Evolutionary ancestry of eukaryotic protein kinases and choline kinases

Shenshen Lai1, Javad Safaei2, Steven Pelech1,3

1Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, British Columbia, V6T 2B5, Canada.
2Department of Computer Science, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada.
3Kinexus Bioinformatics Corporation, Vancouver, British Columbia, V6P 6T3, Canada.

Running title: Ancestry of eukaryotic protein kinases

ABSTRACT

The reversible phosphorylation of proteins catalyzed by protein kinases in eukaryotes supports an important role for eukaryotic protein kinases (ePKs) in the emergence of nucleated cells in the third superkingdom of life. Choline kinases (ChKs) could also be critical in the early evolution of eukaryotes, because of their function in the biosynthesis of phosphatidylcholine, which is unique to eukaryotic membranes. However, the genomic origins of ePKs and ChKs are unclear. The high degeneracy of protein sequences and broad expansion of ePK families have made this fundamental question difficult to answer. In this study, we identified two class-I aminoacyl-tRNA synthetases with high similarities to consensus amino acid sequences of human protein-serine/threonine kinases. Comparisons of primary and tertiary structures supported that ePKs and ChKs evolved from a common ancestor related to glutaminyl aminoacyl-tRNA synthetases, which may have been one of the key factors in the successful emergence of ancient eukaryotic cells from bacterial colonies.

Protein kinases play a pivotal role in communicating intracellular signals in eukaryotes. The family of eukaryotic protein kinases (ePKs) comprises of at least 568 human members, which accounts for more than 2% of the protein coding genes of the entire human genome (1). These kinases are highly conserved in both their primary amino acid sequences (2) and in the 3D structures (3) of their catalytic domains. Because of the central regulatory roles and the high conservation of the ePKs, the ancestry of these enzymes has become an important question in the study of the evolution of eukaryotic organisms.

The majority of the kinases among the ePKs are responsible for the phosphorylation of proteins on serine or threonine residues, while a smaller group of protein kinases catalyze their tyrosine phosphorylation. This branch of protein-tyrosine kinases (PTKs) arose from protein-serine/threonine kinases (STKs), which
Ancestry of eukaryotic protein kinases

is believed to be an important event in early metazoan evolution (4,5). Of all the STKs, there is another lumped group of diverse kinases that are described as atypical protein kinases. With little sequence identity and structural similarity to typical protein kinases, these atypical protein kinases are suggested to diverge early in evolution and have distinct evolutionary histories (6,7). Despite the atypical protein kinases and recently derived PTKs, the rest of the typical protein kinases constitute a major lineage in protein kinase evolution.

Eukaryotic life is believed to have evolved between 1.7 to 2.7 billion years ago and no living representatives of the earliest eukaryotes survive today. Consequently, the actual origin of protein kinases is difficult to establish with a high degree of confidence. Firstly, protein sequences are highly degenerate, which makes the detection of sequence similarities difficult even at the superfAMILY level (8). Secondly, the ePKs comprise a group of very broadly expanded proteins. Loss and expansion of kinase-relatedness tree branches occurs in various species, as well as insertions and deletions inside their catalytic domains. To investigate these problems, we developed novel strategies using consensus sequences from precise amino acid sequence alignments as the initial query in BLAST searches and compared top hits from multiple species. Our conclusions are supported by protein primary and tertiary structure comparisons. Our findings offer new insights into the evolution of ePKs and choline kinases (ChKs) in ancient eukaryotes. The molecular paleontology approach undertaken in this study also provides a broadly applicable strategy to generally investigate the origins of large protein domain families.

EXPERIMENTAL PROCEDURES

Protein sequences and alignment. Sequences of human protein kinases, glutamine-tRNA synthetases and choline kinases are from UniProt database (http://www.uniprot.org). Initial alignment of each group was created using ClustalW (9), followed by manual adjustment of gaps and inserts. Frequency of each amino acid at each position of the alignment was calculated to generate a positional frequency matrix.

