Proceedings of the National Academy of Sciences

Annotated PDFs will now be accepted for proof return. Authors who would like to return their proofs by mail, please print all pages of the proof PDF (use "normal quality"). Note the following directions for correcting and returning your proofs. **Important:** For your convenience, this page contains a shortened version of the content in the proof notification e-mail, and some information is not repeated here. Please read the e-mail letter, which will inform you if your article exceeds our page limit.

Please note: The date a paper appears online in PNAS daily Early Edition is the publication date of record and is posted with the article text online. **All author changes must be made before the paper is published online or will be considered and processed as errata.**

Text

1) Clearly mark all of your changes and answers to author queries in the margins next to the article text in the proofs, either in the annotated PDF or on the hard copy;
2) Review and answer ALL author queries (marked in the margins of the text with AQ: A, etc.) that are listed on the query sheet(s);
3) Proofread tables and equations carefully; and
4) Make sure that any Greek/special characters appear correctly throughout the text.
5) PNAS charges for extensive author alterations on proofs. Six or more author alterations per page will be charged at $4 each. Authors will not be charged to correct printer's errors, copyediting errors, or figure errors made in composition.

Figures

1) Proofs contain low-resolution figures (so proofs can be downloaded and printed quickly). Figure quality will be higher in the printed and online html versions of the journal. Please note any figure quality concerns next to the figure on the proofs.
2) Carefully check fig. numbering, color, text labeling, and cropping; if elements are missing from or moved within a figure, or if your color figure does not appear in color in the PDF, please note this on your proofs and send us a printed copy of the correct figure for comparison;
3) Replacing, deleting, or resizing color figures will cost $150/figure and replacing a black-and-white figure will cost $25/figure.

Supporting Information

If you submitted supporting information (SI) to be posted on the PNAS web site, you will receive a PDF proof of the SI in a separate e-mail. PNAS charges for extensive author corrections in SI. Four or more author alterations per SI page will be charged at $4 each. Replacing SI figs. will cost $25/fig.

Within 2 business days, please either return your annotated proof PDF electronically to the editorial contact listed in query A of your proof, or express mail (by overnight or 2-day delivery, if possible) the following items to the address given below.

1) The original printed copy of the PDF, including query sheet(s), with your corrections marked in the margins next to the article text;
2) The reprint order form (including the price sheet). You can fax this form to the number listed on it instead of mailing the form back to us. If you have trouble printing the form, it is also available on our web site at http://www.pnas.org/misc/author_reprints.pdf.

Please retain a copy of ALL pages of the proof PDF for your records. Please include your manuscript number with all correspondence.

Thank you.
Return address for proofs:
Attn: PNAS
8621 Robert Fulton Dr., Suite 100
Columbia, MD 21046
pnas@cadmus.com
Tel: 410-850-0500 (Use this number for shipping purposes only; see query A on the Author Queries sheet in your proof PDF for contact information for your article.)

The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan

Gerard Manning*†, Susan L. Young‡, W. Todd Miller§, and Yufeng Zhai*

*Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037; ‡Department of Molecular and Cell Biology and Center for Integrative Genomics, University of California, Berkeley, CA 94720; and §Department of Physiology and Biophysics, Stony Brook University, Stony Brook, NY 11794

Edited by Tony Hunter, Salk Institute for Biological Studies, La Jolla, CA, and approved April 28, 2008 (received for review February 11, 2008)

Tyrosine kinase signaling has long been considered a hallmark of intercellular communication, unique to multicellular animals. Our genomic analysis of the unicellular choanoflagellate *Monosiga brevicollis* discovers a remarkable count of 128 tyrosine kinases, 38 tyrosine phosphatases, and 123 phosphotyrosine (pTyr)-binding SH2 proteins, all higher counts than seen in any metazoan. This elaborate signaling network shows little orthology to metazoan counterparts yet displays many innovations reminiscent of metazoans. These include extracellular domains structurally related to those of metazoan receptor kinases, alternative methods for membrane anchoring and phosphotyrosine interaction in cytoplasmic kinases, and domain combinations that link kinases to small GTPase signaling and transcription. These proteins also display a wealth of combinations of known signaling domains. This uniquely divergent and elaborate signaling network illuminates the early evolution of pTyr signaling, explores innovative ways to traverse the cellular signaling circuitry, and shows extensive convergent evolution, highlighting pervasive constraints on pTyr signaling.

choanoflagellate | evolution | genome | kinome | phosphotyrosine

Choanoflagellates such as *Monosiga brevicollis* are unicellular aquatic protists and the closest known relatives of multicellular animals (metazoans). The sequencing of the *Monosiga* genome now provides a key evolutionary node between metazoans and fungi, close to the origin of animal multicellularity (1). The role of the tyrosine-specific group of kinases (TKs) in intercellular signaling and their restriction to metazoans suggested that TKs were key to metazoan evolution (2). Plants and unicellular organisms lack TKs, although they have a small number of dual-specificity kinases and associated tyrosine phosphatases (PTPs) and SH2 phosphotyrosine-binding domains generally not involved in intercellular signaling. The surprising discovery of TKs in choanoflagellates (3–5) showed that invention of these key mediators of intercellular signaling preceded their expansion in metazoans. We show here that choanoflagellates have invested hugely in a largely independent pTyr signaling system, yet many of these genes suggest functional convergence between choanoflagellates and metazoans and new combinations of signaling modules, both of which hint at restricted pathways through the signaling network.

Results

Determination and Classification of *Monosiga* Tyrosine Kinases. Our analysis of the draft *Monosiga* genome predicts 128 TKs within a total kinome of ≈380 protein kinases (http://kinase.com/kinbase). Extensive gene model curation and selected cDNA and genome resequencing allowed us to improve predictions for 102 of these sequences, although several fragments and likely imperfect predictions remain. These constitute the largest known tyrosine kinome and make up over twice the fraction of the proteome than that of any metazoan (6–9), a startling result for a unicellular organism. Sequence analysis of the kinase domain and other regions clusters these TKs into 22 families and 26 singletons (Fig. 1). Their scope is paralleled by their diversity: when compared with metazoan TKs by pairwise and multiple sequence alignment and family profile–profile alignments, the only clearly identifiable specific homologs were of the Src subgroup kinases (Src, Csk, Abl, and Tec).

Receptor TKs (RTKs). Eighty-eight RTKs are predicted, based on predicted signal peptides and transmembrane (TM) regions, known extracellular domains, and paralogy. Most are typical type I TM proteins, but two are multipass (six to nine adjacent predicted TMs), including one encoding an MFS transporter domain [supporting information (SI) Fig. S1]. Seventy-three RTKs belong to 15 families, none of which have obvious metazoan orthologs, although kinase domain profile–profile alignments do show weakly specific similarity between RTKB and RTKC families and the metazoan FGFR and Eph families, respectively. Their domain organization is often similar to that of metazoans, whether due to common origin or convergent evolution (Table S1, Fig. 2). For instance, *Monosiga* lacks the Ig domains found in many metazoan RTKs, but 15 *Monosiga* RTKs have divergent repeats similar to hyalin (HYR) domains, which in turn are predicted to be structurally related to Ig and FN3 domains (10). Similarly, 21 *Monosiga* RTKs have cysteine-rich extracellular repeats and several families of CxxC motifs. These are weakly similar to the TNFR and furin-like domains of some metazoan RTKs. Variant EGF-like domains are also seen (Table S1). Several of these domains are found in other predicted receptor and secreted *Monosiga* proteins. For instance, the *Monosiga*-specific RM1 motif is repeated 8–13 times in three RTKs (Fig. S1) and in 40 other proteins, most of which are predicted to be secreted.

