
The following resources related to this article are available online at <http://stke.sciencemag.org>.
This information is current as of 2 May 2012.

- Article Tools** Visit the online version of this article to access the personalization and article tools:
<http://stke.sciencemag.org/cgi/content/full/sigtrans;5/222/ra35>
- Supplemental Materials** "Supplementary Materials"
<http://stke.sciencemag.org/cgi/content/full/sigtrans;5/222/ra35/DC1>
- Related Content** The editors suggest related resources on *Science's* sites:
<http://stke.sciencemag.org/cgi/content/abstract/sigtrans;5/222/pc9>
<http://stke.sciencemag.org/cgi/content/abstract/sigtrans;5/215/rs1>
<http://stke.sciencemag.org/cgi/content/abstract/sigtrans;3/128/eg5>
<http://stke.sciencemag.org/cgi/content/abstract/sigtrans;3/127/eg4>
<http://stke.sciencemag.org/cgi/content/abstract/sigtrans;2/87/ra50>
- References** This article has been **cited by** 1 article(s) hosted by HighWire Press; see:
<http://stke.sciencemag.org/cgi/content/full/sigtrans;5/222/ra35#BIBL>
This article cites 51 articles, 19 of which can be accessed for free:
<http://stke.sciencemag.org/cgi/content/full/sigtrans;5/222/ra35#otherarticles>
- Glossary** Look up definitions for abbreviations and terms found in this article:
<http://stke.sciencemag.org/glossary/>
- Permissions** Obtain information about reproducing this article:
<http://www.sciencemag.org/about/permissions.dtl>

Genomic Survey of Premetazoans Shows Deep Conservation of Cytoplasmic Tyrosine Kinases and Multiple Radiations of Receptor Tyrosine Kinases

Hiroshi Suga,^{1,2*} Michael Dacre,³ Alex de Mendoza,^{1,2} Kamran Shalchian-Tabrizi,⁴ Gerard Manning,^{3*} Iñaki Ruiz-Trillo^{1,2,5*}

The evolution of multicellular metazoans from a unicellular ancestor is one of the most important advances in the history of life. Protein tyrosine kinases play important roles in cell-to-cell communication, cell adhesion, and differentiation in metazoans; thus, elucidating their origins and early evolution is crucial for understanding the origin of metazoans. Although tyrosine kinases exist in choanoflagellates, few data are available about their existence in other premetazoan lineages. To unravel the origin of tyrosine kinases, we performed a genomic and polymerase chain reaction (PCR)-based survey of the genes that encode tyrosine kinases in the two described filasterean species, *Capsaspora owczarzaki* and *Ministeria vibrans*, the closest relatives to the Metazoa and Choanoflagellata clades. We present 103 tyrosine kinase-encoding genes identified in the whole genome sequence of *C. owczarzaki* and 15 tyrosine kinase-encoding genes cloned by PCR from *M. vibrans*. Through detailed phylogenetic analysis, comparison of the organizations of the protein domains, and resequencing and revision of tyrosine kinase sequences previously found in some whole genome sequences, we demonstrate that the basic repertoire of metazoan cytoplasmic tyrosine kinases was established before the divergence of filastereans from the Metazoa and Choanoflagellata clades. In contrast, the receptor tyrosine kinases diversified extensively in each of the filasterean, choanoflagellate, and metazoan clades. This difference in the divergence patterns between cytoplasmic tyrosine kinases and receptor tyrosine kinases suggests that receptor tyrosine kinases that had been used for receiving environmental cues were subsequently recruited as a communication tool between cells at the onset of metazoan multicellularity.

INTRODUCTION

Multicellularity evolved several times in eukaryotes (1–3). Of particular interest is the multicellular system of metazoans, composed of highly specialized cells that perform coordinated interaction and communication, which leads to various highly complex and mobile forms. Although it has been hypothesized that the evolution of metazoan multicellularity was driven by the emergence of genes whose products were involved in cell adhesion, cell differentiation, and cell-to-cell communication, it is now evident that several of these key genetic modifications occurred well before the origin of animals (2, 4–8).

Many multicellular-specific functions, including cell-to-cell communication and control of cell proliferation and differentiation, are served by protein tyrosine kinases (TKs or PTKs) (9–11). TKs are divided into two types: receptor TKs (RTKs) and nonreceptor or cytoplasmic TKs (CTKs). RTKs mostly receive their specific ligands through their extracellular domains and initiate signal transduction cascades that are mediated by phosphorylated tyrosine residues, whereas CTKs act within cells and transmit the phosphotyrosine signals initiated by receptors (9, 10). One of the most remarkable features of TKs is their high degree of structural diversity. Most TKs are made up of multiple protein domains and motifs other than

the catalytic (or kinase) domain (KD). Such divergent architectures are considered to have been generated by gene duplication and domain shuffling (9, 10, 12). The extracellular regions of RTKs, which are composed of various domains and motifs, are thought to bind to specific ligands directly, and their expression is mostly restricted to specific cell types in the organism (10). Moreover, KDs show an increased versatility (that is, they are frequently combined with other domains), especially in the Metazoa (13–15). Thus, the diversity of TKs, which consist of approximately 30 families [which correspond to subfamilies in some of our previous publications (16–19)] with different protein domain organizations, may reflect the complexity of the phosphotyrosine-based signaling system in metazoans.

TKs are members of the eukaryotic protein kinase (ePK) superfamily and are likely to have evolved from one of the ancestral ePK groups (20–22). TKs phosphorylate substrates only on tyrosine residues, whereas other ePKs phosphorylate mostly on serine and threonine residues and are thus classified as serine-threonine kinases (STKs) (21). Although some STKs can also phosphorylate tyrosine residues (23), the TK group is discriminated from other groups of ePKs by their overall sequence similarity and by the presence of a characteristic catalytic loop motif (24).

The repertoire of CTK and RTK families is highly conserved across all metazoans, including sponges (17, 18, 25). In accordance with their substantial involvement in intercellular communication and the control of cell proliferation and differentiation, TKs had been thought to be exclusive to metazoans (21); however, this notion was dispelled by the discovery of an elaborate collection of TKs in the unicellular and colonial choanoflagellates (7, 19, 24, 26–28), which are the closest relatives to the Metazoa. Moreover, previous reports had suggested the presence of TKs in some

¹Institut de Biologia Evolutiva (UPF-CSIC), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain. ²Departament de Genètica, Universitat de Barcelona, 08024 Barcelona, Spain. ³Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. ⁴Microbial Evolution Research Group, Department of Biology, University of Oslo, 0316 Oslo, Norway. ⁵Institució Catalana per a la Recerca i Estudis Avançats, 08010 Barcelona, Spain.

*To whom correspondence should be addressed. E-mail: hiroshi.suga@ibe.upf-csic.es (H.S.); manning@manninglab.org (G.M.); inaki.ruiz@icrea.es (I.R.-T.)

non-opisthokont lineages, including the amoebozoans *Dictyostelium discoideum* and *Entamoeba histolytica*, the green alga *Chlamydomonas reinhardtii*, and the oomycete *Phytophthora infestans* (Fig. 1) (29, 30), although they are far less numerous than those of metazoans and choanoflagellates. These findings prompted us to explore whether a TK-based signaling system is also found in the Filasterea, the sister group to the Metazoa and Choanoflagellata (MeCh) (Fig. 1) (4, 5). The Filasterea consist of only two known species: *Capsaspora owczarzaki*, a symbiotic amoeba that dwells in the snail *Biomphalaria glabrata* (31), and *Ministeria vibrans*, a free-living marine protist (4, 32). Although several putative TKs have been reported in expressed sequence tag (EST) data from *M. vibrans* (4), a comprehensive genomic analysis of filasterean TKs has not yet been conducted.

Here, we analyzed the whole genome sequence of *C. owczarzaki* and performed targeted polymerase chain reaction (PCR)-based cloning of *M. vibrans* complementary DNAs (cDNAs) to provide a picture of the earlier stage of TK evolution focused on these unicellular relatives of metazoans. Through comprehensive phylogenetic approaches including two filastereans and other premetazoans, together with a detailed comparison of their protein domain organizations, we have elucidated in detail the early evolution of TK diversity. Our data show very different divergence patterns between CTKs and RTKs. The basic repertoire of CTKs had already been established before the separation of filastereans and the MeCh, whereas the RTKs show an extensive and independent diversification in each of the filasterean, choanoflagellate, and metazoan clades.

