Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages

David Pincus*, Ivica Letunic†, Peer Bork†, and Wendell A. Lim*‡

*Department of Cellular and Molecular Pharmacology, University of California, 600 16th Street, San Francisco, CA 94158; and †European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Communicated by Henry R. Bourne, University of California, San Francisco, CA, April 3, 2008 (received for review February 12, 2008)

Multicellular animals use a three-part molecular toolkit to mediate phospho-tyrosine signaling: Tyrosine kinases (TyrK), protein tyrosine phosphatases (PTP), and Src Homology 2 (SH2) domains function, respectively, as ''writers,'' ''erasers,'' and ''readers'' of phospho-tyrosine modifications. How did this system of three components evolve, given their interdependent function? Here, we examine the usage of these components in 41 eukaryotic genomes, including the newly sequenced genome of the choanoflagellate, *Monosiga brevicollis*, the closest known unicellular relative to metazoans. This analysis indicates that SH2 and PTP domains likely evolved earliest—a handful of these domains are found in premetazoan eukaryotes lacking tyrosine kinases, most likely to deal with limited tyrosine phosphorylation cross-catalyzed by promiscuous Ser/Thr kinases. Modern TyrK proteins, however, are only observed in two lineages, metazoans and choanoflagellates. These two lineages show a dramatic coexpansion of all three domain families. Concurrent expansion of the three domain families is consistent with a stepwise evolutionary model in which preexisting SH2 and PTP domains were of limited utility until the appearance of the TyrK domain in the last common ancestor of metazoans and choanoflagellates. The emergence of the full three-component signaling system, with its dramatically increased encoding potential, may have contributed to the advent of metazoan multicellularity.

choanoflagellates | encoding potential | tyrosine kinase | src homology 2 | protein tyrosine phosphatase

Tyrosine phosphorylation is essential for cell–cell communication in animals, mediating hormone, growth factor, immune, and adhesion-based signaling (1–4). Thus, phospho-tyrosine (P-Tyr) signaling has traditionally been linked with metazoan multicellularity (5). Metazoan phospho-tyrosine signaling pathways are built from a three-part system of molecular components: Tyrosine kinases (TyrK) are catalytic domains that add phospho-tyrosine (P-Tyr) modifications, protein tyrosine phosphatases (PTP) are catalytic domains that remove these modifications, and Src Homology 2 (SH2) domains are recognition domains that readout these modifications. These modules play the role of "writer," "eraser," and "reader," respectively, a triad of core functions at the heart of many biological and nonbiological information processing systems (6, 7). By using these three modules in combination, remarkably diverse signaling responses can be generated.

A fundamentally important question is how this and other biological reader/writer/eraser information processing systems could have initially evolved, given the highly interdependent function of the individual parts in modern organisms. Because they act as a synergistic system, an incomplete set of components would be deficient in function. What selective advantage could have sustained a stepwise evolutionary path?

The P-Tyr signaling machinery presents a particularly interesting case, because it appears to have evolved relatively recently compared with other signaling systems. Tyrosine kinases are widespread in metazoans but are absent in most unicellular organisms, placing the evolution of this system close to the advent of multicellularity. Recently, however, it has been discovered that choanoflagellates, the closest known unicellular relative to metazoans, have tyrosine kinases (8). The newly sequenced genome of the choanoflagellate, *Monosiga brevicollis*, presents the opportunity to compare the usage of the phospho-tyrosine regulatory machinery in different metazoan and premetazoan lineages. These comparisons allow us to gain insight into how the system may have evolved and its role in the emergence of metazoan multicellularity.

Results

Frequency of P-Tyr Signaling Domains Across Genomes. We used the SMART domain resource to identify TyrK, PTP, and SH2 domain containing proteins in 41 published eukaryotic genome sequences [supporting information (SI) Fig. S1] (9). Using conservative thresholds, the estimated number of proteins containing these domains in a subset of these genomes is shown in Fig. 1.