Cytoplasmic TKs (CTKs). Most CTKs are associated with membrane and pTyr binding and, as in metazoans, are likely to transduce signals from activated receptors, although frequently with domain combinations. Twenty-nine of the 40 CTKs fall into eight families, seven of which also contain SH2 or phosphotyrosine binding (PTB) domains (Fig. S2). These include homologs of all four Src subgroup families, based on presence of SH2 and SH3 domains and on kinase domain sequence similarity, which averages 60% identity to their closest metazoan homologs, compared with <50% for any other *Monosiga* kinase. As in metazoans, all but Csk have an activation loop phosphorylation site, and all four Src family kinases have

Author contributions: G.M., S.L.Y., and W.T.M. designed research; G.M., S.L.Y., W.T.M., and Y.Z. performed research; G.M., W.T.M., and Y.Z. contributed new reagents/analytic tools; G.M., S.L.Y., W.T.M., and Y.Z. analyzed data; and G.M., S.L.Y., W.T.M., and Y.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. ●●●).

†To whom correspondence should be addressed. E-mail: manning@salk.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0801314105/DCSupplemental.

Cadmus: 3741

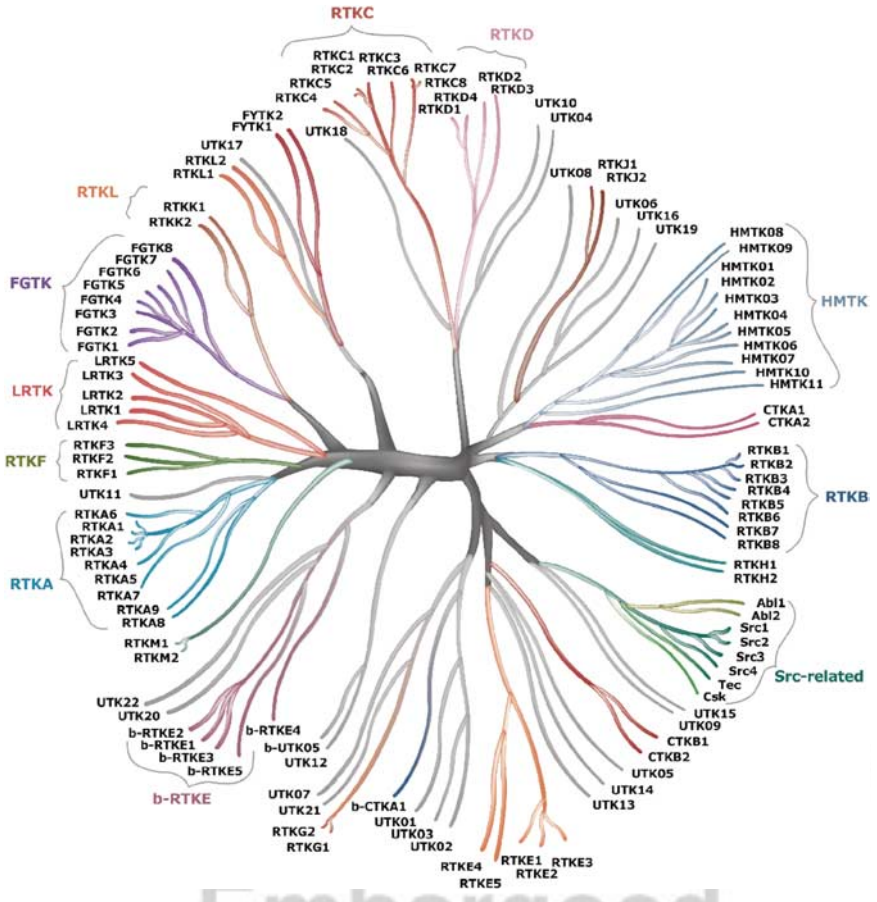


Fig. 1. Phylogenetic tree of *Monosiga* tyrosine kinases, based on alignment of kinase domains, pairwise similarity, and conservation of key residues. Second kinase domains are prefixed by b-. Specific branching patterns between most families are relatively poorly supported.

conserved Csk phosphorylation sites at their C termini. In *Monosiga ovata*, Csk has been shown to phosphorylate and partially repress Src activity through this site (5).

Three of the four Srcs have predicted membrane-anchoring myristoylation sites, indicating they function proximal to receptors, as with their metazoan counterparts. Curiously, the fourth replaces this with a predicted lipid-binding C2 domain that suggests a mechanism of membrane targeting, perhaps similar to the PH domain of Tec kinases (Fig. S2).

Other CTK families have pTyr-binding domains and may be downstream of RTKs. The two FYTK kinases have SH2 and inositol lipid-binding FYVE domains, one CTKA kinase has SH2 and PH domains, and 10 of the 15 HMTK kinases have PTB domains (Fig. S1). Although FYVE and PTB domains have not previously been seen in TKs, they may function analogously to the membrane targeting (PH, myristoylation) and pTyr-binding (SH2) domains of Src subgroup kinases.

Several RTKs contain predicted Src phosphorylation and SH2-binding sites, most notably at four conserved tyrosines in the RM2 motif within the tail of several RTKB kinases (Fig. S1). We tested biochemical activity of *Monosiga* Src1 on peptides generated from two copies of this motif from RTKB2, along with *Monosiga* STAT (a predicted Src substrate) and an optimal vertebrate Src substrate. All showed distinct activity, but the specific activity toward the RTKB2 peptides under these conditions was 6-fold higher (Fig. 3). Kinetic analysis of phosphorylation showed that RTKB2–1 had a k_{cat} of 97.4 min^{-1} and a K_m of 280 μM, whereas the c-Src optimal peptide gave k_{cat} = 6.5 min^{-1} and K_m = 90 μM. Thus, specific recognition of RTKB2–1 by Src1 is driven primarily by a high maximal velocity toward this substrate. These data raise the possibility that the RTKB tail is a Src1 substrate, thus linking RTK and CTK signaling, as in metazoans, and that initial autophosphorylation of one of these sites by the RTK recruits Src for further phosphorylation.

Kinase Domain Conservation and Catalytic Activity. Given their ancient divergence, we tested whether *Monosiga* TK domains had unique sequence features. Comparison of all *Monosiga* TK domains to all human, *Drosophila*, and *Caenorhabditis elegans* TK domains by HMM profiles shows a remarkable similarity (Fig. S3), with no clear difference in the conservation profile at any part of the domain. This suggests that TKs in both lineages are under similar constraints, and that their common ancestor had already taken on a "mature" TK structure. Most appear to be activated by phosphorylation, because 103 TKs conserve one or more tyrosines in the phosphorylable region of the activation loop (Table S2).