RESULTS

Diversity of TKs in the Filasterea

To better understand the early evolution of TKs, we conducted a genomic and PCR-based survey for TK-encoding genes in the Filasterea. We identified 103 putative TK-encoding genes in the whole genome sequence of *C. owczarzaki* (Fig. 2 and fig. S1). Of these, 92 are predicted to be RTKs, which contain an intracellular KD, a transmembrane (TM) segment, a signal peptide, and known protein domains and motifs in the extracellular

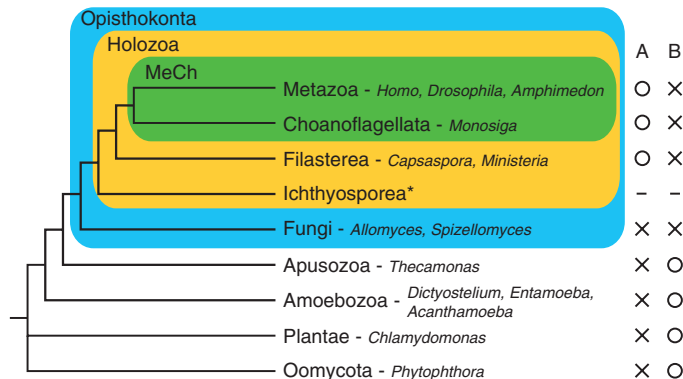


Fig. 1. Schematic phylogenetic tree of eukaryotes depicting the presence of group A and B TKs. A widely accepted consensus phylogeny (32, 55) based on phylogenomic studies is shown. The presence (circle) and absence (x) of the group A and B TKs are shown on the right. Representative genera used in this study are shown after the clade names. The holozoan TKs (group A) most likely originated from the group B TKs. The asterisk indicates that the genome sequencing of ichthyosporeans is ongoing (36).

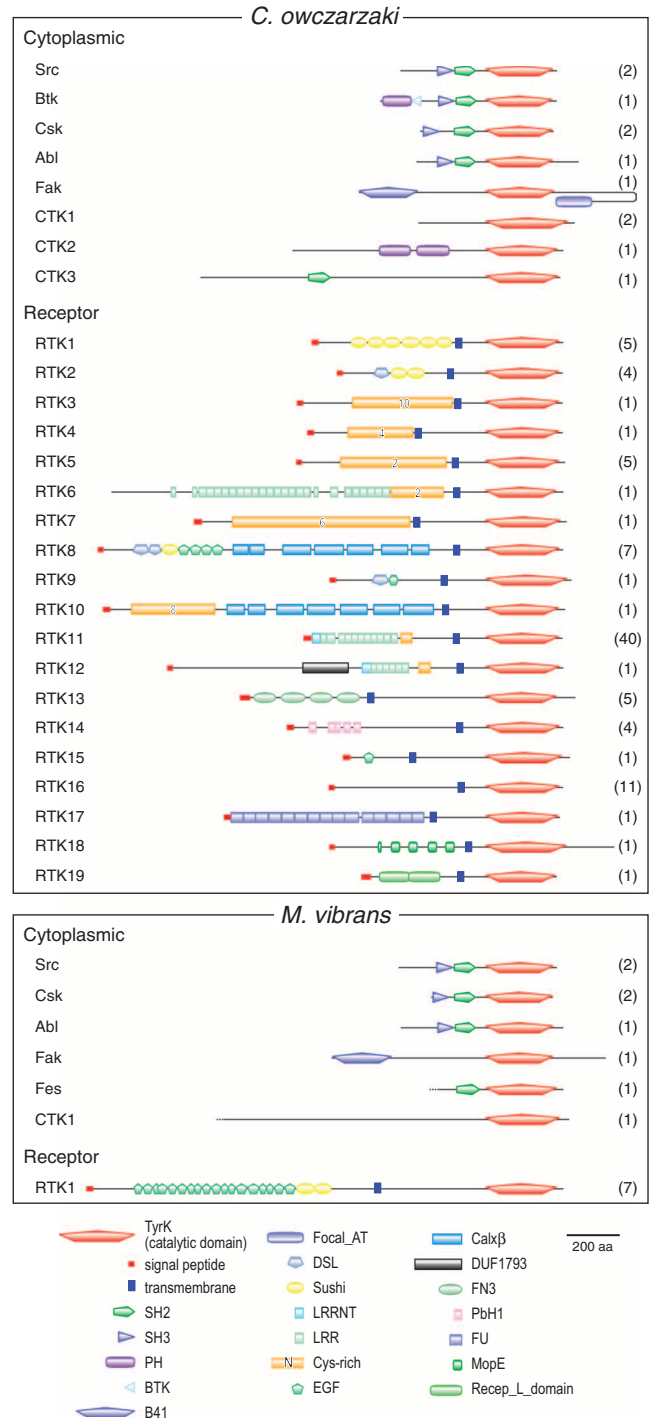


Fig. 2. Classification of *C. owczarzaki* and *M. vibrans* TKs. One hundred three *C. owczarzaki* TKs and 15 *M. vibrans* TKs were divided into 27 and 7 distinct families, respectively, by domain architecture and KD phylogeny. A typical domain organization is schematically displayed for each family, with the number of the genes that belong to each family shown in parentheses. The domain organizations of all the proteins can be found in fig. S1. The Pfam or SMART domain names are shown on the bottom. The sizes of the illustrations are proportional to the actual lengths of amino acid sequences. Scale bar, 200 amino acid residues.

region. The other 11 TKs lack a signal peptide and a TM and are thus classified as CTKs. We also isolated seven RTKs and eight CTKs by a PCR-based survey in *M. vibrans* (Fig. 2 and fig. S1).

According to their protein domain organization and the phylogenetic relationship between KDs, we classified *C. owczarzewski* TKs into 27 families (8 CTK and 19 RTK) and *M. vibrans* TKs into 7 families (6 CTK and 1 RTK) (see Materials and Methods for the classification criteria). The CTK repertoire of *M. vibrans* is similar to that of *C. owczarzewski*, but the only RTK family identified in *M. vibrans* is unique and does not share its domain architecture with any of the known RTK families, including those of *C. owczarzewski*. RTKs of *C. owczarzewski* show an extensive divergence in their architectures, containing 19 families with distinct organizations of protein domains. Among them, the RTK11 family is the largest with 40 genes, which encode 4 to 46 leucine-rich repeats (LRRs) (fig. S1). Many of the extracellular domains found in these filasterean RTKs, such as the epidermal growth factor (EGF)-like and fibronectin type III (FN3) domains, generally interact with other proteins, including extracellular ligands or other receptors.

Phylogenetic position of TKs in the ePK superfamily

To elucidate the origin of TKs, we first conducted a preliminary phylogenetic analysis of the ePK superfamily, including TKs and STKs. We included 23 diverse STK families, with a particular focus on close relatives of TKs (21, 22), as well as a partial TK repertoire from filastereans, metazoans, and the choanoflagellate *Monosiga brevicollis*, in addition to the previously reported TKs from pre-opisthokonts (29, 30). We also included 5 putative KD sequences of TKs that were found in the genome sequence of the apusozoan *Thecamonas trahens* (33), which is the putative sister group to the opisthokonts (Fig. 1) (34), and another 17 found in the genome of the amoebozoan *Acanthamoeba castellanii*. We found no TK-encoding genes in any of the available fungal genomes, including those of the two early-branching species *Allomyces macrogynus* and *Spizellomyces punctatus* (35), which have been sequenced under the UNICORN project (36). This suggests that TKs originated before the divergence between metazoan and fungi and were secondarily lost from fungi (30), if the consensus phylogeny is correct (Fig. 1).

The Bayesian phylogenetic tree inferred from the alignment of KD sequences provided a nearly maximum statistical support (a Bayesian posterior probability of 0.97) to a monophyletic clustering of holozoan [metazoan, choanoflagellate, and filasterean (Fig. 1)] TKs (group A TKs) (Fig. 3 and fig. S2). We deduced a similar tree with the maximum likelihood (ML) method (fig. S3A). Removal of TK sequences from the data set did not alter the overall topology of the ePK tree (fig. S4). The group A TKs branch within another type of TK (group B TKs), which includes all of the putative pre-opisthokont TKs except for some amoebozoan TKs and shows a diversification independent of that of group A. The monophyletic clustering of group A and B TKs is supported with a maximum Bayesian posterior probability of 1.0.