The eukaryotic genomes segregate into two clear classes based on the total number of P-Tyr signaling components (Fig. 1*b*). The first class of genomes, which includes unicellular eukaryotes such as fungi, has very few of these proteins. These genomes have a handful (usually 1–15) of SH2 or PTP domain containing proteins but no TyrK proteins. The second class of genomes, which includes metazoans, has a large number (25 to >100) of all three types of proteins (TyrK, PTP, and SH2). There are no genomes that show an intermediate number or distribution of these proteins, suggesting that the appearance of the TyrK domain marks a sharp phase transition associated with coexpansion of all three domains (see *Discussion*). Interestingly, however, these classes do not breakdown along metazoan vs. non-metazoan lines. The newest sequenced genome, that of the choanoflagellate *M. brevicollis* (10), has an estimated 43 TyrK proteins, 34 PTP proteins, and 100 SH2 proteins. These numbers were determined by using fairly conservative SMART domain identification cutoff values (see *Methods*) and thus represent an estimated lower bound. Less conservative algorithms predict that this choanoflagellate may in fact have >100 TyrK proteins (data not shown). Genomic analysis indicates that Choanoflagellates are not merely a derived outgroup of metazoans; rather, they embody a unique evolutionary lineage that branched before the appearance of the first metazoans (10–13). This unicellular non-metazoan organism has a comparable number of phospho-tyrosine signaling proteins to complex multicellular metazoans such as humans (85 TyrK, 40 PTP, and 111 SH2), which strongly

Author contributions: P.B. and W.A.L. designed research; D.P. and I.L. performed research; I.L. and P.B. contributed new reagents/analytic tools; D.P. and W.A.L. analyzed data; and D.P. and W.A.L. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

‡To whom correspondence should be addressed. E-mail: lim@cmp.ucsf.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0803161105/DCSupplemental.

