cell division and morphogenesis. *FEMS Microbiol Rev* 40: 41-56.
22. Kimura Y, Urata M (2016) Characterization of a eukaryotic-like protein kinase, DspB, with an atypical catalytic loop motif from *Myxococcus xanthus*. *Arch Microbiol* 198: 219-226.
23. Flayhan A, Bergé C, Bailo N, Doublet P, Bayliss R, Terradot L (2016) The structure of *Legionella pneumophila* LegK4 type four secretion system (T4SS) effector reveals a novel dimeric eukaryotic-like kinase. *Sci Rep* 6: 20746.
24. Depardieu F, Didier JP, Bernheim A, Sherlock A, Molina H, Duclos B, Bikard D (2016) A eukaryotic-like serine/threonine kinase protects *Staphylococci* against phages. *Cell Host Microbe* 20: 471-481.
25. Domazet-Lošo T, Brajković J, Tautz D (2007) A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet* 23: 533-539.
26. Domazet-Lošo T, Tautz D (2010) Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC Biol* 8: 66.
27. Domazet-Lošo T, Tautz D (2010) A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468: 815-818.
28. Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charlotteaux B, Hidalgo CA, Barbette J, Santhanam B, et al. 2012. Proto-genes and de novo gene birth. *Nature* 487: 370–374.
29. Domazet-Lošo T, Carvunis A-R, Mar Albà M, Šestak MS, Bakarić R, Neme R, Tautz D (2017) No evidence for phylostratigraphic bias impacting inferences on patterns of gene emergence and evolution. *Mol Biol Evol* 34: 843-856.
30. Kobir A, Poncet S, Bidnenko V, Delumeau O, Jers C, Zouhir S, Grenha R, Nessler S, Noirot P, Mijakovic I (2014) Phosphorylation of *Bacillus subtilis* gene regulator AbrB modulates its DNA-binding properties. *Mol Microbiol* 92: 1129-1141.
31. Correia FF, D'Onofrio A, Rejtar T, Li L, Karger BL, Makarova K, Koonin EV, Lewis K (2006) Kinase activity of overexpressed HipA is required for growth arrest and multidrug tolerance in *Escherichia coli*. *J Bacteriol* 188: 8360-8367.
32. Gomez-Valero L, Rusniok C, Jarraud S, Vacherie B, Rouy Z, Barbe V, Medigue C, Etienne J, Buchrieser C (2011) Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. *BMC Genomics* 12: 536-560.

33. Patrick S, Jobling KL, O'Connor D, Thacker Z, Dryden DT, Blakely GW (2011) A unique homologue of the eukaryotic protein-modifier ubiquitin present in the bacterium *Bacteroides fragilis*, a predominant resident of the human gastrointestinal tract. *Microbiology* 157: 3071-3078.
34. Putonti C, Nowicki B, Shaffer M, Fofanov Y, Nowicki S (2013) Where does *Neisseria* acquire foreign DNA from: an examination of the source of genomic and pathogenic islands and the evolution of the *Neisseria* genus. *BMC Evol Biol* 13: 184-196.
35. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, Bäckhed F, Nielsen J (2012) Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun* 3: 1245.
36. Buongiorno Pereira M, Wallroth M, Kristiansson E, Axelson-Fisk M (2016) HattCI: fast and accurate *attC* site identification using hidden Markov models. *J Comp Bio* 23: 891-902.
37. Scheeff ED, Bourne PE (2005). Structural evolution of the protein kinase-like superfamily. *PLoS Comput Biol* 1: e49.
38. Leipe DD, Koonin EV, Aravind L (2003). Evolution and classification of P-loop kinases and related proteins. *J Mol Biol* 333: 781–815.
39. Shi, L., Ji, B., Kolar-Znika, L., Boskovic, A., Jadeau, F., Combet, C., Grangeasse C, Franjevic D, Talla E, Mijakovic, I. (2014). Evolution of bacterial protein-tyrosine kinases and their relaxed specificity toward substrates. *Genome Biol Evol* 6: 800–817.
40. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38: W545-549.
41. Holm L, Laakso LM (2016) Dali server update. *Nucleic Acids Res* 44: W351-355.