BLAST search for the STK ancestor. BLAST search using STK consensus sequence was performed by NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) or local BLAST+ (10) with a filter to exclude all the protein kinase hits. List of top hits from each organism was generated separately and then compared. Candidates that were common in at least two species with a BLAST score higher than 18.0 were selected for manual alignment to identify the ancestor gene.

Structure prediction and comparison. We utilized the Phyre server (11) for secondary structure prediction and alignment. RCSB PDB Protein Comparison Tool (12,13) and other related applications from RCSB Protein Data Bank (http://www.pdb.org) were applied to view and align the 3D structures.

Conservation of human STKs. To search for the most ancient STK, the conservation of all the human protein-serine/threonine kinases were calculated by retrieving from PhosphoNET (http://www.phosphonenet.ca) the calculated identities and similarities between each human protein kinase and its closest homologs in twenty other species. This list was compared with the ranking of human STKs by their similarities to GluRS consensus sequence and STK consensus sequence.

RESULTS

Alignment of human protein kinase domains - The amino acid sequences of all the typical human protein kinase catalytic domains were collected and precisely aligned based on
primary and secondary structural data (Supplemental Table S1). The alignment contained 12 catalytic subdomains made up of about 30 highly conserved amino acids, and 10 gaps, which represented more variable regions responsible for the specificity of individual kinase. The initial alignment was facilitated by the early work of Hanks et al. (2), which was further refined with more secondary structure information that has arisen from x-ray crystallographic structures of more than 50 protein kinases. The STKs and the PTKs were separated to two groups. Despite of the preponderance of conserved residues in both STKs and PTKs, major differences between the two groups often occurred near Subdomains VI (HRD) and VIII (APE) (Supplemental Tables S3 and S4).

To explore the origin of ePKs, the alignment of the 393 human STK catalytic domains were used to generate a consensus sequence. We calculated the frequency of each of the 20 common amino acids at each position. The average frequency of the most common amino acid at each position was 36%, and two-thirds of them were higher than 20%, indicating very high conservation among the catalytic domains of these protein kinases. An STK consensus sequence of 247 amino acids in length was defined using the amino acid with highest frequency at each position.

A protein kinase domain alignment with 56,691 sequences from Pfam database (http://pfam.sanger.ac.uk/) was also downloaded (14,15). This alignment included the catalytic domains of both protein-serine/threonine and protein-tyrosine kinases, which were not easily resolvable in view of the diversity of species represented and the similarity shared between these subgroups of ePKs. The consensus sequence of protein kinase domains from all species was shown to be highly similar to our human STK consensus sequence with 85% overall homology (Supplemental Table S5). Since the development of the protein-tyrosine kinase group from protein-serine/threonine kinases is proposed to be a relatively late event during evolution with the emergence of metazoans (4,5), we believe our human STK consensus sequence was a closer representation of the earliest protein kinases.

Proteins most closely related to protein kinases - To identify the proteins that were most closely related to protein kinases, our STK consensus sequence was employed as the query in BLAST searches performed in six diverse species including Escherichia coli, Saccharomyces cerevisiae, Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster and Homo sapiens. For each organism, the top 100 non-protein kinase subjects with a BLAST score higher than 18.0 and an alignment length longer than 35 amino acids were considered as top hits and compared among species. In search of the most likely ancestor of protein kinase, we calculated the average BLAST scores of proteins present among the top hits in more than one species to define candidates with consistent similarity to our consensus sequence. As listed in Table 1, 11 proteins were identified with the range of average score from 19.9 to 22.6.

Given that the alignment of these distant genes might be missed by automatic BLAST search due to insertions and deletions, we then manually checked through all these listed candidates on their alignment with STK consensus sequence. Among the 11 proteins, only choline/ethanolamine kinase (ChK) and glutaminyl-tRNA (GlnRS) synthetase exhibited particular similarities at the kinase catalytic subdomains, which are highly conserved and critical for phosphotransferase activity. ChKs have already been reported to share high similarity in 3D structure with protein kinases (16). As a group of proteins, tRNA synthetases appeared among the top hits in all the six species employed for BLAST search.
Glutaminyl-tRNA synthetase, as well as glutamyl- and alanyl-tRNA synthetase, belong to class-I aminoacyl-tRNA synthetase family, which was the group most similar with the STK consensus sequences. Like ePKs and ChKs, aminoacyl-tRNA synthetases utilize ATP to accomplish their functions. These results pointed to a possible evolutionary relationship between tRNA synthetases, choline/ethanolamine kinases and protein kinases.