To further corroborate this topology, we performed two additional tests. First, we challenged the position of group A by statistically evaluating the 144 alternative topologies that are different from each other by the group A position; all topologies that placed group A outside group B were statistically rejected (fig. S5). Second, we statistically assessed the diversification pattern between group A TKs, group B TKs, and the STK group also by bootstrapping: 2000 hypothetical ML trees generated by bootstrapping mostly support the relationship among these three groups suggested by the ML tree in Fig. 3 (fig. S6). It is thus likely that the canonical TKs of metazoans, choanoflagellates, and filastereans originated from pre-opisthokont TKs and independently diversified. Any of the experimentally proven TKs from *D. discoideum* (29) and a few predicted TKs from

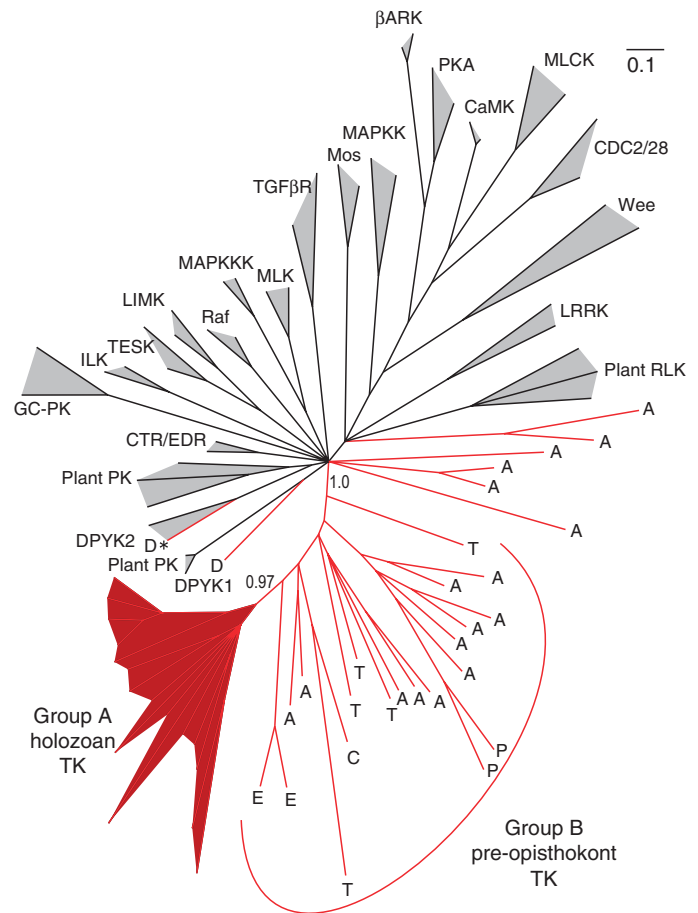


Fig. 3. Phylogenetic positions of TKs in the ePK superfamily. A Bayesian tree was inferred from 173 amino acid sites of the KD alignment. Twenty-three diverse STK families were chosen (21, 56). MLK, mixed lineage kinase; GC-PK, guanylate cyclase-coupled protein kinase; CTR, constitutive triple response; EDR, enhanced disease resistance; DPYK, *D. discoideum* protein tyrosine kinase; ILK, integrin-linked kinase; CaMK, calcium/calmodulin-dependent protein kinase; MLCK, myosin light chain kinase; β ARK, β -adrenergic receptor kinase; LRRK, leucine-rich repeat kinase; Plant RLK, plant receptor-like kinase; TESK, testis-specific protein kinase. The Bayesian posterior probabilities are shown at two key branches. Predicted or experimentally proven TKs are shown by red lines. The asterisk indicates that the TK motif is not clear but was experimentally shown to have only TK activity (29). Holozoan group A TKs are highlighted by red shading. Abbreviated species names are as follows: A, *A. castellanii*; C, *C. reinhardtii*; D, *D. discoideum*; E, *E. histolytica*; T, *T. trahens*; and P, *P. infestans*. Further details can be found in fig. S2.

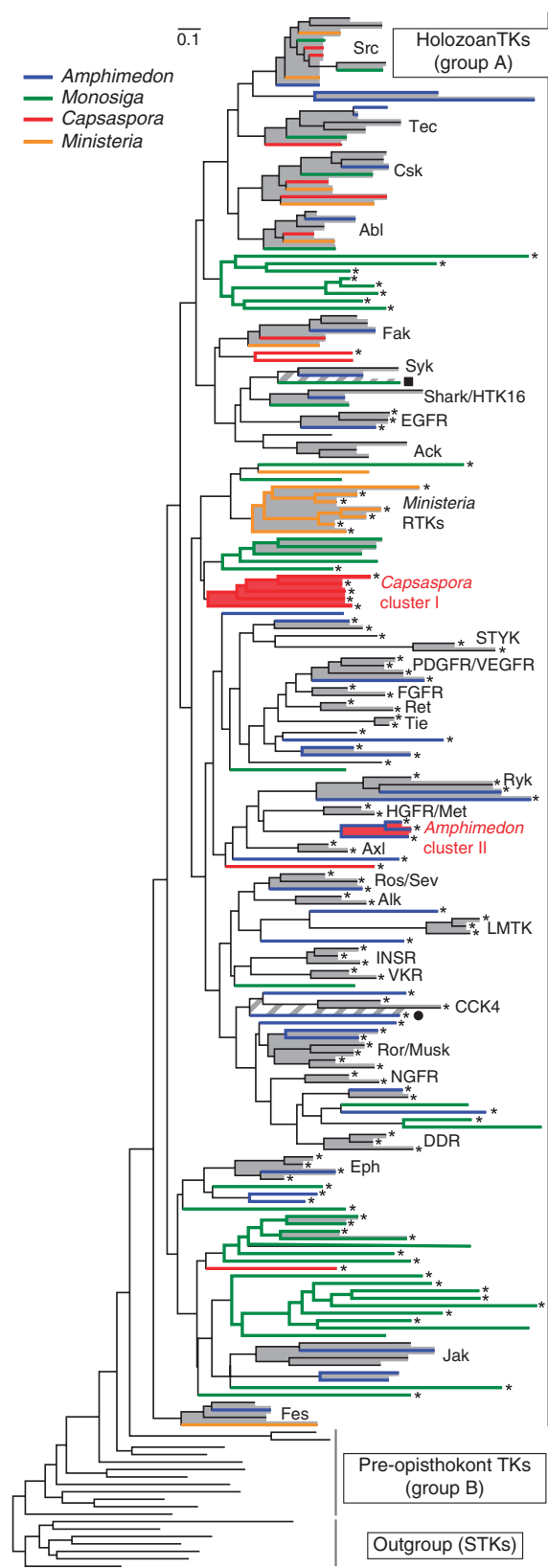
A. castellanii do not branch within either group A or group B, but they are included in the ancestral STK class, reminiscent of multiple origins of TKs. However, ML trees optimized either under the constraint that all of the *A. castellanii* TKs included within group A and B, or under the constraint that all the amoebozoan TKs (that is, those from both *A. castellanii* and *D. discoideum*) included within group A and B are not denied by the one-tailed, Kishino-Hasegawa test (fig. S3) (37). Therefore, we cannot confidently conclude whether TKs and their catalytic loop motifs evolved once or multiple times in eukaryote evolution.

Fig. 4. Comprehensive phylogenetic tree of holozoan TKs. The ML tree was inferred from an alignment consisting of 190 holozoan TKs (group A), 12 pre-opisthokont TKs (chosen from the group B TKs in Fig. 3), and 7 ancestor STKs (chosen from Fig. 3 as an outgroup). All TKs found in the genomes except for highly similar paralogs were included. The genes belonging to clusters I (*C. owczarzaki*) and II (*A. queenslandica*) (red shading) are not fully included but were separately analyzed. The sequences of *A. queenslandica*, *M. brevicollis*, *C. owczarzaki*, and *M. vibrans* are in blue, green, red, and orange lines, respectively. Families are shaded in gray, and the names of 29 major metazoan families are indicated. Black square indicates a putative *M. brevicollis* ortholog of the Syk family, which is classified into an independent family as a result of a minor architectural difference (gray stripe). Jak family proteins share the same domain architecture with this protein. It can thus also be a Jak homolog whose KD had been replaced with that of a Syk-related family. Black circle indicates a possible *A. queenslandica* ortholog of the CCK4 family, as suggested by another ML analysis excluding biased or fast-evolving sequences (fig. S10). Asterisks indicate RTKs. The detailed tree with individual gene names, bootstrap values, and protein architectures are shown in fig. S7.