In other species, several RTKs have lost key catalytic residues and are thought to act as scaffolds or coreceptors, including the EGF receptor ErbB3 and several Eph receptors (7). By this measure, 13 *Monosiga* TKs are inactive (Table S2). Most belong to the RTKB or RTKM families or are unclassified. Unlike in human, three of the inactive *Monosiga* kinases appear to be cytoplasmic.

Nine kinases have dual catalytic domains, including the six RTKE receptors and the two CTKA cytoplasmic kinases. In all cases, one of the two domains is predicted to be catalytically inactive and is usually very divergent or fragmentary. This situation is analogous to but distinct from metazoan Jak kinases, whose inactive second kinase domains are autoinhibitory (11).

Other Phosphotyrosine Signaling Proteins. The richness and diversity of tyrosine kinases are reflected in downstream pTyr-dependent

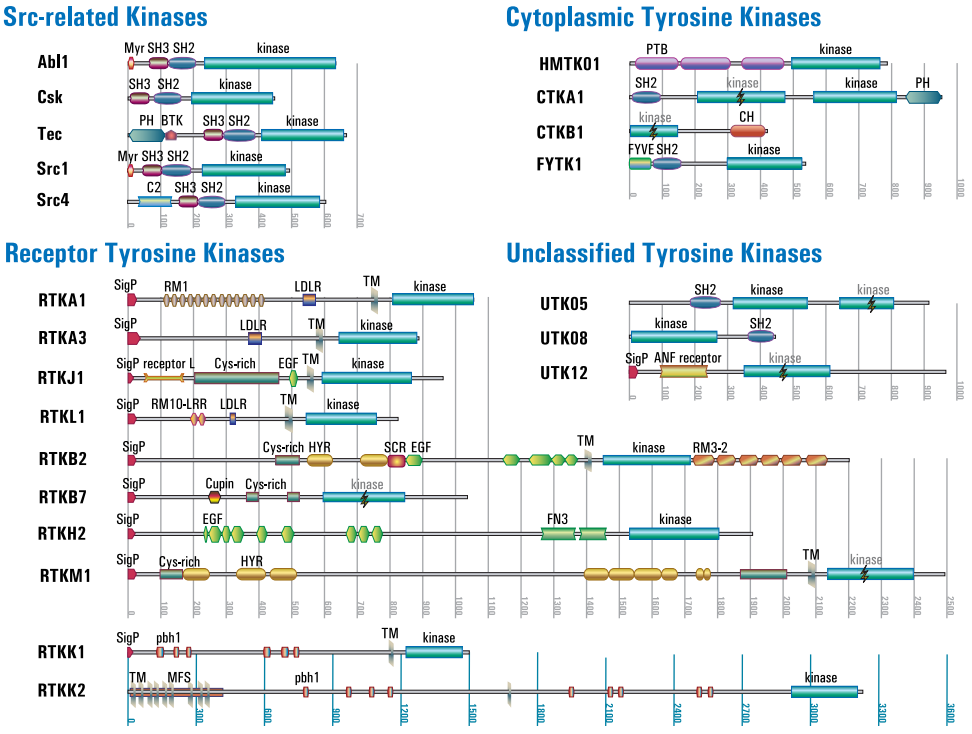


Fig. 2. Domain architecture of representative tyrosine kinases. Predicted inactive kinase domains indicated by lightning bolt, fragments or partial matches to domains indicated by shortened icons. SigP: signal peptide; Myr, myristoylation site; other names from Pfam, SMART, or Table S1. For fuller tyrosine kinome, see Fig. S1 and http://kinase.com/kinbase.

proteins. Conventional tyrosine-specific phosphatases (PTP) and pTyr-binding domains (SH2, PTB) are also greatly expanded in number and domain complexity when compared with yeast, *Dictyostelium*, or *Tetrahymena* and surpass even the human counts for PTP and SH2 proteins (Table 1). As with TKs, we see limited orthology to metazoans, tremendous diversity and several recurrent themes and variations in domain architecture (Fig. 4).

Unlike the few other unicellular PTPs, 4 of the 39 *Monosiga* classical PTPs have clear human orthologs. These include SHP, PTPN13 (PTP-BAS), PTP23 (HD-PTP), and PTP N3/N4. Curiously, *Drosophila* lacks PTPN13, and both *Drosophila* and *C. elegans* lack PTPN23, so, although ancient, these are not evolutionarily indispensable. Both SHP and PTPN13 have been shown to dephosphorylate Src in mammals (12). As in metazoans, some PTPs appear to be catalytically inactive, and four have lost their HCxxxxxR active site motifs (Fig. S2).

By contrast, over one-fifth (26 of 123) of the SH2 proteins have human orthologs covering 15 classes (13) and all 11 major functional categories (Table 2, Fig. 4, Fig. S2). This indicates that much of the cellular pTyr-modulated circuitry was present in the unicellular common ancestor, despite the limited orthology in TKs and PTPs. These shared SH2 proteins mediate pTyr modulation of major signaling pathways, including Ras, Rho, Rac, and Cdc42 small GTPases, phospholipid and calcium signaling, transcription, cytoskeletal interactions, Src subgroup tyrosine kinase, and SHP phosphatase signaling, and several adaptors and scaffolds. The remaining 98 SH2 proteins and 35 PTPs lack metazoan orthologs, but many have domain combinations that suggest common themes and the development of new circuits within a constrained set of signaling interactions. These include 10 receptor PTPs (rPTP), previously unique to metazoans, and 15 cases of a previously

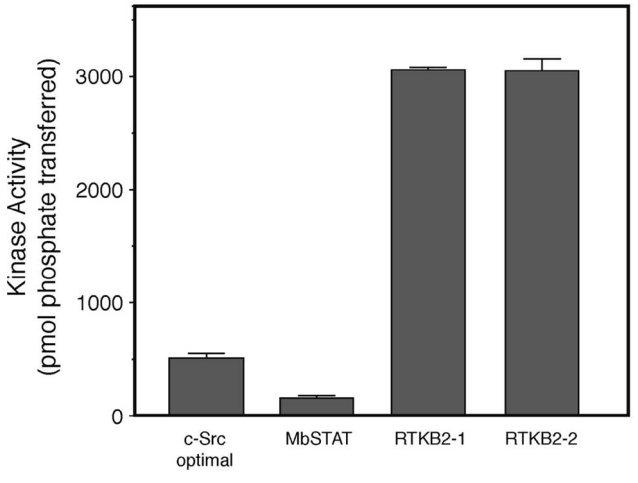


Fig. 3. *Monosiga* Src1 kinase efficiently phosphorylates two RM2 motifs in the cytoplasmic tail of RTKB2 (MbRTK1). The higher efficiency relative to a site on a *Monosiga* STAT homolog or a consensus c-Src substrate suggests that these are specific Src1 phosphorylation sites.