**Ancestry of ePK, ChK and GlnRS** - Among the three tRNA synthetases from class-I aminoacyl-tRNA synthetase family, glutaminyl-tRNA synthetase is believed to be evolved from glutamyl-tRNA synthetase (GluRS) (17). Most bacteria employ an alternative two-step pathway to synthesis glutaminyl-tRNA without GlnRS. Phylogenetic analyses indicate that GlnRS arose from the duplication of ancient GluRS after the split of bacteria and archaea/eukarya branches and acquired an N-terminal non-specific RNA binding domain later during evolution (18). GlnRS occurrence in a few bacterial species is the result of horizontal transfer from eukaryotes before the domain acquisition event (19).

Glutaminyl-tRNA synthetase (GlnRS) is the only candidate we found with particular conservation at the protein kinase catalytic subdomains from the BLAST results. A consensus sequence for this protein was created using sequences from various species and compared with the STK consensus sequence. From the alignment, the catalytic domain of GlnRS showed strong similarities with the kinase subdomains near the activation loop, including the LxxLH and DFG motifs. The GlnRS N-terminal domain aligned with the ATP-binding subunits of ePKs. We also applied the same strategy to generate glutamyl- and alanyl-tRNA synthetase (GluRS and AlaRS) consensus sequences. Both of the two genes lack the N-terminal fragment that aligns with kinase Subdomain I to V. At the same time, their catalytic domains also share much lower similarities with kinase Subdomains VI to IX when compared to GlnRS. All these results reveal a closer relationship of ePK to GlnRS than to the other aminoacyl-tRNA synthetases.

To investigate the evolutionary relationship between GlnRS, ePK and ChK, we took the consensus sequence of ChK into the alignment (Table 2). GlnRS and STK consensus sequences shared the highest identity of 24%, among which 18 of the 30 conserved amino acids were identical. The GlnRS consensus sequence displayed 20% identity with ChK. The similarities of the two pairs were both around 34%. The identity of STK and ChK was 16%, which was comparable to the identity of two randomly chosen human protein kinases with each other. These percentages strongly support the possibility that ePK, ChK and GlnRS were evolved from a common ancestor, which probably functioned as an aminoacyl-tRNA synthetase.

To further characterize the evolutionary links between GlnRS, ePK and ChK, we employed 3D structure comparison tools from RCSB Protein Data Bank (http://www.pdb.org) to align the structures of these three groups of proteins (12). Available human structures of GlnRS and ChK were used for comparison. For STKs, the candidates were from those sharing most similar amino acid sequences with the consensus sequence. The STK structures of highest similarity scores with ChK and GlnRS are shown in Figure 1 and Figure 2. The human choline kinase displayed high structural similarity with human PKA (p-value = 0.024, calculated by the algorithm). The human GlnRS had slightly lower scores with both ePK and ChK (p-value = 0.1~0.2, calculated by the algorithm).

We also used the Phyre server (11) to predict the secondary structure of the part of GlnRS that aligns with ePK catalytic domain.
Although some of the important beta strands were missing in the predicted GlnRS secondary structure, the overall pattern was similar, especially near Subdomains VI to VIII, which was also the most conserved region in GlnRS. For the regions with more dissimilar secondary structures, most of the key amino acid residues critical for maintaining the kinase catalytic core were conserved in GlnRS sequences, including E91 and H158 in human cAMP-dependent protein kinase, which is consistent with the possibility of generating the ePK catalytic structure through a series of point mutation events starting with a duplicated GlnRS gene.

In summary, our data indicated that ePKs and ChKs both emerged from an ancient aminoacyl-tRNA synthetase, which was also the ancestor of contemporary GlnRS.