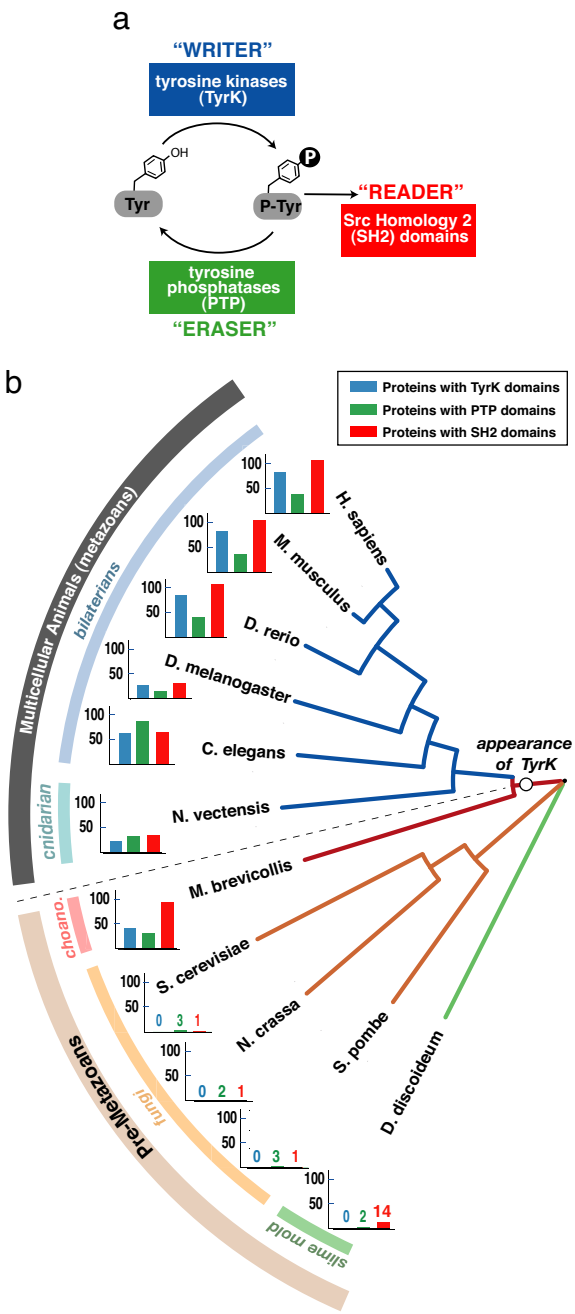

Fig. 1. Phospho-tyrosine signaling machinery in different eukaryotic lineages. (*a*) P-Tyr signaling systems are built from a three-component system comprised of tyrosine kinase (writer), tyrosine phosphatase (eraser) and Src homology 2 (reader) domains. (*b*) Number of proteins containing TyrK, PTP, or SH2 domains by species. Only choanoflagellates and metazoans have high numbers of all three domains. All other premetazoans only have small numbers of PTP and SH2 domain proteins (no TyrK). These data imply an early evolution of PTP and SH2 domains, followed by an expansion in all domains only after invention of the TyrK domain (white circle). Protein numbers are lower-bound estimates as predicted by the SMART domain identification resource.

suggests that phospho-tyrosine signaling can be used extensively for functions other than multicellular communication.

Analysis of Domain Combinations Across Genomes. How is phospho-tyrosine signaling used across different genomes? Are there differences between the functions of phospho-tyrosine signaling in metazoans and choanoflagellates? A simple way to assess functional use is to see in what higher order domain architectures P-Tyr signaling domains are used. Thus, as a simple metric for functional domain usage, we have analyzed pairwise domain combinations (14)—the sets of other domains that are found within the same ORFs as TyrK, PTP, and SH2 domains (Fig. 2*a*). The number of distinct domain combinations may more accurately reflect the diversity of functional usage of a domain (compared with total number of occurrences), because this metric avoids redundant counting of proteins with the same or related domain architectures.

Fig. 2*b* shows the occurrence of specific domain combinations within each genome. Before the evolution of modern tyrosine kinases, there is relatively little variation and diversity of SH2 and PTP domain combinations. For example, in the fungal genomes, PTP domains only occur either as single domain proteins or in combination with Rhodanase domains. However, after the emergence of modern tyrosine kinase domains, there appears to be considerable expansion in the combinatorial usage of all three P-Tyr regulatory domains; there is a core set of domain combinations that are shared in lineages from the unicellular *M. brevicollis* to humans. This core shared set includes examples like the Src family of cytoplasmic kinases that contain SH2, SH3, and TyrK domains and PDZ- and SH2-containing PTP proteins. These were presumably the most ancient domain combinations that evolved after the emergence of the modern three-part P-Tyr toolkit.

Despite this core set of shared combinations, however, we also observe a large set of lineage-specific domain combinations. In particular, the choanoflagellate *M. brevicollis* and the cnidarian *Nematostella vectensis* (star anemone) both show a large number of novel domain pair combinations that are absent in all known bilateral metazoans. For example, of the 66 domains found in combination with SH2 domains across all eukaryotes, 22 (33%) are unique to *M. brevicollis*, 9 (14%) are unique to *N. vectensis*, and 10 (15%) are unique to higher metazoans (38% are shared). Some of the domains that cooccur with SH2 domains in *M. brevicollis* suggest functions that are absent in bilateral metazoan SH2 proteins, such as extracellular communication [TNFR domain: repeats involved in growth factor binding (these domains are separated from the SH2 domain by a transmembrane domain)]; adhesion (cadherin domain: calcium-mediated adhesion); and other posttranslational modification systems, such as acetylation (histone deacetylation interaction domain: chromatin remodeling compexes). We cannot determine which of these domains combinations are divergent innovations and which might have been lost in bilaterians. Nonetheless, these data clearly show that the P-Tyr signaling machinery has evolved to fulfill distinct sets of functions in choanoflagellates (*M. brevicollis*), radially symmetric metazoans (*N. vectensis*), and bilaterian metazoans. Examples of shared and unique domain combinations in distinct lineages are shown in Fig. 3*a*. Within these genomes, one can track examples of expansion and divergence for particular domain combination architectures (Fig. 3*b*).