Figure legends

Fig. 1. Hanks-type kinase are equally conserved among the three domains of life, and most of them have originated in LUCA. A) Multiple alignment of the catalytic domains of Hanks-type kinases from archaea, bacteria and eukarya, respectively. Sequences were randomly selected from the dataset of 1457 sequences retrieved from UniProtKB/Swiss-Prot. Blue shading corresponds to highly conserved motifs, as defined by Hanks *et al.* (1), marked by roman numerals. Five sequences of HipA kinases, distantly related to Hanks-type kinases, are included at the bottom of the alignment. Kinases represented in this alignment, listed with their species of origin and UniProt codes are as follows: *Picrophilus torridus* (Q6L2T2), *Methanocaldococcus vulcanius* (C9RH09), *Methanotorris igneus* (F6BEU8), *Thermococcus thioeducens* (A0A0Q2RDP2), *Methanotorris formicicus* (H1L0Y3), *Thermococcus barophilus* (F0LHW7), *Thermococcus gammatolerans* (C5A5G9), *Thermococcus peptonophilus* (A0A142CXX3), *M. vulcanius* (C9RGZ6), *Thermococcus nautili* (W8NTA3), *Bacillus subtilis* (O34507), *Streptococcus pneumoniae* (Q8KY50), *Mycobacterium bovis* (P0A5S5), *Mycobacterium tuberculosis* (P9WI81), *Mycobacterium leprae* (P54744), *Bacillus anthracis* (Q81WH6), *Lactococcus lactis* subsp. *lactis* (Q9CEF5), *Caldanaerobacter subterraneus* subsp. *tengcongensis* (Q8R9T6), *Mycobacterium smegmatis* (A0QNG1), *Staphylococcus aureus* (A6QGC0), *Drosophila melanogaster* (Q9V315), *Rattus norvegicus* (Q63638), *Mus musculus* (Q62407),

Schizosaccharomyces pombe (Q09170), *Arabidopsis thaliana* (P42818), *A. thaliana* (Q39030), *Mus musculus* (Q9QY01), *Homo sapiens* (Q8IYT8), *A. thaliana* (Q9FJ55), *Dictyostelium discoideum* (Q8T2I8), *Escherichia coli* (P23874), *E. coli* (P39410), *Shewanella oneidensis* (Q8EIX3), *Sinorhizobium fredii* (P55412), *Haemophilus influenzae* (P44033). B) Phylostratigraphy analysis of Hanks-type kinases from *A. thaliana*, *H. sapiens* and *M. tuberculosis*. Origin of individual sequences is mapped on the reference evolutionary tree for each species.

Fig. 2. Structural similarity matrix (Dali Z-scores) obtained from the Dali server (Holm & Rosenström 2010). The heat map represents the values of the Dali Z-score, a pairwise similarity metric. The bottom left square is constituted of 13 P-loop kinases (Leipe et al. 2003) and 2 BY-kinases (Shi et al., 2014). The larger upper right square is constituted of Hanks-type kinases of bacterial (27) and eukaryal (32) origin. The UniProt identifiers and PDB codes of the kinases used for this analysis, along with species names, are provided in Supplementary Table 4.