The most ancient STKs and PTKs - We hypothesized that the most ancient protein kinases should be conserved across species and be more closely related to other PKs in primary structure. We determined the percent amino acid identity and similarity scores for the full length forms of 388 human protein-serine/threonine kinases in 22 other diverse species, and observed that casein kinases 1 and 2, various cyclin-dependent protein kinases, glycogen synthase kinase 3 and the p38 and ERK MAP kinases were the most evolutionary conserved of the human protein kinases (Supplemental Table S8). However, scores from BLAST search using STK consensus sequence were very close mainly due to the high similarity among all the human STKs, which rendered the result lacking resolution. Therefore, catalytic domains of all the human protein kinases were also aligned with GlnRS consensus sequence by BLAST. The top 20 hits aligned with GlnRS consensus sequence have the scores range from 19.4 to 23.3. Seven of them including AMPKs and some RSK family members, also appeared among the top 100 hits in the BLAST search against STK consensus sequence (Table 3). Among the seven candidates, AMPKs, the metabolic stress-sensing protein kinases switching off biosynthetic pathways when AMP level rises due to fuel limitation or hypoxia (20), had the highest conservation scores. Additionally, the AMPKs were also consistently found in kinomes from yeast to human (21), indicating that these kinases are most closely related to the ancient protein kinases.

Similarly, we generated a human PTK catalytic domain consensus sequence from the alignment (Supplemental Table S3). From BLAST searches with this consensus sequence, EPH and Src families were identified as the most ancient PTKs. In fact, these two families appeared to be the closest to the merging point of receptors and non-receptors in the evolutionary tree of human kinome. Moreover, they were also the most broadly spanned PTK families, with 221 EPH receptors and 172 Src family members identified from 37 metazoans. Thus, we concluded that EPH- and Src-like kinases were the most ancient receptor and non-receptor PTKs, respectively.

DISCUSSION
A few eukaryotic protein kinase-like genes have been identified in archaea (22) and prokaryotes (23, 24). The widespread distribution of protein kinase genes has led to suggestions that the ancestry of these catalytic domains predated the divergence of the three domains of life (6). However, these eukaryotic-like protein kinases lack some of the essential motifs of ePKs. Other studies have indicated some of the eukaryotic-like protein kinases had distinct evolutionary histories, which might be even more ancient than ePKs (9, 25).

Signal transduction in prokaryotes is mainly conducted through the two-component system by histidine kinases instead of by protein-serine/threonine or protein-tyrosine
kinases. These histidine kinases commonly possess a conserved C-terminal kinase core domain that features the phospho-accepting histidine as well as homology boxes (H-, N-, D-, F-, G- and X) that are not evident in typical eukaryotic protein kinases and display no resemblance to the highly conserved kinase catalytic subdomains in ePKs (26).

With recent data generated from the sequencing of many whole genomes, it is believed genes actively undergo horizontal transfers across species, which contribute significantly to the flows of genes in evolution (27,28). Horizontal gene transfers most likely account for many of the eukaryotic-like protein kinases that have been identified in bacteria. These proteins, such as the PknB kinases (29,30) and the aminoglycoside phosphotransferase APH(3’)-IIIa (31,32), are usually limited to a few branches of the entire bacterial kingdom. Thus, ePKs are still likely to have an eukaryotic origin.

The human protein kinase complement is a well studied group of regulatory enzymes that is expanded broadly in relatedness trees in all investigated eukaryotes. As a result, we selected all of the human STK catalytic domains and precisely aligned them to generate a representative consensus sequence for ancient ePKs. The strategy of comparing BLAST results from various well studied organisms and aligning the extremely conserved key residues made it possible to detect long distant relationships. The supportive results from primary sequence analysis and structural comparison provide high confidence in the evolutionary linkages between glutaminyl-tRNA synthetase, protein kinases and choline/ethanolamine kinases.