Table 1. Number of proteins with pTyr associated signaling domains in selected genomes

Species	TK	PTP	SH2	PTB
Tetrahymena thermophila	0	3	1	0
Dictyostelium discoideum	0	3	13 (14)	0
S. cerevisiae	0	7	1	0
M. brevicollis	128 (136)	39 (40)	123 (143)	20 (31)
Drosophila melanogaster	33 (34)	16 (23)	28 (34)	10
Human	90 (94)	38 (50)	110 (120)	46 (51)
Human--*Monosiga* orthologs	4	4–5	19	1

Parentheses indicate domain count due to multidomain proteins. Human counts from RefSeq analysis and published studies (13, 16, 28).

Table 2. Human orthologs of *Monosiga* SH2-containing proteins and their functions

Ortholog	Function
Crk	SH2-SH3 adaptor (RAP/RAC GEFs for adhesion)
Grb2	SH2-SH3 adaptor (SOS, Gab1-MAPK/PI3K)
SHP	PTP phosphatase: Src activator, RTK signaling
STAT	Transcription factor
Cbl	Ubiquitination, receptor trafficking
PIK3R (p85)	Phosphoinositide signaling; PI3K regulatory subunit
PLCγ	Phospholipase: PI3K/Ca signaling adaptor
SHIP2	Phosphoinositide phosphatase
RASA1	Small GTPase: Ras adaptor
Rin	Small GTPase: CDC42 adaptor?
Vav	Small GTPase: Rho adaptor
Src/Abl/Csk/Tec	Src kinase signaling
TNS1	Cytoskeleton
SH2D4	Unknown
Supt6 h	Regulator of chromatin structure. Conserved in yeast, probably non-ptyr-binding

ZSF2 undescribed receptor-SH2 (rSH2) combination (Fig. S2). These bring to 103 the count of pTyr-linked receptors. Two rPTPs and three rSH2s are cadherins (2 PTP, 3 SH2), a class best known as metazoan cell adhesion proteins (4, 14). Other extracellular domains include the *Monosiga*-specific variant HYR and cysteine-rich regions also seen in RTKs and several more conventional extracellular domains (FN3, TIG, VWA, TSP, EGF). The single dual-domain rPTP is a possible homolog of the metazoan LAR family, but as with RTKs, the other rPTPs have no clear orthologs.

Several PTP and SH2 domains are fused to Class III myosins. This class was previously found only in combination with the NinaC subfamily of Ser/Thr kinases, which function in both phototransduction and hearing (15). Two PTPs are fused to the kinase–myosin combination, whereas seven SH2 domains are fused to the myosin
ZSF2 but lack the kinase (Fig. 4, Fig. S2).

Many more PTP and SH2 proteins are linked to other signaling domains but in unique architectures or with no specific homology
ZSF2 to human counterparts (Fig. S2). Partner domains consist mostly of protein, lipid, and calcium-binding adaptor modules, including SH2, SH3, PDZ, SAM, WW, C1, PH, ankyrin, and EF hand domains. *Monosiga* lacks orthologs of the metazoan SH2-RasGEF and SH2-C1-RhoGAP proteins but has unique and possibly analogous RasGEF-SH2-SH2 and RhoGAP-SH2 combinations. Similarly, *Monosiga* and metazoans have several nonorthologous proteins containing SH2 and PH domains, which may share related functions, although lacking obvious common ancestry. A few metazoan proteins have dual SH2 domains, but *Monosiga* has 27 multiSH2 proteins, with up to six domains seen in a single protein. Many more proteins (13 PTP, 28 SH2) consist of only one recognizable domain and probably include many fragmentary gene predictions.

PTB domains are prevalent in *Monosiga* and metazoans (Table 1) but are absent from lower organisms. Metazoan PTBs can bind peptides, phosphopeptides, or phospholipids (16). The specific ligands in *Monosiga* could not be predicted by sequence analysis, although domain combinations indicate several are associated with pTyr signaling. The PTB–kinase association in the HMTK family
AQ: H is novel, although reminiscent of the pairing of SH2 with the Src subgroup kinases and a likely case of convergent evolution. *Monosiga* has one PTB-SH2 protein that might be a homolog of the Shc adaptor, but other PTB proteins are unique and have no other domains.

Discussion

The *Monosiga* genome has revealed a treasure trove of diverse tyrosine kinases and associated pTyr signaling proteins. These demonstrate an unprecedented diversity relative to all known (metazoan) TK-based signaling yet reveal several common themes that suggest convergent evolution and a limited set of recurring molecular themes favored by signaling pathways. These data also highlight the unresolved puzzle of why a unicellular organism has such an elaborate signaling system based on external cues. Some choanoflagellates such as *Monosiga ovata* do form colonies, and it may be that such a colonial ancestor drove the evolution of this system, yet it is clear from sequence conservation that pTyr signaling proteins continue to be essential for the current unicellular lifestyle of *M. brevicollis*. Possible functions include response to prey, predators, mates, and the abiotic environment.