Evolution of architectural diversity of the holozoan TKs

To further explore the evolutionary history of holozoan TKs, we performed a more focused ML phylogenetic analysis on the TKs from *C. owczarzaki* and *M. vibrans* as well as those from the choanoflagellate *M. brevicollis* and some representative metazoans, including *Homo sapiens*, *Drosophila melanogaster*, and the sponge *Amphimedon queenslandica* (Fig. 4 and fig. S7). TKs showing identical domain architecture and highly similar KD sequences with others in a single organism are considered to be recently duplicated paralogs, which were removed to simplify the analysis. Two large monophyletic clusters of *C. owczarzaki* and *A. queenslandica* TKs (clusters I and II, respectively) were also trimmed and analyzed separately (Fig. 5 and figs. S8 and S9). Caution should be exercised in using published genome annotations, which were usually automatically performed, because long protein sequences such as those of TKs are often mispredicted. Thus, we manually revised the published *M. brevicollis* and *A. queenslandica* TK-encoding gene predictions (24, 25) and optimized the predictions of their domain organizations. To minimize phylogenetic artifacts, we also inferred an ML tree that excluded sequences with biased amino acid composition and rapidly evolving sequences, using as an outgroup the group B TKs, which are the closest relatives of group A TKs (fig. S10); the tree topology was mostly in agreement with the original tree including the complete data set (Fig. 4), except for the position of *A. queenslandica* immunoglobulin (Ig)7-RTK, which is suggested to be a colon carcinoma kinase 4 (CCK4) family homolog (fig. S10). Consistent with the preliminary analysis (Fig. 3), the tree showed that all holozoan TKs (group A) form a clade distinct from that of the group B pre-opisthokont TKs (Fig. 4).

As shown by previous PCR-based studies of a freshwater sponge (17, 18), the *A. queenslandica* genome strengthened the notion that the basic repertoire of the CTK and RTK families found in most eumetazoan lineages (cnidarians and bilaterians) had already been established before the divergence of eumetazoans and sponges (25); 15 (16 if the CCK4 family is included) of the 29 major eumetazoan TK families were identified in the genome of *A. queenslandica* (Fig. 4). Phylogenetic analyses suggest gene losses in sponge rather than innovations in eumetazoans, even for eumetazoan families that sponge lacks (17, 18). Their protein domain



Downloaded from stke.sciencemag.org on May 2, 2012

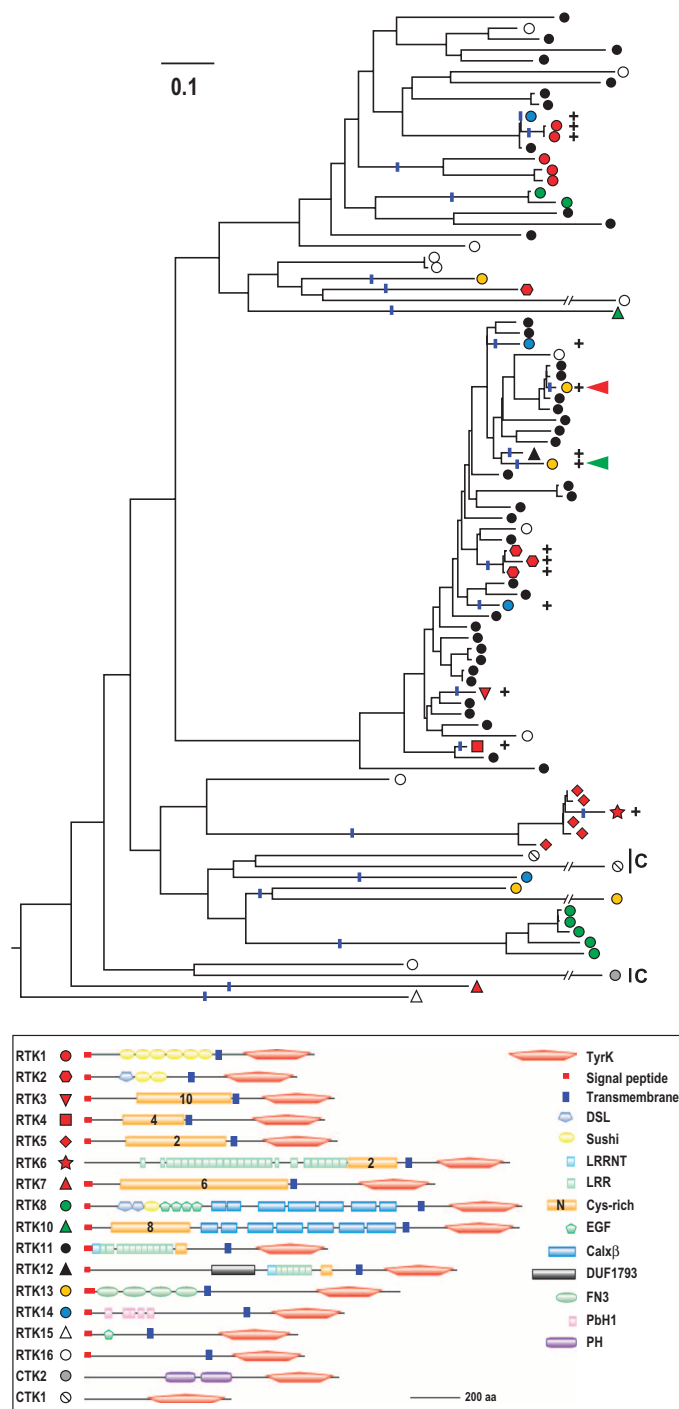


Fig. 5. Expansion of *C. owczarzaki*-specific TKs in cluster I. The tree was inferred by the ML method. Different *C. owczarzaki* TK families are indicated by distinct symbols and colors. A typical domain organization is shown for each family in the lower panel. The abbreviated names of the protein domains or motifs were taken from the SMART or Pfam databases, except for "Cys-rich," which are not mapped to any of the known protein domains or motifs but just contain periodically arranged cysteines. The TKs with Cys-rich domains were classified by the number of cysteines (shown in the diagrams) usually seen in a repeating unit. C indicates CTKs. A cross indicates an RTK whose extracellular domain organization has been entirely altered by domain duplication, shuffling, and conversion that occurred relatively recently (more than 85% KD identity with its counterpart). Blue bars indicate the evolutionary points or branches where extracellular domain alterations are supposed to have occurred by parsimonious inference. Red and green arrowheads indicate CoPTK-207 and CoPTK-219, respectively, whose extracellular sequences are highly similar (see fig. S14). The detailed tree with the bootstrap supports and protein architectures is presented in fig. S8.

the 10 major metazoan CTK families (Src, Tec, Csk, Abl, Fak, and Fes) on the basis of KD sequence similarity and overall domain architecture, whereas none of the RTKs can be confidently assigned to any metazoan family by the same criteria (Figs. 2 and 4 and fig. S11). Similarly, the PCR-based sampling of the TK repertoire of *M. vibrans* identified eight CTKs, of which seven are assigned to five metazoan families, and seven RTKs, all of which belong to a unique family (Figs. 2 and 4 and fig. S11). A similar situation was also seen in choanoflagellates, which have orthologs of at least six metazoan CTK families (Fig. 4 and fig. S11), and a large RTK expansion that appears independent of both filastereans and metazoans (19, 24).