Discussion

Stepwise Emergence: SH2 and PTP Domains Precede the TyrK Domain. The occurrence of a few SH2 and PTP domains in many premetazoan genomes that lack TyrK domains (Fig. 1*b*) indicates that these two domains most likely evolved before TyrK domains. In several cases, these non-metazoan SH2 and PTP domains have been experimentally shown to have their traditional biochemical functions, P-Tyr binding and P-Tyr removal, respectively (15). Why would there be a functional advantage to have these reader and eraser domains even in the absence of the TyrK writer domain? This conundrum can be explained by the observation that, in some of the fungal species, Ser/Thr kinases (a more ancient family) are able to carry out sporadic tyrosine

EVOLUTION

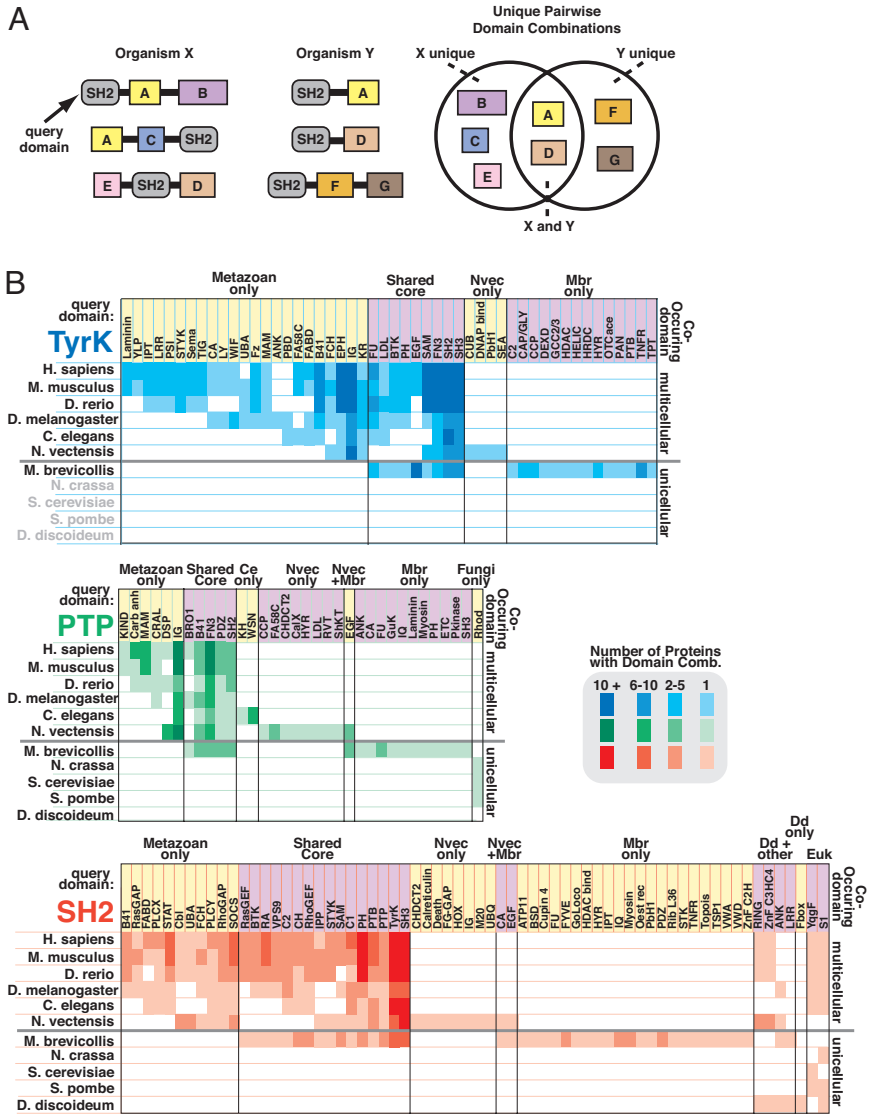

Fig. 2. Pairwise domain combinations in P-Tyr signaling proteins by species. (*a*) Domains that cooccur in the same ORF as TyrK, SH2, and PTP domains provide functional contexts for these query domains and give clues to the usage of the P-Tyr signaling machinery across different genomes. Comparison of domain combinations between species reveals conserved and divergent functions. (*b*) Clustering diagram of pairwise combinations by species, using TyrK, PTP, and SH2 as query domains. Darker boxes indicate higher number of occurences of that combination in the genome. Combinations are clustered based on cooccurence in similar sets of species. Combinations easily cluster into classes, such as shared core, metazoan only, *M. brevicollis* (choanoflagellate) only, and *N. vectensis* (cnidarian) only. For all three query domains, *M. brevicollis* (Mbr) and *N. vectensis* (Nvec) have a large set of highly divergent domain combinations that are not observed in higher metazoans.