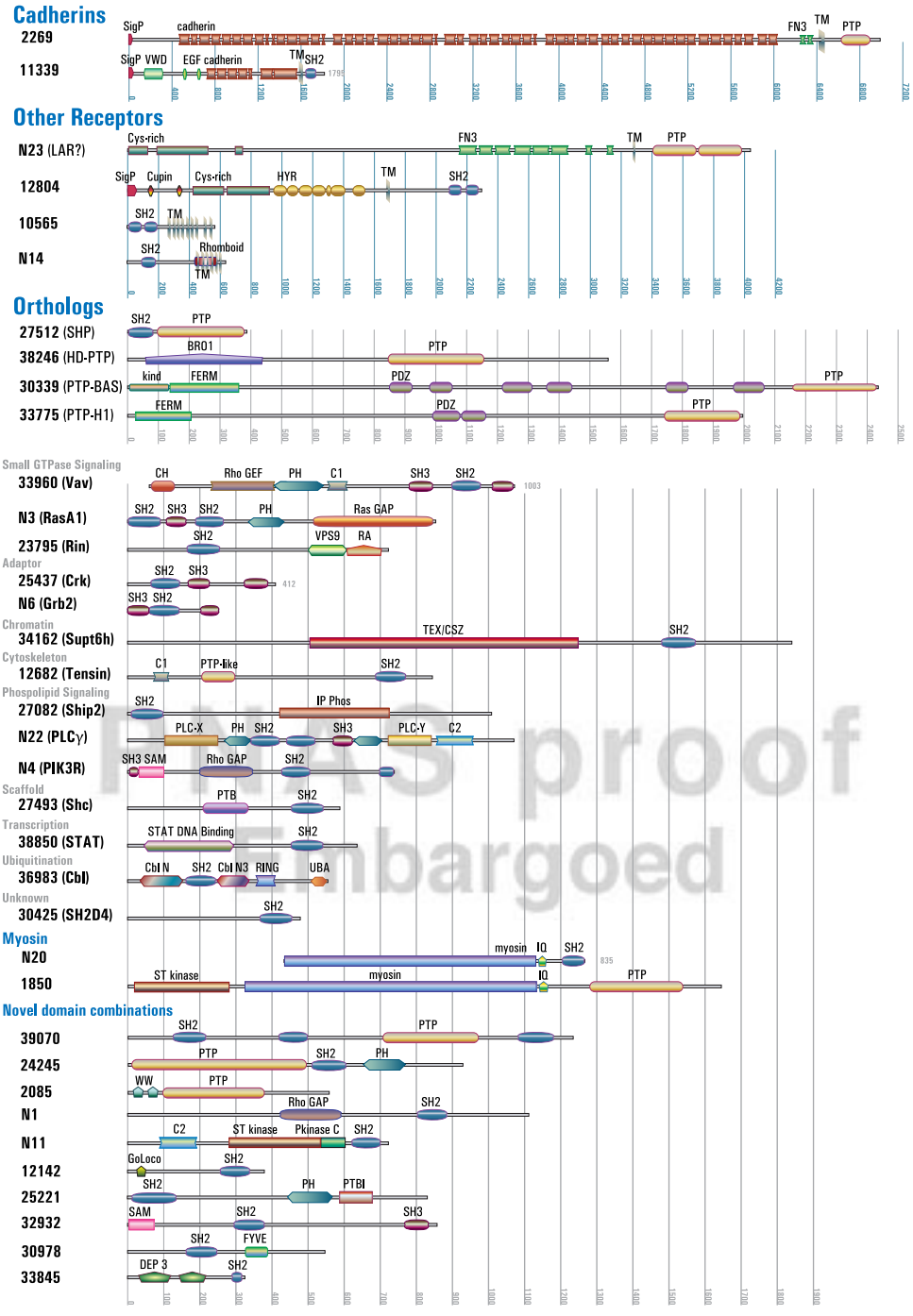
Only the four Src-subgroup kinase families have detectable metazoan orthologs, although possible divergent homologs of FGFR and Eph RTKs may also be present. By contrast, most metazoan TK families are clearly visible in sponges (EGFR, FGFR, Eph, InsR, Ret, Musk, Sev, DDR RTKs and Jak, Syk, and Fer CTKs) (17, 18). This suggests that the choanoflagellate–metazoan common ancestor had a mature Src signaling and some RTKs, but that most metazoan TK families were established closer to the base of metazoans. The story is similar with PTPs, but the common ancestor apparently had a very extensive set of SH2 proteins, with many more classes invented within the choanoflagellate lineage. Extracellular domains evolve rapidly and can swap between families, such as in the RM1 domains found in both RTKA and RTKG1 kinases. Even more remarkably, one, or possibly two, of the receptor SH2 proteins have extracellular domains that are highly similar (>90% sequence identity) to RTKB3 and RTKB5, indicating these are recent fusions and suggesting a kinase–SH2 interaction by receptor heterodimerization. *Monosiga*-specific extracellular motifs are also seen in many other receptor proteins, including rPTPs and rSH2s, and secreted proteins, suggesting they have common ligands or may interact homotypically.

Intracellular domains are more evolutionarily stable and are dominated by pTyr, lipid, and protein interaction domains, as seen in metazoans. In addition, *Monosiga* pTyr-associated proteins have a strong association with the cytoskeleton, as evidenced by an unusual abundance of myosin, CAP_Gly, calponin homology domains, and a variety of small GTPase GAPs and GEFs.

Common themes and possible convergent evolution are seen in the domain structures of many pTyr signaling proteins. These include the swapping of a myristoylation site for a putative lipid-binding C2 domain in Src4, the common occurrence of HYR domains reminiscent of Ig domains, the use of cysteine-rich motifs in extracellular regions, and the use of PTB domains as membrane or phosphopeptide anchors that may be analogous to SH2 and myristoylation domains in Src kinases. The development of dual-domain and catalytically inactive kinases are also probably independent innovations in both lineages. In other cases, *Monosiga* proteins are associated with signaling domains not found in metazoans or found in a different architecture, indicating it has successfully explored new paths within signaling space.

Many of the combinatorial aspects of metazoan pTyr signaling are also found in *Monosiga*, including the widespread occurrence of activation loop phosphorylation sites, the likely phosphorylation of RTKs by Src and of Srcs by Csk, the predicted membrane localization of most CTKs, the conservation of most major classes of SH2 domain proteins, and the occurrence of many multiSH2 proteins that may link distinct pTyr signals. Conversely, the absence of many metazoan components may allow experimental investigation of pathway alternatives, such as the likely specific activation of STAT by Src kinases in the absence of JAKs or the possible link between RTKs and MAPK signaling given the absence of any Raf kinase in *Monosiga* (1).

Future Prospects. This analysis of the draft *Monosiga* genome is surely just an exploratory step in understanding this elaborate and divergent network. The sequence divergence in *Monosiga* and the



ZSF2 **Fig. 4.** Domain organization of selected PTP, PTB and SH2 domain-containing proteins. For fuller details, see Fig. S2 and http://kinase.com/kinbase.

presence of many short exons hamper gene prediction. We manually improved 102 the kinase sequences over the genome predictions, but several are still clearly incomplete. Impending large-scale choanoflagellate EST and genome sequencing, including those for *Proterospongia sp.* and *Monosiga ovata*, will greatly improve our predictions and provide an evolutionary context. New proteomic technologies to identify TK substrate sites and signaling protein interactions (19–21) could quickly fill in much of the signaling network and allow large-scale comparisons to other systems. A greater understanding of pTyr signaling in choanoflagellates promises to reveal both variations on an important biological theme and commonalities that indicate common origin or convergent evolution.

Materials and Methods

Gene Identification. Protein sequences were predicted from release 1.0 of the *Monosiga* genome (1). Protein kinases, PTP, SH2, and PTB-containing proteins were identified by profile HMM searches against genomic, EST, and predicted gene sequences, using HMMer, GeneWise, and Gene Detective (a hardware-accelerated implementation of GeneWise). Individual hits were merged by sequence comparison and mapping to genomic sequence using Blat (22).

AQ: I

TKs were identified by their characteristic HrD[IVLM]Aa**R**N motif [uppercase letters are invariant; Ser/Thr kinases (STKs) are typically HrD[**K**PEN] and by scoring against kinase group-specific HMMs. These TKs also strongly conserved the [KR]**W**m[as]PE motif ([KR]**Y**M[AS]PE in STKs) (Table S2).

ZSF4

All sequences were extensively curated using ESTs, sequence similarity to *Monosiga* and published proteins and to Pfam/SMART domains in surrounding genomic regions. Seven questionable cases were improved by targeted cDNA or

genomic sequencing. Kinase domains were compared with metazoan kinase families by multiple alignment and tree building, by pairwise blast analysis, and by comparison of profile HMMs built from metazoan and *Monosiga* families using PRC (http://supfam.org/PRC).

Domain Profiles. HMM searches on Pfam, SMART, TIGR, and in-house HMMs with Global and Glocal models were performed with a hardware-accelerated DeCypher system (Active Motif). E value cutoffs of e = 10 were used to pick up repeated elements whose individual scores were very low. Sequence-level scores of e >0.001 were discarded and scores of e >1e-8 inspected manually. Overlapping domains from different profile families were merged. Cysteine-rich regions were identified by multiple overlapping hits to the Pfam and SMART profiles GCC2_GCC3, TNFR_3, TNFR_c6, NCD3G, and to internal models for RM5, RM6, RM9, RM15, and RM15t. Adjacent cysteine-rich regions were merged when separated by <10 residues.