Diversity and commonality of the holozoan TK repertoires

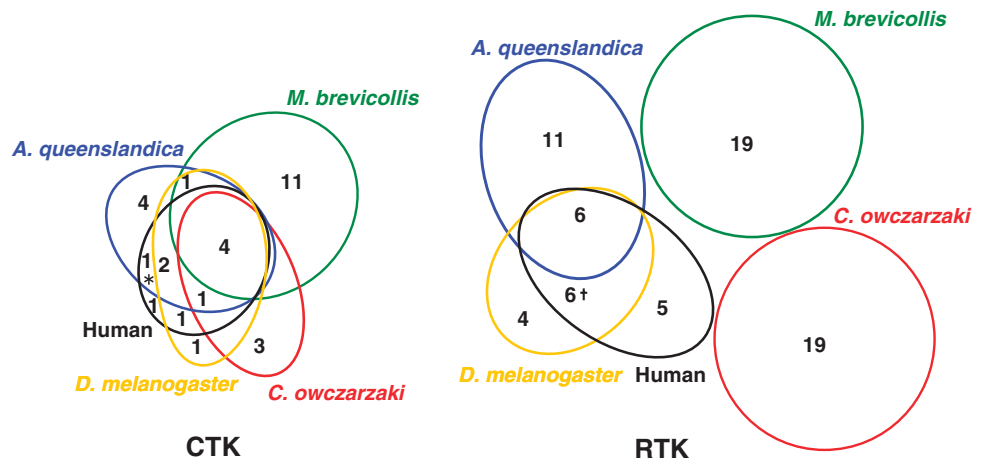
We summarized the numbers of unique and shared TKs among these five holozoan genomes, *H. sapiens*, *D. melanogaster*, *A. queenslandica*, *M. brevicollis*, and *C. owczarzaki* (Fig. 6 and fig. S11), highlighting the ancient establishment of CTK families and the rapid turnover of RTKs in the premetazoan stage. Of the 10 CTK families that are mostly in common among the three metazoans (fig. S11), 6 are also present in at least one filasterean. Another ML analysis focusing only on CTK families confirmed this finding (fig. S12). The basic Src homology 2 (SH2)–SH3–KD architecture of the Src-related families (Src, Tec, Csk, and Abl) was established before the divergence of filastereans and the MeCh and had already diversified into the four families. The presence of homologs of Fes and Fak in filastereans demonstrates that they were secondarily lost from *M. brevicollis*, although a putative Fes homolog is present in the choanoflagellate *Codosiga gracilis* (19). In addition, we identified a previously unreported Shark/HTK16 homolog in the *M. brevicollis* genome. This gene, together with the possible Syk ortholog or Jak homolog of this choanoflagellate (Fig. 4 and fig. S11), brings the number of metazoan CTK families with a premetazoan origin to eight. In addition to these common CTK families, each holozoan clade has also increased the repertoire independently by gene duplication and domain shuffling (fig. S11). In contrast, we did not observe a clear orthology between the RTKs of metazoans, choanoflagellates, and filastereans (Fig. 6), either by the phylogeny of KD sequences or by comparison of domain organization (Fig. 4 and figs. S7 and S11).

We then focused on the pattern of clade-specific RTK expansion within the Choanoflagellata. As described earlier, most metazoan-specific families

organizations are also well conserved (figs. S7 and S11) (17, 25). Moreover, this anciently established repertoire has not suffered a marked change in metazoan evolution (17, 18). Some metazoans, however, exceptionally diversified their TK repertoire independently in each lineage (22, 25) by creating TKs with diverse architectures, for example, in *A. queenslandica* (Fig. 4, cluster II, and fig. S9).

The TK repertoire of filastereans shows a divergence pattern markedly different from that of sponges; most CTKs are clearly orthologous to 6 of

Fig. 6. Diversity and commonality of holozoan TK families. Numbers of TK families of five holozoans are summarized in Venn diagrams. The numbers were counted on the basis of the classification shown in fig. S11. Gene duplications that occurred within a family were not taken into account. An asterisk indicates that although the Syk family is shared only by *H. sapiens* and *A. queenslandica*, having the same domain architecture, a possible ortholog is present in *M. brevicollis* with a B41 domain added. This gene could also be a Jak homolog. See the legend to Fig. 4 for further details. The dagger indicates that ML analysis on the data set excluding biased or fast-evolving sequences (see fig. S10) suggests the *A. queenslandica* Ig7-RTK as a CCK4 family gene, although we do not include it here according to the ML analysis on the complete data set (Fig. 4 and fig. S7).



seem to have been generated within early metazoans before the divergence between sponges and the rest of the metazoans. To determine whether a similar pattern was seen in choanoflagellates, we compared the complete TK repertoire of *M. brevicollis* (family Codonosigidae) with partial ones of *Monosiga ovata* (family Salpingoecidae) (38), *C. gracilis* (family Codonosigidae), and *Stephanoeca diplocostata* (family Acanthoecidae), which were obtained by PCR surveys (fig. S13) (19). Five choanoflagellate-specific CTK families and three RTK families are shared between *M. brevicollis* and other choanoflagellate species (figs. S11 and S13). In contrast, three CTK and seven RTK families found in *M. ovata*, *C. gracilis*, and *S. diplocostata* are not present in the whole genome of *M. brevicollis* (fig. S13). Thus, unlike in metazoans, the generation of novel TK families early in choanoflagellate evolution continued to at least beyond the divergence of these three taxonomic families of the Choanoflagellata.

Extensive lineage-specific diversification of the *C. owczarzaki* RTKs

Of the 92 *C. owczarzaki* RTKs, 88 belong to a single cluster, cluster I (Figs. 4 and 5 and fig. S8), on the basis of KD sequence similarity. They display diverse domain organizations, which is indicative of frequent domain shuffling. Given that their common ancestor belonged to the RTK11 family (Fig. 5), which is most common in this cluster (Fig. 2), we parsimoniously estimated 22 extracellular architecture alterations in total (Fig. 5). Fourteen RTKs show particularly high KD sequence identities (more than 85%) to their close relatives that have distinct extracellular domain organizations, which suggest recent duplication and domain swapping (Fig. 5). It is thus likely that frequent gene duplication and domain shuffling that occurred in *C. owczarzaki* augmented the diversity of RTKs of this protist. The *A. queenslandica* RTKs in cluster II show a similar pattern of diversification (fig. S9), indicating that the rapid expansion of RTKs is not a phenomenon specific to unicellular protists but occurred also in the earliest branching metazoans with primitive multicellularity (39, 40). These mechanisms of quick expansion of RTKs are likely to have contributed to increasing the variety of protein tyrosine phosphatases (PTPs), which are antagonists of TK signaling (41). For example, the extracellular region of the *C. owczarzaki* RTK CoPTK-207 (Fig. 5) shows a high sequence identity not only to that of the other RTK CoPTK-219 (Fig. 5) but also to that of a PTP (CoPTP-14FN3) that has two PTP catalytic domains instead of a KD (fig. S14).

DISCUSSION

Our data show that holozoan TKs originated from ancestral pre-opisthokont TKs and diversified independently, and that fungi secondarily lost TKs. The two known filastereans, *C. owczarzaki* and *M. vibrans*, have TK repertoires that rival those of choanoflagellates and metazoans in size. This suggests that an elaborate TK repertoire emerged before the divergence of filastereans from choanoflagellates and metazoans. Moreover, our data show that the early evolution of RTKs is markedly different from that of CTKs. Of the 10 common metazoan CTK families, eight were already in place even before the divergence of the Filasterea from the McCh. In contrast to the early maturation of CTKs, no orthology was found among RTKs in metazoans, choanoflagellates, and filastereans. This indicates that the last common ancestor of holozoans had (i) one or a few RTKs, from which each clade has expanded the repertoire by frequent gene duplication, and duplication and shuffling of the extracellular domains, or (ii) many RTKs, but extensive domain shuffling, gene loss, or KD conversion in each clade has obscured the orthology on the basis of the architectures of the extracellular regions and KD sequence homology. In all of these three lineages, RTKs substantially outnumber CTKs, and some families are highly expanded.

Within metazoans, we saw substantial RTK generation at the earliest stage of metazoan evolution before the split between sponges and eumetazoans but far fewer changes in domain architecture in more evolutionarily recent times. Although some metazoan lineages (for example, sponges) seem to have expanded their TK repertoire also recently (fig. S9), most recent large-scale changes are limited to simple duplications, including whole-genome duplications within vertebrates (12, 16, 42, 43). On the other hand, choanoflagellates seem to have continued the diversification of TKs even after the divergence of their three taxonomic families. However, more data are necessary to assess the continued TK diversification within choanoflagellates and filastereans.