phosphorylation (16). Phospho-tyrosine has been detected in yeast (17) and the mitogen activated protein kinases (MAPKs), which are found in all of these species, are examples of well characterized fungal proteins that require phosphorylation on both a Thr and Tyr residue on the activation loop by the upstream MAPKKs (18). Several PTP species in yeast play a key physiological role in dephosphorylating MAPKs (19). In addition, in some species such as *D. discoideum*, there is an expansion of a branch of the Ser/Thr kinase family known as tyrosine kinase-like (TKL); in some cases, the TKLs appear to carryout tyrosine phosphorylation (20). Together, these findings offer a potential explanation for why the maintenance of PTP and SH2 proteins would provide a fitness advantage: In some of the earliest unicellular eukaryotes, P-Tyr was probably used as a limited but functionally useful signaling moiety, even though there was no highly efficient catalytic machinery (e.g., the modern TyrK domain) specifically dedicated to generating this modification.

Phase Transition: Explosion of P-Tyr Machinery After Appearance of TyrK Domain. Eukaryotes segregate into two discrete classes based on number of phospho-tyrosine signaling proteins found in their genome: those that have very few, and those that have many (Fig. 1). The observation that there are no organisms with an intermediate number of P-Tyr signaling proteins suggests an all-or-none phase transition whereby an explosion in usage of all three parts occurs only upon the appearance of the third component, the modern tyrosine kinase (presumably in the last common ancestor of metazoans and choanoflagellates). This observation suggests that the full functional potential of the PTP and SH2 domains remained largely untapped until the evolution of an efficient modular tyrosine kinase. We hypothesize that