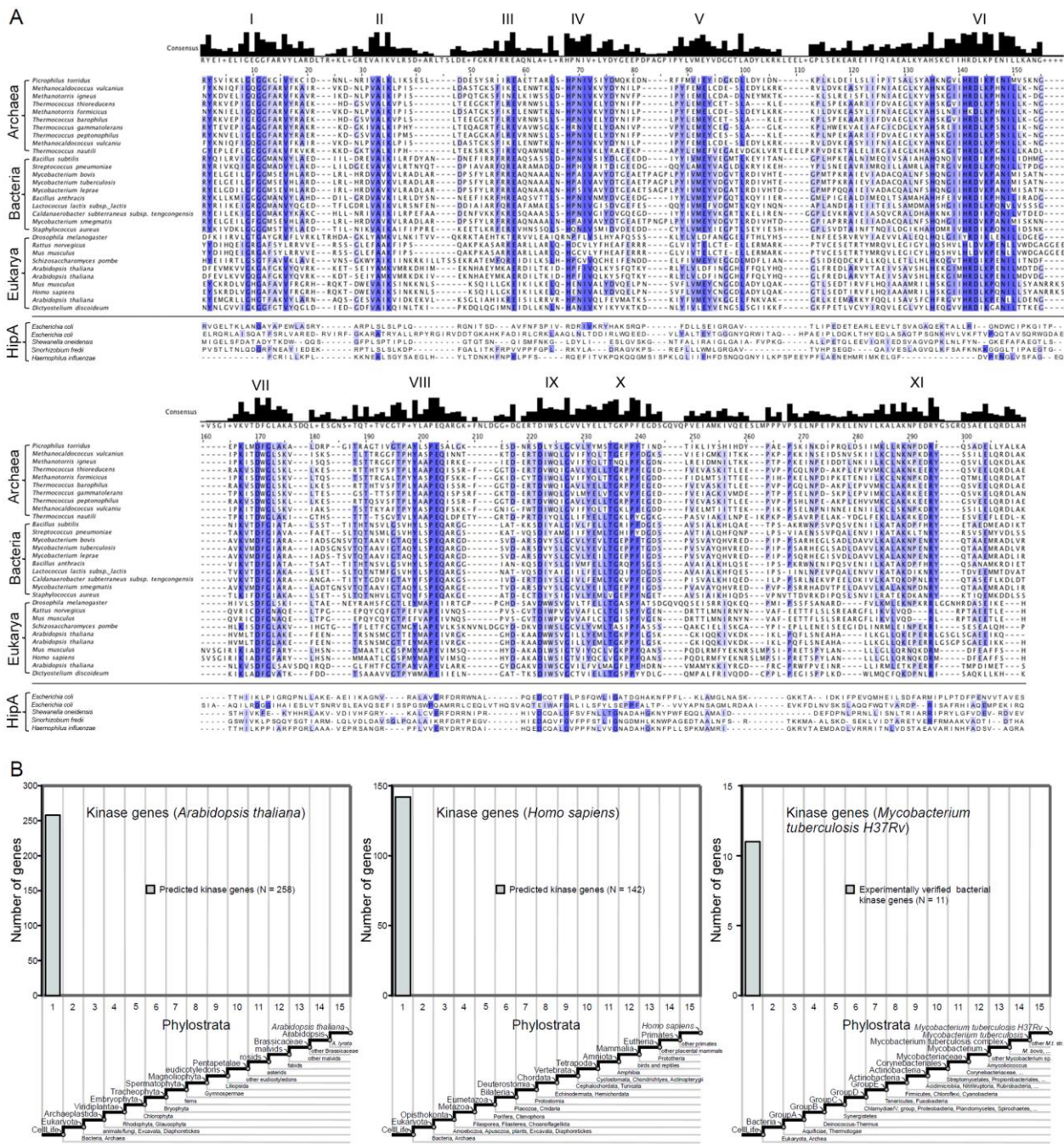
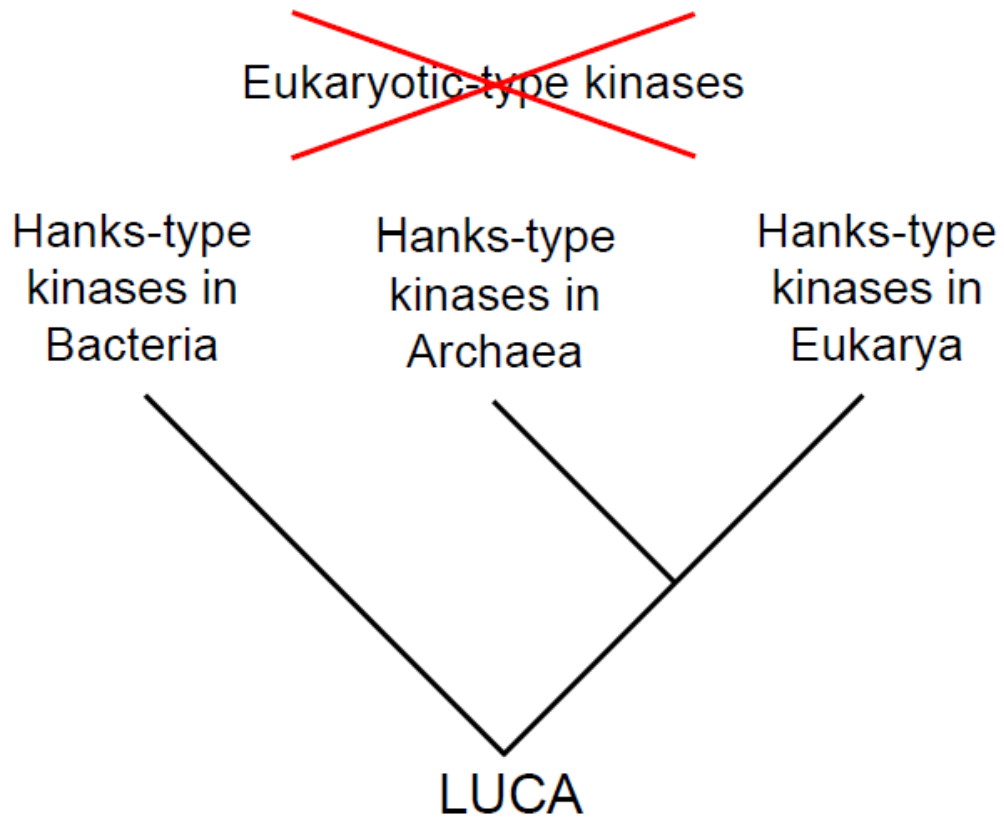


Figure 1



Graphical abstract

ACCEPTED

Highlights

- We argue that the term “eukaryotic-type protein kinases” is a misnomer
- These kinases are present and abundant in bacteria, archaea, and eukarya
- Our phylostratigraphy analysis places the origin of this protein family in LUCA
- Convergent evolution and massive horizontal transfer scenarios are ruled out
- We propose to replace the term “eukaryotic-type” by “Hanks-type” protein kinases

ACCEPTED MANUSCRIPT