This contention could be further supported in future studies by site-directed mutagenesis experiments, ideally starting with the deduced consensus sequence of GlnRS or possibly human glutaminyl aminoacyl-tRNA synthetase as this would be technically easier. Based on our comparisons of the consensus sequences of the ePKs and GlnRS shown in Table 2, there are at least 8 highly conserved amino acids found in the catalytic subdomains of ePKs that were missing in GlnRS. Replacement of these amino acid residues in GlnRS with those that are conserved in the ePKs in their catalytic subdomains and which are generally involved in ATP binding and catalysis would be a first step. Additional amino acids residue replacements may be needed for improving recognition of the protein substrate. Our Protein Kinase Substrate Prediction Algorithm Version 2.0 predicts substrate determining residues (SDR’s) that might also be altered to improve the prospects of successful conversion of a GlnRS into a protein kinase (33).

Our results indicated that ePKs and ChKs share a common ancestor, which is consistent with previous 3D structure studies on these proteins. GlnRS exhibited higher sequence identities with ePKs and ChKs than these did with each other, as well as moderate structural similarities. It appears to be the contemporary gene most closely related to the ancestor of both ePK and ChK. Although GlnRS appears exclusively in eukarya and archaea, the aminoacyl-tRNA synthetases comprise a most ancient group of genes that are believed to undergo horizontal transfers early in evolution and gave rise to many of the contemporary genes (33,35). We are compelled to believe ePKs and ChKs also have an early eukaryotic origin and both play an important part in early evolution of highly complex eukaryotic cells.

Here we propose ePKs and ChKs arose from a common ancestor that is an ancient gene involved in the mRNA translation process as an aminoacyl-tRNA ligase. The emergence of ChKs offered additional phospholipid constituents for construction of more complex membrane structures that provide for intracellular compartmentalization as well as...
sources of intracellular mediators of signaling. At the same time, ePKs made the communication among different compartments more specific and efficient, which facilitated the specialization of various organelles. The emergence of protein kinases and choline/ethanolamine kinases may well have been critical to the development and success of eukaryotic organisms.

Acknowledgements: SL and JS were supported by University Graduate Fellowships from the University of British Columbia and Kinexus Bioinformatics Corporation. This work was funded by Kinexus.

Conflict of interest: SP is the president, chief scientific officer and majority shareholder of Kinexus Bioinformatics Corporation, which offers proteomics services and products for sale.

Author contributions: SP conceived the project. SL and SP designed and performed most of the analyses. JS carried out the pfam domain alignments and kinase domain evolutionary conservation analyses. SL wrote the initial draft of the manuscript. SL and SP prepared the figures and tables. All authors analyzed the results and approved the final version of the manuscript.

REFERENCES

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. Science 298, 1912-1934
2. Hanks, S. K., Quinn, A. M., and Hunter, T. (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241, 42-52
3. Taylor, S. S., and Radzio-Andzelm, E. (1994) Three protein kinase structures define a common motif. Structure 2, 345-355
4. Darnell, J. E., Jr. (1997) Phosphotyrosine signaling and the single cell:metazoan boundary. Proc. Natl. Acad. Sci. U S A 94, 11767-11769
5. Rokas, A., Kruger, D., and Carroll, S. B. (2005) Animal evolution and the molecular signature of radiations compressed in time. Science 310, 1933-1938
6. Leonard, C. J., Aravind, L., and Koonin, E. V. (1998) Novel families of putative protein kinases in bacteria and archaeb: evolution of the "eukaryotic" protein kinase superfamily. Genome Res. 8, 1038-1047
7. Middelbeek, J., Clark, K., Venselaar, H., Huynen, M. A., and van Leeuwen, F. N. (2010) The alpha-kinase family: an exceptional branch on the protein kinase tree. Cellular Molecular Life Sciences : CMLS 67, 875-890
8. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536-540
9. Ortiz-Lombardia, M., Pompeo, F., Boitel, B., and Alzari, P.M. (2003) Crystal structure of the catalytic domain of the PknB serine/threonine kinase from Mycobacterium tuberculosis. J. Biol. Chem. 278, 13094-13100
10. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.

11. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363-371.

12. Ye, Y., and Godzik, A. (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* **19 Suppl 2**, ii246-255.

13. Holland, R.C., Down, T.A., Pocock, M., Prlic, A., Huen, D., James, K., Foisy, S., Drager, A., Yates, A., Heuer, M., and Schreiber, M.J. (2008) BioJava: an open-source framework for bioinformatics. *Bioinformatics* **24**, 2096-2097.

14. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L. L., Eddy, S. R., and Bateman, A. (2010) The Pfam protein families database. *Nucleic Acids Res.* **38**, D211-222.

15. Sonnhammer, E. L., Eddy, S. R., and Durbin, R. (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28**, 405-420.

16. Peisach, D., Gee, P., Kent, C., and Xu, Z. (2003) The crystal structure of choline kinase reveals an eukaryotic protein kinase fold. *Structure* **11**, 703-713.

17. Lamour, V., Quevillon, S., Diriong, S., N'Guyen, V. C., Lipinski, M., and Mirande, M. (1994) Evolution of the Glx-tRNA synthetase family: the glutaminyl enzyme as a case of horizontal gene transfer. *Proc. Natl. Acad. Sci. U S A* **91**, 8670-8674.

18. Saha, R., Dasgupta, S., Basu, G., and Roy, S. (2009) A chimaeric glutamyl:glutaminyl-tRNA synthetase: implications for evolution. *Biochem. J.* **417**, 449-455.

19. Siatecka, M., Rozek, M., Barciszewski, J., and Mirande, M. (1998) Modular evolution of the Glx-tRNA synthetase family---rooting of the evolutionary tree between the bacteria and archaea/eukarya branches. *Eur. J. Biochem.* **256**, 80-87.

20. Carling, D., Aguan, K., Woods, A., Verhoeven, A.J., Beri, R.K., Brennan, C.H., Sidebottom, C., Davidson, M.D. and Scott, J. (1994) Mammalian AMP-activated protein kinase is homologous to yeast and plant protein kinases involved in the regulation of carbon metabolism. *J. Biol. Chem.* **269**, 11442-11448.

21. Manning, G., Plowman, G. D., Hunter, T., and Sudarsanam, S. (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514-520.

22. Smith, R. F., and King, K. Y. (1995) Identification of a eukaryotic-like protein kinase gene in Archaeabacteria. *Protein Sci.* **4**, 126-129.

23. Tyagi, N., Anamika, K., and Srinivasan, N. (2010) A framework for classification of prokaryotic protein kinases. *PLoS One* **5**, e10608.

24. Perez, J., Castaneda-Garcia, A., Jenke-Kodama, H., Muller, R., and Munoz-Dorado, J. (2008) Eukaryotic-like protein kinases in the prokaryotes and the myxobacterial kinome. *Proc. Natl. Acad. Sci. U S A* **105**, 15950-1595.

25. Scheeff, E. D., and Bourne, P. E. (2005) Structural evolution of the protein kinase-like superfamily. *PLoS Computational Biology* **1**, e49.

26. Grebe, T.W., and Stock, J.B. (1999) The histidine protein kinase superfamily. *Advances in Microbial Physiology* **41**, 139-227.

27. Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C., and Manning, G. (2007) Structural and functional diversity of the microbial kinome. *PLoS Biol.* **5**, e17.
Ancestry of eukaryotic protein kinases

28. Syvanen, M. (2012) Evolutionary implications of horizontal gene transfer. *Amm. Rev. Genet.* **46**, 341-358
29. Keeling, P. J., and Palmer, J. D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**, 605-618
30. Kennelly, P. J. (2002) Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEBS Microbiol. Lett.* **206**, 1-8
31. Daigle, D. M., McKay, G. A., Thompson, P. R., and Wright, G. D. (1999) Aminoglycoside antibiotic phosphotransferases are also serine protein kinases. *Chem. Biol.* **6**, 11-18
32. Walsh, C. (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775-781
33. Safaei, J., Manuch, J., Gupta, A., Stacho, L. and Pelech, S. (2011) Prediction of 492 human protein kinase substrate specificities. *Proteome Sci.* **9** Suppl 1:S6.
34. Theobald, D. L. (2010) A formal test of the theory of universal common ancestry. *Nature* **465**, 219-222
35. Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E., and Stanhope, M. J. (2001) Universal trees based on large combined protein sequence data sets. *Nat. Genet.* **28**, 281-285

FOOTNOTES

1. The abbreviations used are: AMPK, AMP-dependent protein kinase; ChKs, choline kinases; ePKs, eukaryotic protein kinases; GlnRS, glutaminyl-tRNA synthetase; PKA, cAMP-dependent protein kinase; PTKs, protein-tyrosine kinases; STKs; protein-serine/threonine kinases.