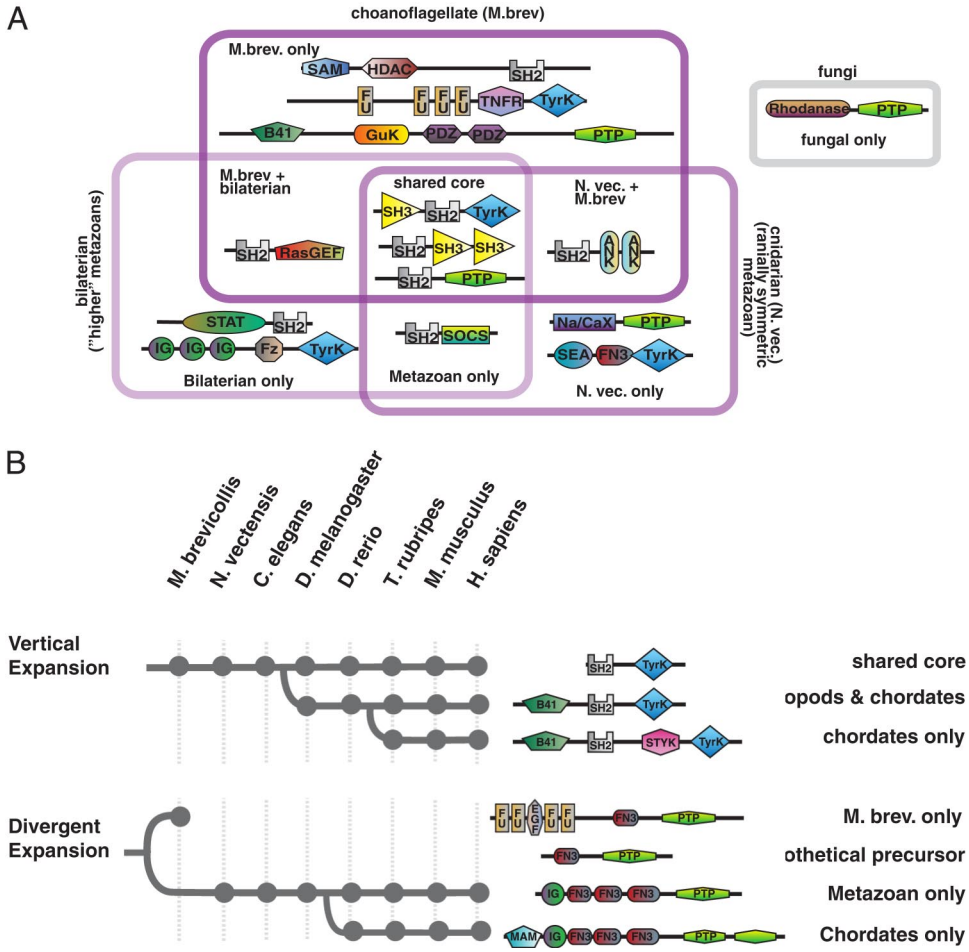

Fig. 3. Evolution of domain architectues: conservation, expansion and divergence. (*a*) Examples of domain architectures by class. Examples are organized in a Venn diagram, showing overlapping and unique architectures in different lineages. A key of domains types is given in Fig. S2. (*b*) Circles in the trees indicate occurrence of the given architecture in a particular species. (*Upper*) Gradual accretion of more complex architecture in higher organisms (vertical expansion). (*Lower*) A case where simple two domain precursors diverged into independent branches between choanoflagellates and metazoans.

PTP and SH2 domains were of finite but limited utility until the three-component toolkit of a P-Tyr eraser, reader, and writer (TyrK) was complete.

This type of abrupt phase-transition mirrors patterns observed in technology development, in cases of interdependent or complementary technologies. For example, the use of lasers for point-to-point communication, its most common use today, only began ≈20 years after the invention of the laser. This is because the laser could not be used for this function until the invention of the complementary technologies of fiber optics and optical switches (21). These two very different examples of codependent multipart systems are both characterized by punctuated expansion: There is limited usage of the early occurring components, followed by an explosive burst in usage after the occurrence of the last component.

New Signaling Currency May Provide Significant New Encoding Potential. Why does such a dramatic increase in usage of all three signaling domains occur once the TyrK domain appears? As soon as the complete three-component system is in place, suddenly there is a highly modular and efficient new system for using phosphorylated tyrosine to store and transmit information. At this key point in evolution, P-Tyr likely represented an orthogonal information storing system—it was chemically unique from other systems—and therefore might have provided ample means to generate new signaling connections free of cross interference with other regulatory systems already present in the cell. Essentially, P-Tyr may have represented a new untapped signaling currency.

Any cellular signaling currency inherently has a finite amount of information it can store and transmit; a particular reader/writer/eraser system can only be used so many times before some level of deleterious cross-talk will arise. Variation in specificity, compartmentalization, localization, and combinatorial use can increase the number of distinct pathways a signal currency can be used for, but at some point the system is still likely to become saturated. Thus, any specific signaling currency will have finite signal encoding potential (22, 23). At the time of the appearance of the TyrK domain, phospho-tyrosine was a novel signal currency that did not cross-talk with other cellular signaling systems—i.e., it was completely orthogonal—and therefore possessed the maximum signal encoding potential. We posit that the explosion in phospho-tyrosine domain usage we observe upon the appearance of the TyrK domain is the result of filling out this untapped signal encoding potential (Fig. 4).

Divergent Use of the Orthogonal P-Tyr System in Distinct Lineages. The signal encoding potential of the phospho-tyrosine signaling system has been allocated differently in *M. brevicollis*, *N. vectensis*, and the bilaterian metazoans (Fig. 2*b* and 4*a*). The divergent use of the P-Tyr machinery in the choanoflagellate lineage suggests that the branchpoint between these unicellular