We have no experimental evidence for the biological functions of filasterean TKs. Although a previous publication suggested that the phosphotyrosine-mediated signaling system of choanoflagellates might serve a function related to the cell cycle or cell survival (44), still nothing is known about the extracellular signals received by the RTKs. However, the clade-specific divergence pattern of RTKs is consistent with the hypothesis that they act to detect changes in the extracellular environment or

to recognize and catch prey (44), because each organism has to be adapted to its own environment and nutrient conditions. This might explain why the extracellular regions of RTKs in filastereans and choanoflagellates, which are usually exposed to the environment, are highly divergent in each lineage. The metazoan RTK repertoire, however, seems to be largely stable after the initial expansion, with a unique set of metazoan RTKs retained after the emergence of multicellularity. The recruitment of this initial set for functions such as intercellular communication and control of cell adhesion, instead of for receiving environmental cues, might be a key to the evolution of metazoan multicellularity. The sponge-specific RTK repertoire, which diversified independently of the common RTK set of metazoans, may reflect the constant exposure of their tissues to the environment (45), in contrast to eumetazoans, in which the RTKs are generally exposed to internal fluids. The more stable evolution of CTKs would then reflect a relatively stable intracellular environment; CTKs generally act downstream of RTKs or other receptors to transmit extracellular signals (46).

We have shown the extensive diversification of holozoan TKs after the split from fungi, as well as the differences in the patterns of expansion between RTKs and CTKs before and after the evolution of multicellularity. We hypothesize that the generation of a new RTK repertoire and its use to encode cell-to-cell communication tools has been one of the key genetic changes at the transition from the unicellular to the multicellular system. The genome sequencing of the other filasterean *M. vibrans*, as well as the remaining holozoan taxon, the Ichthyosporaea (Fig. 1), and other preopisthokonts (36), together with functional assays on TKs, will further clarify the evolutionary origin and functional transition of TK signaling.

MATERIALS AND METHODS

Cultures of filastereans

Live cultures of *C. owczarzaki* and *M. vibrans* were purchased from the American Type Culture Collection (ATCC) and maintained at 23°C in ATCC medium 1034 and at 17°C in ATCC medium 1525.

Cloning of TK-encoding genes from *C. owczarzaki* and *M. vibrans*

M. vibrans cDNAs were obtained either by PCR with degenerate primers as described previously (17, 19) or by searching available ESTs (4). The 15 obtained cDNAs were extended to the 5' and 3' termini by rapid amplification of cDNA ends (RACE) (47). All of the putative TK-encoding sequences previously found in ESTs (4), except the clearly redundant ones, were included in this study. The uncertain exon-intron boundaries and 5' and 3' ends of five *Capsaspora* TKs (CoPTK-96, CoPTK-102, CoPTK-219, CoPTK-235, and CoPTK-238) were confirmed by PCR and RACE. Sequences were deposited in GenBank under accession numbers AB591048 to AB591054.

Data mining

The genome sequences and predicted protein collections of *H. sapiens*, *D. melanogaster*, *M. brevicollis*, and *C. owczarzaki* were obtained from <http://www.ncbi.nlm.nih.gov/RefSeq/>, <http://flybase.org/>, <http://genome.jgi-psf.org/Monbr1/>, and http://www.broadinstitute.org/annotation/genome/multicellularity_project/, respectively. The predicted TK-encoding sequences of *M. brevicollis* and *A. queenslandica* were obtained from <http://kinase.com/>. The genome sequence of *A. queenslandica* was provided by B. Degan (University of Queensland). The predicted proteomes were searched for TKs by the BLAST and HMMER (48) programs. The highly characteristic motif in the catalytic loop [typically HRDLAARN/HRDLRAAN (24)] was also used to find TKs. Note that none of the pub-

lished gene predictions automatically performed on whole genome sequences is perfect even though they are covered by theoretically sufficient raw reads, and it is still important to analyze directly on the DNA sequences and the raw reads even if official gene predictions are available. We therefore manually revised all the published prediction of TK-encoding sequences of *A. queenslandica*, *M. brevicollis*, and *C. owczarzaki* by searching the genomic DNA sequences with two ab initio gene prediction programs, GENSCAN (49) and Augustus (50). For the Augustus prediction, we optimized the model parameters specifically for each organism. Their genomic DNA assemblies were also revised and reconstructed from the raw data by the use of the wgs-assembler (<http://sourceforge.net/apps/mediawiki/wgs-assembler/>) and phrap (<http://www.phrap.org>) programs when necessary. We then checked all of the predictions manually by comparing with closely related sequences and by inspecting the independent predictions, including suboptimal ones, by GENSCAN and Augustus, paying special attention to obtaining complete sequences. Such reannotations were indispensable both to the classification of TKs based on their protein domain organization and to the reliable phylogeny by KD sequence alignment. The integrity of each protein domain and the presence of the signal peptide in the N terminus of RTK are good indicators of the correct prediction. Comparison of domain architectures of closely related proteins also improved the gene prediction and even the genome sequence assembly. The word “mod” or “new” is added to the name of a *M. brevicollis* gene if the previous prediction (24) is modified, or if the gene is newly discovered in the genome. We totally renamed the *A. queenslandica* genes that were originally predicted in the whole genome sequence (25) because most of them have been radically modified in this study. Not all *A. queenslandica* and *M. brevicollis* TKs were included for the analyses, but they were selected such that the representative sequences cover all the diversity of TK families. We also searched the whole genome sequences of five preopisthokonts, *E. histolytica*, *T. trahens*, *A. castellanii*, *C. reinhardtii*, and *P. infestans*, which were retrieved from <http://pathema.jcvi.org/cgi-bin/Entamoeba/PathemaHomePage.cgi>, ftp://ftp.ncbi.nlm.nih.gov/pub/TraceDB/amastigomonas_sp_atcc_50062, http://www.hgsc.bcm.tmc.edu/microbial-detail.xsp?project_id=163, <http://genome.jgi-psf.org/chlamy/>, and http://www.broadinstitute.org/annotation/genome/phytophthora_infestans/, respectively. Here, we completed, if possible, only the KD sequences, because the quality of assembly was not always satisfactory for obtaining the full-length sequences. We also went back to the raw data and assembled them if necessary. The whole genome sequences of two basal fungal species *A. macrosporus* and *S. punctatus* were retrieved from http://www.broadinstitute.org/annotation/genome/multicellularity_project. All of the sequences (re)annotated in this study are found in the Supplementary Materials.

Protein domain architecture

Protein domain organization was analyzed with HMMER software by searching the SMART (<http://smart.embl-heidelberg.de>) and Pfam (<http://pfam.sanger.ac.uk/>) databases. We confirmed the presence of these predicted domains by manually inspecting the alignments by HMMER. Cysteine-rich sequences that were not mapped to any known domain were classified by the number of the cysteines in a repeating unit, which was identified by aligning the repeats. The TM and signal peptides of RTKs were predicted by the TMHMM and SignalP programs (<http://www.cbs.dtu.dk/services/>), respectively (51).

Classification of TKs

We defined “cognate” TKs as those that had identical domain organization or differed by the loss of one or more copies of a single domain. Repeats of a domain were counted as one (for example, proteins with five and seven Ig-like repeats are cognate, as in the CCK4 family; Fig. 4 and

fig. S7). Noncognate TKs include those that differ by more than one deletion or by one or more domain insertions. Insertions were distinguished from deletions by phylogenetic analysis of their timing. Domain insertions that occurred specifically in bilaterian lineages after the separation from nonbilaterians (or protists) were not taken into account, because such domains might have been defined only by alignments of bilaterian sequences, which are generally much better investigated than others. We classified TKs into families according to both their domain organization and, with some exceptions (discussed later), their phylogenetic relationship on the basis of the KD sequence similarity. Cognate TKs showing a close phylogenetic relationship were classified into a single family. Those that show a cognate domain organization but do not take the neighboring position in the phylogenetic tree were thus classified into distinct families because they were considered to be products of evolutionary convergence. However, cognate TKs of filasterean, choanoflagellate, or sponge lineages that were not mapped to any common metazoan family (thus, they are likely to have diversified specifically in each lineage) were classified into a single family even if they do not show the closest phylogenetic relationship to each other. As seen in Fig. 5 and figs. S8 and S9, frequent domain swapping and conversion have obscured the orthology of such TK genes during their diversification process. We therefore classified them only on the basis of the domain organizations to avoid overestimation of clade-specific expansion of the TK families of these lineages. By the same reasoning, we also classified TKs from a single species having no known domain but KD (or KD, TM, and signal peptide in the case of RTKs) into a single family, which can be subdivided into multiple families when new protein domains or motifs are described in the future.