Custom HMM profiles were built for several unique conserved regions, found by manual inspection and the MEME motif-finder (23), followed by HMM searches of *Monosiga* and GenBank protein and EST sequences to diversify the motifs found and occasional merging of adjacent motifs into gapped profiles. The Vav PH domain and Supt6 CSZ domain were detected by alignment to proteins with these domains but did not score significantly on the HMMs.

Signal peptides and TM segments were predicted by SignalP (24) and TMHMM (25). TMs that overlapped kinase domains or signal peptides were eliminated. Likely receptors that lacked either signal were subjected to gene reprediction and evaluated in part on the basis of these motifs; this may have lead to some overprediction of such motifs. Myristoylation sites were predicted with NMT (http://mendel.imp.ac.at/myristate) (26).

Kinase Domain Conservation. The alignment of *Monosiga* and metazoan TK domain HMMs was built from a hand-edited alignment of all *Monosiga* TK kinase domains (Fig. S1, SOM File 1) and an alignment of published TK domains of *C. elegans*, *Drosophila*, and human (6, 7, 9). The logo was generated with Logomat-M (27).

ZSF1

Other Genomic Searches. Sequence files used for profile searches included *Dictyostelium*: ''dicty_predicted_proteins'' (http://dictybase.org, June 2007 download) *Saccharomyces cerevisiae*: ''SGD1.01.45.known.pep'' (www.ensembl.org/index.html); *Drosophila* BDGP.4.3.46 ''all.pep'' (www.ensembl.org/index.html); *Tetrahymena thermophila*''gene_prediction'' (http://ciliate.org, May 2005 download), and human RefSeq proteins from GenBank, June 2007 download.

AQ: J

Sequencing. Resequencing used either a *M. brevicollis* cDNA library (3) or cDNA. To generate cDNA, *M. brevicollis* (American Type Culture Collection 50154) was cocultured with *Enterobacter aerogenes* at 25°C in natural seawater infused with cereal grass (5 g/liter) in 150- × 25-mm polystyrene dishes (Falcon). Total RNA was extracted and DNase treated with RNeasy Midi-prep kit (Qiagen). This was reverse-transcribed with an oligo(dT) primer (Invitrogen) and amplified using gene specific primers. PCR amplicons were cloned into pCR-Blunt II-TOPO (Invitrogen) and sequenced by using the vector specific primers M13F (5′-TGTAAAACGACGGCCAGT-3′) and M13R (5′-AACAGCTATGACCATG-3′) or the gene-specific PCR primers.

Src1 Phosphorylation Assay. Src1 was expressed and purified from insect cells (Li *et al.*, personal communication). Phosphorylation assays were carried out in total volumes of 25 μl at 30°C, containing 50 mM Tris·HCl (pH 7.4), 10 mM $MgCl_2$, 1 mg/ml BSA, 400 μM [γ-^{32}P]ATP (15 cpm/pmol), and 750 μM peptide. Peptide sequences were: c-Src optimal, AEEEIYGEFEAKKKG; MbStat, KKKASGYVMADIA; RTKB2–1, SEEVYGAVVDKKK; RTKB2–2, AEEVYEAIADKKK. Reactions were initiated by addition of purified Src1 to 1.5 μM and terminated with 45 μl of cold 10% trichloroacetic acid at 20 min. This time point was within the linear range of the enzyme assay. Samples were centrifuged for 1 min, and 35-μl aliquots of the supernatants were spotted onto 2.1-cm phosphocellulose paper circles (27). The circles were washed three times with cold 0.5% phosphoric acid and once with acetone, dried, and counted dry in a liquid scintillation counter to measure incorporation of ^{32}P into peptide. Reactions were carried out in duplicate and are presented ± standard deviation. For kinetic measurements, reactions were carried out with varying concentrations of peptide substrates (5–1,000 μM). Kinetic parameters were calculated by fitting data to the Michaelis–Menten equation.

AQ: K

ACKNOWLEDGMENTS. We thank Nicole King for advice and for spearheading the *Monosiga* genome project and its use as a model organism and Anne Ashley for superb graphical skills and responsiveness. This work was supported by National Institutes of Health Grants 1 R01 HG004164-01 and P30 CA014195 and by the Razavi Newman Center for Bioinformatics. All data and additional analysis are freely available through KinBase: http://kinase.com/kinbase.

1. King N, *et al.* (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788.
2. Hunter T (2000) Signaling–2000 and beyond. *Cell* 100:113–127.
3. 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.
4. King N, Hittinger CT, Carroll SB (2003) Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301:361–363.
5. Segawa Y, *et al.* (2006) Functional development of Src tyrosine kinases during evolution from a unicellular ancestor to multicellular animals. *Proc Natl Acad Sci USA* 103:12021–12026.
6. Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 27:514–520.
7. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. *Science* 298:1912–1934.
8. Caenepeel S, Charydczak G, Sudarsanam S, Hunter T, Manning G (2004) The mouse kinome: Discovery and comparative genomics of all mouse protein kinases. *Proc Natl Acad Sci USA* 101:11707–11712.
9. Manning, G (2005) *WormBook*, ed *C. elegans* Research Community, available at http://wormbook.org, accessed ●●●.
10. Callebaut I, Gilges D, Vigon I, Mornon JP (2000) HYR, an extracellular module involved in cellular adhesion and related to the immunoglobulin-like fold. *Protein Sci* 9:1382–1390.
11. Saharinen P, Vihinen M, Silvennoinen O (2003) Autoinhibition of Jak2 tyrosine kinase is dependent on specific regions in its pseudokinase domain. *Mol Biol Cell* 14:1448–1459.
12. Roskoski R, Jr (2005) Src kinase regulation by phosphorylation and dephosphorylation. *Biochem Biophys Res Commun* 331:1–14.
13. Liu BA, *et al.* (2006) The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Mol Cell* 22:851–868.
14. Abedin M, King N (2008) The premetazoan ancestry of cadherins. *Science* 319:946–948.
15. Berg JS, Powell BC, Cheney RE (2001) A millennial myosin census. *Mol Biol Cell* 12:780–794.
16. Uhlik MT, *et al.* (2005) Structural and evolutionary division of phosphotyrosine binding (PTB) domains. *J Mol Biol* 345:1–20.
17. Suga H, Katoh K, Miyata T (2001) 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.
18. Nichols SA, Dirks W, Pearse JS, King N (2006) Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci USA* 103:12451–12456.
19. Machida K, *et al.* (2007) High-throughput phosphotyrosine profiling using SH2 domains. *Mol Cell* 26:899–915.
20. Olsen JV, *et al.* (2006) Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127:635–648.
21. Machida K, Mayer BJ, Nollau P (2003) Profiling the global tyrosine phosphorylation state. *Mol Cell Proteomics* 2:215–233.
22. Kent WJ (2002) BLAT–The BLAST-like alignment tool. *Genome Res* 12:656–664.
23. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36.
24. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
25. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580.
26. Maurer-Stroh S, Eisenhaber B, Eisenhaber F (2002) *N*-terminal N-myristoylation of proteins: Prediction of substrate proteins from amino acid sequence. *J Mol Biol* 317:541–557.
27. Schuster-Bockler B, Bateman A (2005) Visualizing profile-profile alignment: pairwise HMM logos. *Bioinformatics* 21:2912–2913.
28. Alonso A, *et al.* (2004) Protein tyrosine phosphatases in the human genome. *Cell* 117:699–711.