EVOLUTION

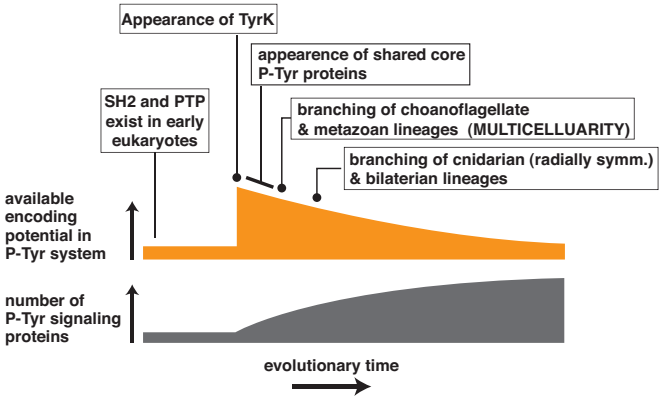

Fig. 4. Model: Timeline for the evolution of the P-Tyr signaling system. Encoding potential of the P-Tyr system is shown in orange. The number of P-Tyr signaling proteins is shown in gray. Timeline starts with early eukaryotes, which have only PTP and SH2 signaling molecules and thus only low encoding potential. Upon the appearance of TyrK, however, encoding potential of the system abruptly increases, leading to expansion of number of P-Tyr signaling proteins. Branching between choanoflagellate and metazoan lineages is likely to have occurred while there was still significant untapped encoding potential in P-Tyr signaling systems.

organisms and their multicellular relatives (metazoans) occurred after the evolution of the modern three-component P-Tyr system but before the signal-encoding potential of the P-Tyr system had been saturated. For the last common ancestor of metazoans, this new encoding potential was allocated to such novel functions as cell–cell communication. For the ancestor of the choanoflagellates, this new potential was used divergently, although extensive cell biological and genetic studies will be required to understand exactly what these functions are. We cannot rule out a model in which bilaterians, cnidarians, and choanoflagellates shared more P-Tyr functions, which were subsequently lost. Nonetheless, today the usages of these domains are still clearly partially distinct. Thus, perhaps most significant is the conclusion that P-Tyr signaling is not uniquely dedicated to any specific class of signaling events but, instead, likely represents an effective generic molecular information currency that could in principle be used for many alternative functions.

Role of P-Tyr Signaling in the Advent of Metazoan Multicellularity. With a new signaling currency, such as P-Tyr, comes a dramatic expansion of new encoding potential. This new encoding potential may bring the possibility of generating innovative functions. Therefore, we propose a model (Fig. 4) where the untapped signal encoding potential made available by the appearance of the complete reader/writer/eraser P-Tyr system provided the last common ancestor of metazoans and choanoflagellates with the potential to evolve innovative new functions. One of the most successful lineages to evolve from this common ancestor exploited this encoding potential for cell–cell communication and became the precursor to metazoans. This lineage subsequently experienced selection pressures that led to the increased diversity in signaling that was required to coordinate complex development and intricate anatomies. Thus, phospho-tyrosine signaling may have played a key role in fostering the major evolutionary transition to multicellular animals.

Abstracting this model, it may not have been P-Tyr *per se* that was required for this major transition to metazoan multicellularity. Rather, it was the signal encoding potential made available by any new, modular reader/writer/eraser system. Thus, the signal currency could have just as easily been any other orthogonal modification (e.g., alkylation and acetylation). It is also unlikely that P-Tyr signaling alone was sufficient for the evolution of metazoan multicellularity. Including P-Tyr, there are eight signaling systems that have traditionally been thought to be exclusive to metazoans (P-Tyr, nuclear hormone receptors, WNT, TGF-β, JAK/STAT, Notch/Delta, hedgehog, and toll-like receptors) (24). Except for P-Tyr, however, there is little evidence that mature versions of the canonical metazoan signaling pathways were present in the last common ancestor of choanoflagellates and metazoans (10). Thus, it may have been a convergence of several new orthogonal signaling systems that permitted the dramatic morphological innovation of multicellularity. We suspect that other major evolutionary transitions, like the emergence of eukaryotes, may have also been preceded by similar quantum increases in signal encoding potential associated with the emergence of orthogonal regulatory systems.

Methods

All domain analysis was performed with the SMART domain prediction resource using publicly available genomic data. For a detailed description, see *SI Methods* and Tables S1 and S2.

ACKNOWLEDGMENTS. We thank Nicole King, Susan Young, Kenneth Oye, Larry McCray, Sergio Peisajovich, Andrew Horwitz, Noah Helman, and Eric Meltzer for assistance and critical reading of the manuscript. This work was funded by the National Institutes of Health, the National Science Foundation, the Sandler Family Foundation, and the Rogers Family Foundation.