Inference of phylogenetic trees

Phylogenetic trees were inferred on the basis of the comparison of KD sequences by the ML method with the four categories WAG- Γ model by the use of RAxML v7.0.4 software (52). For large trees with more than 90 operational taxonomic units, we refined the initial near-ML topology found by the RAxML program with the same program in a genetic algorithm (GA)-based heuristic approach, similar to a method previously described (53). The difference of our approach is that, after the normal GA-based optimization, we applied a final tuning of the discovered near-ML topology by the tree bisection and reconnection (TBR) and crossover (tree recombination) algorithms (53); we explored all possible topologies generated by TBR from the GA-optimized topology and all of the possible topologies generated by crossing-over all topologies of the GA population in every possible combination, and confirmed that there was no more improvement. With this approach, we could effectively avoid the local maxima of likelihood for large trees (53). The calculation was performed with a computer cluster composed of six iMacs (Apple) with two cores, a MacPro (Apple) with eight cores, and two PCs (Dell) with four cores. The bootstrap replicates were also performed by RAxML. One *C. owczarzaki* TK (CoPTK-243) was not included because its KD is highly divergent. Phylogenetic inference of the holozoan TK position in the protein kinase superfamily was performed also by the Bayesian method with MrBayes v3.1.2 (54) with the same model. We ran the program for 10,000,000 generations (average SD of split frequencies < 0.013). The confidence of clustering is represented by Bayesian posterior probability. The alignments are shown in figs. S15 and S16.

Statistical test on the group-level clustering

ML bootstrap values for the trees including TKs and STKs (fig. S3A) are relatively low, mostly as a result of the use of divergent sequences and the low number (173) of alignment sites. To test the robustness of higher-level topology (that is, the relationship between group A TKs, group B TKs,

and members of the STK group), we developed a new method, which assesses the overall group-level diversification in each tree of bootstrap replicates. We first inferred 2000 bootstrap-replicated trees, each optimized by the ML method (one example is shown in fig. S6B). We then determined for each tree two putative boundary branches that most likely separate group A and B TK clusters from STKs and group A TK clusters from group B TKs, allowing some violations of included members. We tracked the tree branches by the minimum path starting from a group A TK-encoding gene (1 in fig. S6A) to an STK-encoding gene (84 in fig. S6A) and sought a branch (STK-TK boundary, blue bars in fig. S6, A and B) that maximized the number of group A and group B TKs below this boundary and the number of STKs on the other side of this boundary. We then sought another branch (group A–group B boundary, red bars in fig. S6, A and B) that maximized the number of group A TKs below this boundary and the number of group B TKs above this boundary but still below the STK-TK boundary. We defined the area below the group A–group B boundary as the red area, and that above the group A–group B boundary, but still below the STK-TK boundary, as the blue area. The numbers of the group A TKs (1 to 16), group B TKs (17 to 37), and TKs that belong to neither group A nor B (38 to 45) were counted in each of the red and blue areas and presented as histograms.

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/5/222/ra35/DC1

- Fig. S1. Protein domain organizations of all of the TKs of *C. owczarzaki* and *M. vibrans*.
 Fig. S2. Phylogenetic position of TKs in the ePK superfamily.
 Fig. S3. ML analysis of the phylogenetic positions of TKs and a test of the hypothesis of the multiple origins of TKs.
 Fig. S4. Bayesian and ML analyses of ePKs without TKs.
 Fig. S5. Statistical tests of the group A position.
 Fig. S6. Statistical test of the group-level clustering of TKs.
 Fig. S7. Phylogenetic tree and domain architectures of holozoan TKs.
 Fig. S8. ML tree and protein domain organizations of cluster I TKs of *C. owczarzaki*.
 Fig. S9. Diversification of *A. queenslandica* RTKs in cluster II.
 Fig. S10. An ML tree excluding biased sequences and rapidly evolving sequences.
 Fig. S11. TK families of six holozoans.
 Fig. S12. Phylogenetic tree of CTKs.
 Fig. S13. Protein domain organizations of TKs from three choanoflagellates and their comparison with the *M. brevicollis* TK repertoire.
 Fig. S14. Comparison of three *C. owczarzaki* receptor proteins with very similar extracellular regions.
 Fig. S15. Alignment of ePKs.
 Fig. S16. Alignment of holozoan TKs.
 References
 Sequence data

REFERENCES AND NOTES

1. J. T. Bonner, The origins of multicellularity. *Integr. Biol.* **1**, 27–36 (1998).
2. N. King, The unicellular ancestry of animal development. *Dev. Cell* **7**, 313–325 (2004).
3. A. Rokas, The molecular origins of multicellular transitions. *Curr. Opin. Genet. Dev.* **18**, 472–478 (2008).
4. K. Shalchian-Tabrizi, M. A. Minge, M. Espelund, R. Orr, T. Ruden, K. S. Jakobsen, T. Cavalier-Smith, Multigene phylogeny of choanozoa and the origin of animals. *PLoS One* **3**, e2098 (2008).
5. I. Ruiz-Trillo, A. J. Roger, G. Burger, M. W. Gray, B. F. Lang, A phylogenomic investigation into the origin of metazoa. *Mol. Biol. Evol.* **25**, 664–672 (2008).
6. A. Seb -Pedr s, A. J. Roger, F. B. Lang, N. King, I. Ruiz-Trillo, Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10142–10147 (2010).
7. N. King, M. J. Westbrook, S. L. Young, A. Kuo, M. Abedin, J. Chapman, S. Fairclough, U. Hellsten, Y. Isogai, I. Letunic, M. Marr, D. Pincus, N. Putnam, A. Rokas, K. J. Wright, R. Zuzow, W. Dirks, M. Good, D. Goodstein, D. Lemons, W. Li, J. B. Lyons, A. Morris, S. Nichols, D. J. Richter, A. Salamov, J. Sequencing, P. Bork, W. A. Lim, G. Manning, W. T. Miller, W. McGinnis, H. Shapiro, R. Tjian, I. V. Grigoriev, D. Rokhsar, The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788 (2008).