AQ: L

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES 1

A—Au: Please contact Betty Cherniak (e-mail: cherniakb@cadmus.com; phone: 410-691-6452) if you have questions about editorial changes, this list of queries, or the figures in your article. Please review the author affiliation and footnote symbols carefully and indicate whether they are correct by writing 'OK' in the margin next to the author line. Please also check the spelling of all author names and affiliations and indicate whether they are correct (write 'OK' in the margin).

B—Au: Titles are limited to three lines in the printed journal. Please shorten your title if it runs over three lines. Please note that nonstandard abbreviations must be spelled out in the title.

C—Au: Journal style discourages statements of priority in research findings ('new,' 'novel,' 'the first,' etc.). 'Previously undescribed, 'previously uncharacterized,' 'recently discovered,' etc., are acceptable substitutes; please reword.

D—Au: If your article contains figures that were submitted in an unsupported or problematic format (e.g., any format other than .eps or .tif), please check the quality of the figures closely. Please note that if figures need to be replaced, the standard charges listed on the reprint form may be incurred.

E—Au: Please spell out MFS.

F—Au: Please verify that all supporting information (SI) citations are correct. A PDF containing the edited and composed SI file can be downloaded from our web site by following the instructions contained within another notification e-mail received with the proofs for the main text. If you never received such an e-mail, please let us know as soon as possible. All print and online-only (SI) materials should be mailed back to us together.

G—Au: Please spell out HMTK.

H—Au: See priority query above; please reword.

I—Au: Please provide supplier names for HMMer, GeneWise, and Gene Detective.

J—Au: Please confirm this web site is correct ('ensembl.org' did not return anything).

K—Au: Please provide all author names, up to 10, for this personal communciation; per journal style, do you have permission for this personal communication? Can this be updated to 'in

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

press‘ and added to the ref. list?

L—Au: Please provide date web site was last accessed.

M—Au: Please review the information in the author contribution footnote carefully. Please make sure that the information is correct and that the correct author initials are listed. Note that the order of author initials matches the order of the author line per journal style. You may add contributions to the list in the footnote; however, funding should not be an author’s only contribution to the work.

N—Au: Please provide GenBank accession nos; please make sure the sequences are released by GenBank or another publicly accessible database before your page proofs are returned. It is PNAS policy that the data be released BEFORE the paper can appear in print.

PNAS
Proceedings of the National Academy of Sciences of the United States of America www.pnas.org

2008 Reprint and Publication Charges

Reprint orders and prepayments must be received no later than 2 weeks after return of your page proofs.

PUBLICATION FEES

Page Charges
(Research Articles Only)

Page charges of $70 per journal page are requested for each page in the article. PNAS charges for extensive author alterations on proofs. Six or more author alterations per page will be charged at $4 each. Authors will not be charged for correcting printer's errors, copyediting errors, or figure errors made in composition.

Articles Published with Figures
(Research Articles Only)

If your article contains color, add $325 for each color figure or table. Replacing, deleting, or resizing color will cost $150 per figure or table. Replacing black-and-white figures will cost $25 per figure. State the exact figure charge on the following page and add to your payment or purchase order accordingly.

Supporting Information
(Research Articles Only)

Supporting information for the web will cost $250 per article.

PNAS Open Access Option

Authors may pay a surcharge of $1200 to make their paper freely available online immediately upon publication. If your institution has a 2008 Site License, the open access surcharge is $850. If you wish to choose this option, please notify the Editorial Office (pnas@nas.edu) immediately, if you have not already done so.

Shipping

UPS ground shipping within the continental U.S. (1–5 days delivery) is included in the reprint prices, except for orders over 1,000 copies. Orders are shipped to authors outside the continental U.S. via expedited delivery service (included in the reprint prices).

Multiple Shipments

You may request that your order be shipped to more than one location. Please add $45 for each additional address.

Delivery

Your order will be shipped within 2 weeks of the journal publication date.

Tax Due

For orders shipped to the following locations, please add the appropriate sales tax:

Canada – 6%; in the U.S.: CA – 7.25% plus the county rate; CT – 6%; DC – 5.75%; FL – 6% sales tax plus local surtax, if you are in a taxing county; MD – 5%; NC – 4.5%; NY – state and local sales taxes apply; VA – 5%; WI – 5%. A copy of the state sales tax exemption certificate must accompany the order form; otherwise, sales tax will be assessed (billed).

Ordering

Prepayment or a signed institutional purchase order is required to process your order. You may use the following page as a Proforma Invoice. Please return your order form, purchase order, and payment to:

PNAS Reprints
PO Box 631694
Baltimore, MD 21263-1694
FEIN 53-0196932

Please contact Robin Wheeler by e-mail at wheelerr@cadmus.com, phone 1-800-407-9190 (toll free) or 1-410-819-3903, or fax 1-410-820-9765 if you have any questions.

Covers are an additional $75 regardless of the reprint quantity ordered. Please see reprint rates and cover image samples below.

Rates for Black-and-White Reprints* (Minimum order 50. Includes shipping.)

Quantity	50	100	200	300	400	500	Add'l 50s over 500
Domestic	$440	$595	$620	$675	$730	$775	$55
Foreign	$475	$630	$685	$775	$850	$920	$75

*** Color covers may be ordered for black-and-white reprints; however, color reprint rates (below) will apply.**

For Black-and-White and Color Reprint Covers add $75

Rates for Color Reprints† (Minimum order 50. Includes shipping.)

Quantity	50	100	200	300	400	500	Add'l 50s over 500
Domestic	$490	$615	$825	$1,110	$1,425	$1,740	$160
Foreign	$550	$665	$885	$1,200	$1,575	$2,015	$215

†Please return your order form promptly.

Covers for black-and-white reprints will display the volume, issue, page numbers, and black-and-white PNAS masthead with the reprint article title and authors imprinted in the center of the page.‡

Covers for color reprints will display the volume, issue, page numbers, and the color PNAS masthead and will include the issue cover image with the reprint article title and authors imprinted in the center of the page.‡

‡Covers for all reprints will be printed on the same paper stock as the article.

PNAS
Proceedings of the National Academy of Sciences of the United States of America www.pnas.org

2008 Reprint Order Form or Proforma Invoice

(Please keep a copy of this document for your records.)

Reprint orders and payments must be received no later than 2 weeks after return of your proofs.

1 Publication Details

Reprint Order Number 3358573

Author's Name

Title of Article

Number of Pages Manuscript Number 08-01314

Are there color figures in the article? ❑ Yes ❑ No

2 Reprint Charges (Use Rates Listed on Previous Page)

Indicate the number of reprints ordered and the total due. Minimum order is 50 copies; prices include shipping.