FIGURE LEGENDS

**FIGURE 1.** Structural comparisons of protein-serine/threonine kinases, choline kinases and glutaminyl t-RNA synthetases. The identity and similarity of each pair (grey text) were calculated based on primary amino acid consensus sequences. The PDB ID of each 3D structure used for comparison were shown in dark circles and linked with scores generated from RCSB PDB Protein Comparison Tool.

**FIGURE 2.** Comparison of 3D structures of protein-serine/threonine kinases, choline kinase and glutaminyl t-RNA synthetase. (A) glutaminyl-tRNA synthetase (PDB ID: 4R3Z Chain C, orange) with choline kinase (PDB ID: 2IG7, blue); (B) glutaminyl-tRNA synthetase (PDB ID: 4R3Z Chain C, orange) with microtubule affinity-regulating kinase (PDB ID: 2R0I, blue); (C) cAMP-dependent protein kinase PKA (PDB ID: 3DND, orange) with choline/ ethanolamine kinase (PDB ID: 2IG7, blue). All sequences were from human. The comparisons were done by RCSB PDB Protein Comparison Tool using jFatCat – Java Web Start (rigid) method. Two different angles of each alignment are shown.
A consensus sequence of the human protein-serine/threonine kinase catalytic domain was used to search for most closely related non-kinase proteins in six diverse species. The STK consensus sequence used was shown in Table 2 and Supplemental Table S2. Top subjects from BLAST results of each organism (score $\geq 18.0$) were compared, and those present among the top hits in more than one species were listed above. Entries were not provided when no significant similarities were detected and no BLAST score could be calculated by the BLAST algorithm.
TABLE 2. Alignment of consensus sequences of protein-serine/threonine kinases, choline kinases and glutaminyl-tRNA synthetases

The primary structures of glutaminyl-tRNA synthetases (GlnRS), protein-serine/threonine kinases (STK) and choline/ethanolamine kinases (ChK) from diverse species were aligned to generate consensus sequences. Twelve gaps (G1 – G12) were created at the same positions where the sequences were more variable in protein kinase alignments and have also been indicated with dashes. “X”s stand for positions that were not well conserved in the original consensus sequences. Highly conserved protein kinase Subdomains (I - XII) are marked and indicated in yellow. Other highly conserved amino acids between the GlnRS, STK and ChK are highlighted in dark orange (identical) and light orange (similar). Alignments for the derivation of the consensus sequences for GlnRS, STK and ChK are provided in Supplemental Tables S7, S2 and S6, respectively.
### TABLE 3. List of most probable ancient protein kinases

| Name      | Uniprot ID | Identity (%) to STK consensus sequence | Conservation score |
|-----------|------------|----------------------------------------|--------------------|
| AMPK2     | P54646     | 42.25                                  | 715                |
| p70S6K1   | P23443     | 45.90                                  | 570                |
| AMPK1     | Q13131     | 42.25                                  | 527                |
| KAMK2D    | Q13557     | 42.21                                  | 523                |
| MAST2     | Q6P0Q8     | 37.68                                  | 385                |
| RSK4      | Q9UK32     | 45.08                                  | 347                |
| p70S6K2   | Q9UBS0     | 44.40                                  | 264                |

Consensus sequences of human protein-serine/threonine kinases and glutamyl-tRNA synthetases were used as queries in BLAST search separately to identify the most closely related human protein kinases. Human protein kinases presented in both of the BLAST results were listed and sorted by their conservation scores, which were generated by summing up the identity of each human protein kinase with its homologues in ten diverse species (data from HomoloGene (www.ncbi.nlm.nih.gov/homologene)).
FIGURE 1.
FIGURE 2.