8. A. Seb -Pedr s, A. de Mendoza, B. F. Lang, B. M. Degnan, I. Ruiz-Trillo, Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. *Mol. Biol. Evol.* **28**, 1241–1254 (2011).
9. W. J. Fantl, D. E. Johnson, L. T. Williams, Signalling by receptor tyrosine kinases. *Annu. Rev. Biochem.* **62**, 453–481 (1993).
10. P. van der Geer, T. Hunter, R. A. Lindberg, Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* **10**, 251–337 (1994).
11. J. Gerhart, 1998 Warkany lecture: Signaling pathways in development. *Teratology* **60**, 226–239 (1999).
12. H. Suga, K. Kuma, N. Iwabe, N. Nikoh, K. Ono, M. Koyanagi, D. Hoshiyama, T. Miyata, Intermittent divergence of the protein tyrosine kinase family during animal evolution. *FEBS Lett.* **412**, 540–546 (1997).
13. G. Apic, J. Gough, S. A. Teichmann, Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **310**, 311–325 (2001).
14. E. V. Koonin, Y. I. Wolf, G. P. Karev, The structure of the protein universe and genome evolution. *Nature* **420**, 218–223 (2002).
15. S. Wuchty, Scale-free behavior in protein domain networks. *Mol. Biol. Evol.* **18**, 1694–1702 (2001).
16. T. Miyata, H. Suga, Divergence pattern of animal gene families and relationship with the Cambrian explosion. *Bioessays* **23**, 1018–1027 (2001).
17. H. Suga, K. Katoh, T. Miyata, Sponge homologs of vertebrate protein tyrosine kinases and frequent domain shufflings in the early evolution of animals before the parazoan–eumetazoan split. *Gene* **280**, 195–201 (2001).
18. H. Suga, M. Koyanagi, D. Hoshiyama, K. Ono, N. Iwabe, K. Kuma, T. Miyata, Extensive gene duplication in the early evolution of animals before the parazoan–eumetazoan split demonstrated by G proteins and protein tyrosine kinases from sponge and hydra. *J. Mol. Evol.* **48**, 646–653 (1999).
19. H. Suga, G. Sasaki, K. Kuma, H. Nishiyori, N. Hirose, Z. H. Su, N. Iwabe, T. Miyata, Ancient divergence of animal protein tyrosine kinase genes demonstrated by a gene family tree including choanoflagellate genes. *FEBS Lett.* **582**, 815–818 (2008).
20. S. K. Hanks, A. M. Quinn, T. Hunter, The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52 (1988).
21. S. K. Hanks, T. Hunter, The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576–596 (1995).
22. G. Manning, G. D. Plowman, T. Hunter, S. Sudarsanam, Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514–520 (2002).
23. R. A. Lindberg, A. M. Quinn, T. Hunter, Dual-specificity protein kinases: Will any hydroxyl do? *Trends Biochem. Sci.* **17**, 114–119 (1992).
24. G. Manning, S. L. Young, W. T. Miller, Y. Zhai, The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9674–9679 (2008).
25. M. Srivastava, O. Simakov, J. Chapman, B. Fahey, M. E. Gauthier, T. Mitros, G. S. Richards, C. Conaco, M. Dacre, U. Hellsten, C. Larroux, N. H. Putnam, M. Stanke, M. Adamska, A. Darling, S. M. Degnan, T. H. Oakley, D. C. Plachetzki, Y. Zhai, M. Adamski, A. Calcino, S. F. Cummins, D. M. Goodstein, C. Harris, D. J. Jackson, S. P. Leys, S. Shu, B. J. Woodcroft, M. Vervoort, K. S. Kosik, G. Manning, B. M. Degnan, D. S. Rokhsar, The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010).
26. N. King, S. B. Carroll, A receptor tyrosine kinase from choanoflagellates: Molecular insights into early animal evolution. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15032–15037 (2001).
27. Y. Segawa, H. Suga, N. Iwabe, C. Oneyama, T. Akagi, T. Miyata, M. Okada, Functional development of Src tyrosine kinases during evolution from a unicellular ancestor to multicellular animals. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12021–12026 (2006).
28. W. Li, S. L. Young, N. King, W. T. Miller, Signaling properties of a non-metazoan Src kinase and the evolutionary history of Src negative regulation. *J. Biol. Chem.* **283**, 15491–15501 (2008).
29. J. L. Tan, J. A. Spudich, Developmentally regulated protein-tyrosine kinase genes in *Dictyostelium discoideum*. *Mol. Cell. Biol.* **10**, 3578–3583 (1990).
30. S. H. Shiu, W. H. Li, Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. *Mol. Biol. Evol.* **21**, 828–840 (2004).
31. L. A. Hertel, C. J. Bayne, E. S. Loker, The symbiont *Capsaspora owczarzaki*, nov. gen. nov. sp., isolated from three strains of the pulmonate snail *Biomphalaria glabrata* is related to members of the Mesomycetozoa. *Int. J. Parasitol.* **32**, 1183–1191 (2002).
32. J. Paps, I. Ruiz-Trillo, Animals and their unicellular ancestors, in *Encyclopedia of Life Sciences* (John Wiley and Sons, Chichester, UK, 2010), pp. 1–8.
33. T. Cavalier-Smith, E. E. Chao, Phylogeny and evolution of Apusomonadida (Protozoa: Apusozoa): New genera and species. *Protist* **161**, 549–576 (2010).
34. G. Torruella, R. Derelle, J. Paps, B. F. Lang, A. J. Roger, K. Shalchian-Tabrizi, I. Ruiz-Trillo, Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single copy protein domains. *Mol. Biol. Evol.* **29**, 531–544 (2012).
35. T. Y. James, P. M. Letcher, J. E. Longcore, S. E. Mozley-Standridge, D. Porter, M. J. Powell, G. W. Griffith, R. Vilgalys, A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* **98**, 860–871 (2006).
36. I. Ruiz-Trillo, G. Burger, P. W. Holland, N. King, B. F. Lang, A. J. Roger, M. W. Gray, The origins of multicellularity: A multi-taxon genome initiative. *Trends Genet.* **23**, 113–118 (2007).
37. H. Kishino, M. Hasegawa, Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidae. *J. Mol. Evol.* **29**, 170–179 (1989).
38. M. Carr, B. S. Leadbeater, R. Hassan, M. Nelson, S. L. Baldauf, Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16641–16646 (2008).
39. P. R. Bergquist, in *Invertebrate Zoology*, D. T. Anderson, Ed. (Oxford Univ. Press, Victoria, 2001), pp. 10–27.
40. L. Margulis, K. V. Schwartz, *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth* (W. H. Freeman and Company, New York, 1998), pp. 246–249.
41. N. K. Tonks, Protein tyrosine phosphatases: From genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* **7**, 833–846 (2006).
42. H. Suga, D. Hoshiyama, S. Kuraku, K. Katoh, K. Kubokawa, T. Miyata, Protein tyrosine kinase cDNAs from amphioxus, hagfish, and lamprey: Isoform duplications around the divergence of cyclostomes and gnathostomes. *J. Mol. Evol.* **49**, 601–608 (1999).
43. P. Dehal, J. L. Boore, Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
44. N. King, C. T. Hittinger, S. B. Carroll, Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301**, 361–363 (2003).
45. T. L. Simpson, *The Cell Biology of Sponges* (Springer-Verlag, New York, 1984), pp. 255–340.
46. T. Mustelin, P. Burn, Regulation of src family tyrosine kinases in lymphocytes. *Trends Biochem. Sci.* **18**, 215–220 (1993).
47. J. Sambrook, D. W. Russell, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 2001).
48. S. R. Eddy, A new generation of homology search tools based on probabilistic inference. *Genome Inform.* **23**, 205–211 (2009).
49. C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
50. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (suppl. 2), ii215–ii225 (2003).
51. J. D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).
52. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
53. K. Katoh, K. Kuma, T. Miyata, Genetic algorithm-based maximum-likelihood analysis for molecular phylogeny. *J. Mol. Evol.* **53**, 477–484 (2001).
54. F. Ronquist, J. P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
55. A. J. Roger, A. G. Simpson, Evolution: Revisiting the root of the eukaryote tree. *Curr. Biol.* **19**, R165–R167 (2009).
56. T. Hunter, Protein kinase classification. *Methods Enzymol.* **200**, 3–37 (1991).

Acknowledgments: The genome sequences of *C. owczarzaki*, *A. macrogynus*, *S. punctatus*, and *T. trahens* are being sequenced by the Broad Institute, Massachusetts Institute of Technology and Harvard University, under the auspices of the National Human Genome Research Institute. We thank B. Degnan (University of Queensland) for access to the *A. queenslandica* genome sequence database; K. Worley and her colleagues in the Human Genome Sequencing Center of the Baylor College of Medicine for allowing us to analyze the *A. castellanii* genome sequence; and F. Lang (University of Montreal) for access to the unpublished genome data of *T. trahens*, *A. macrogynus*, and *S. punctatus*. **Funding:** H.S. is supported by the Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme. A.d.M. is supported by the grant Formaci n del Personal Investigador from Ministerio de Ciencia e Innovaci n awarded to I.R.-T. K.S.-T. is supported by starting grants from the University of Oslo. This study was financially supported by the European Research Council Starting Grant ERC-2007-StG-206883 (to I.R.-T.), grant BFU2008-02839/BMC from Ministerio de Ciencia e Innovaci n (to I.R.-T.), and NIH grant HG004164 (to G.M.). **Author contributions:** H.S. and A.d.M. performed the experiments; H.S., M.D., G.M., and A.d.M. performed computational work; H.S. and I.R.-T. designed the experiments; H.S., I.R.-T., and G.M. designed computational analyses; and H.S., I.R.-T., G.M., M.D., and K.S.-T. wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Sequences have been deposited in GenBank under accession numbers AB591048 to AB591054.

Submitted 28 November 2011

Accepted 13 April 2012

Final Publication 1 May 2012

10.1126/scisignal.2002733

Citation: H. Suga, M. Dacre, A. de Mendoza, K. Shalchian-Tabrizi, G. Manning, I. Ruiz-Trillo, Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. *Sci. Signal.* **5**, ra35 (2012).