Research, Special Feature Research, From the Academy, and Colloquium Articles:

_____ Reprints (black/white only) $

_____ Color Reprints (with or without color figures) $

_____ Covers $

For Commentary, Inaugural, Solicited Review, and Solicited Perspective Articles Only:

_____ First 100 Reprints (free; black/white or color)

_____ Covers $

Subtotal $

Sales Tax* $

Total $

*For orders shipped to the following locations, please add the appropriate sales tax: Canada – 6%; in the U.S.: CA – 7.25% plus the county rate; CT – 6%; DC – 5.75%; FL – 6% sales tax plus local surtax, if you are in a taxing county; MD – 5%; NC – 4.5%; NY – state and local sales tax apply; VA – 5%; WI – 5%. A copy of the state sales tax exemption certificate must accompany the order form; otherwise, sales tax will be assessed (billed).

3 Publication Fees (Research Articles Only)

Pages in article @ $70 per page requested $

Color figures or tables in article @ $325 each $

Replacement or deletion of color figures @ $150 each $

Replacement of black/white figures @ $25 each $

Supporting information @ $250 per article $

Open Access option @ $1200 ($850 if your institution has a 2008 Site License/Open Access Membership) per article $

Subtotal $

TOTAL AMOUNT due for reprint and publication fees $

4 Invoice Address It is PNAS policy to issue one invoice per order.

Name

Institution

Department

Address

City State Zip

Country

Phone Fax

Purchase Order Number

5 Shipping Address (if different from Invoice Address)

Name

Institution

Department

Address

City State Zip

Country

Quantity of Reprints

Phone Fax

6 Additional Shipping Address †

Name

Institution

Department

Address

City State Zip

Country

Quantity of Reprints

Phone Fax

†Add $45 for each additional shipping address.

7 Method of Payment

❑ Credit Card ❑ Personal Check (enclosed) ❑ Institutional Purchase Order (enclosed)

8 Credit Card Payment Details

Total Due

❑ Visa ❑ MasterCard ❑ AMEX

Card Number

Exp. Date

Signature

9 Payment Authorization

❑ I assume responsibility for payment of these charges.
(Signature is required. By signing this form, the author agrees to accept responsibility for payment of all charges described in this document.)

Signature of Responsible Author

Phone Fax

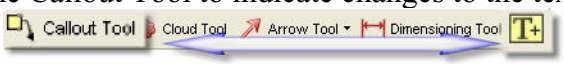
3358573 08-01314

Reprint Order Number Manuscript Number

Send payment and order form to **PNAS Reprints**, PO Box 631694, Baltimore, MD 21263-1694 FEIN 53-0196932
Please e-mail wheelerr@cadmus.com, call 1-800-407-9190 (toll free) or 1-410-819-3903, or fax 1-410-820-9765 if you have any questions.

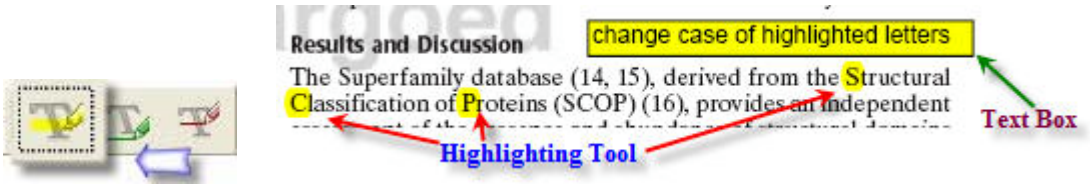
Instructions for Annotating Your .PDF Proof

- Use Text Boxes and the Callout Tool to indicate changes to the text.

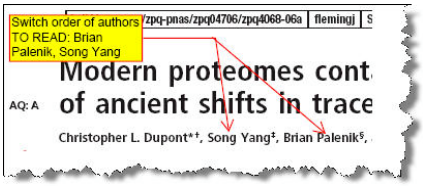


- Use the Strike-Out tool to indicate deletions to the text.

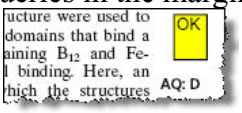
- Use the Highlighting Tool to indicate font problems, bad breaks, and other textual inconsistencies.



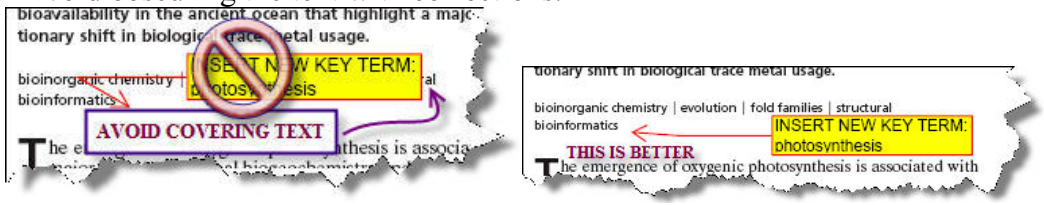
- Clearly indicate where changes need to be made using arrow, lines, and the Call-Out Tool.



- Mark changes and answer queries in the margins and other areas of white space.



- Avoid obscuring the text with corrections.



Proofreader's Marks

MARK	EXPLANATION	EXAMPLE
	TAKE OUT CHARACTER INDICATED	 Your proof.
	LEFT OUT, INSERT	 Your proof.
	INSERT SPACE	 Yourproof.
	TURN INVERTED LETTER	Your proof.
	BROKEN LETTER	Your proof.
	EVEN SPACE	A good proof.
	CLOSE UP: NO SPACE	Your proof.
	TRANSPOSE	A proof good
	WRONG FONT	 Your proof.
	LOWER CASE	Your proof.
	CAPITALS	Your proof.  Your proof.
	ITALIC	Your proof. Your proof.
	ROMAN, NON ITALIC	 Your proof.
	BOLD FACE	Your proof. Your proof.
	LET IT STAND	Your proof.  Your proof.
	DELETE, SEE COPY	She Our proof.
	SPELL OUT	 Queen Eliz.
	START PARAGRAPH	 read. Your
	NO PARAGRAPH: RUN IN	marked. Your proof.
	LOWER	 Your proof.

MARK	EXPLANATION	EXAMPLE
	RAISE	 Your proof.
	MOVE LEFT	Your proof.
	MOVE RIGHT	Your proof.
	ALIGN TYPE	Three dogs. Two horses.
	STRAIGHTEN LINE	Your proof.
	INSERT PERIOD	Your proof
	INSERT COMMA	Your proof
	INSERT COLON	Your proof
	INSERT SEMICOLON	Your proof
	INSERT APOSTROPHE	Your mans proof.
	INSERT QUOTATION MARKS	Marked it proof
	INSERT HYPHEN	A proofmark.
	INSERT EXCLAMATION MARK	Prove it
	INSERT QUESTION MARK	Is it right
	QUERY FOR AUTHOR	Your proof read by
	INSERT BRACKETS	The Smith girl
	INSERT PARENTHESES	Your proof 1
	INSERT 1-EM DASH	Your proof.
	INDENT 1 EM	Your proof
	INDENT 2 EMS	Your proof.
	INDENT 3 EMS	Your proof.