1. Myers MG, Jr. *et al.* (1994) Insulin receptor substrate-1 mediates phosphatidylinositol 3′-kinase and p70S6k signaling during insulin, insulin-like growth factor-1, and interleukin-4 stimulation. *J Biol Chem* 269:28783–28789.
2. Hunter T, Cooper JA (1981) Epidermal growth factor induces rapid tyrosine phosphorylation of proteins in A431 human tumor cells. *Cell* 24:741–752.
3. Weiss A, Littman DR (1994) Signal transduction by lymphocyte antigen receptors. *Cell* 76:263–274.
4. Shattil SJ, Brugge JS (1991) Protein tyrosine phosphorylation and the adhesive functions of platelets. *Curr Opin Cell Biol* 3:869–879.
5. Darnell JE, Jr (1997) Phosphotyrosine signaling and the single cell:metazoan boundary. *Proc Natl Acad Sci USA* 94:11767–11769.
6. Kouzarides T (2000) Acetylation: A regulatory modification to rival phosphorylation? *EMBO J* 19:1176–1179.
7. Rossman KL, Der CJ, Sondek J (2005) GEF means go: Turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat Rev Mol Cell Biol* 6:167–180.
8. King N, Carroll SB (2001) A receptor tyrosine kinase from choanoflagellates: Molecular insights into early animal evolution. *Proc Natl Acad Sci USA* 98:15032–15037.
9. Letunic I, *et al.* (2006) SMART 5: Domains in the context of genomes and networks. *Nucleic Acids Res* 34:D257–D260.
10. King N, *et al.* (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature*, in press.
11. Lang BF, *et al.* (2002) The closest unicellular relatives of animals. *Curr Biol* 12:1773–1778.
12. Lavrov DV, *et al.* (2005) Mitochondrial genomes of two demosponges provide insights into an early stage of animal evolution. *Mol Biol Evol* 22:1231–1239.
13. Rokas A, Kruger D, Carroll SB (2005) Animal evolution and the molecular signature of radiations compressed in time. *Science* 310:1933–1938.
14. Chothia C, Gough J, Vogel C, Teichmann SA (2003) Evolution of the protein repertoire. *Science* 300:1701–1703.
15. Zhukovskaya NV, *et al.* (2004) Dd-STATb, a Dictyostelium STAT protein with a highly aberrant SH2 domain, functions as a regulator of gene expression during growth and early development. *Development* 131:447–458.
16. Schieven G, Thorner J, Martin GS (1986) Protein-tyrosine kinase activity in Saccharomyces cerevisiae. *Science* 231:390–393.
17. Errede B, Gartner A, Zhou Z, Nasmyth K, Ammerer G (1993) MAP kinase-related FUS3 from S. cerevisiae is activated by STE7 in vitro. *Nature* 362:261–264.
18. Zhou Z, Gartner A, Cade R, Ammerer G, Errede B (1993) Pheromone-induced signal transduction in Saccharomyces cerevisiae requires the sequential function of three protein kinases. *Mol Cell Biol* 13:2069–2080.
19. Zhan XL, Deschenes RJ, Guan KL (1997) Differential regulation of FUS3 MAP kinase by tyrosine-specific phosphatases PTP2/PTP3 and dual-specificity phosphatases MSG5 in Saccharomyces cerevisiae. *Genes Dev* 11:1690–1702.
20. Goldberg JM, *et al.* (2006) The dictyostelium kinome—Analysis of the protein kinases from a simple model organism. *PLoS Genet* 2:e38.
21. Townes CH (December 23, 1968) The laser—What it can do. *US News and World Report*, p 81.
22. Zarrinpar A, Bhattacharyya RP, Lim WA (2003) The structure and function of proline recognition domains. *Sci STKE* 2003:RE8 (abstr).
23. Itzkovitz S, Tlusty T, Alon U (2006) Coding limits on the number of transcription factors. *BMC Genomics* 7:239.
24. Pires-daSilva A, Sommer RJ (2003) The evolution of signalling pathways in animal development. *Nat Rev Genet* 4:39–49.