A pre-metazoan origin of the CRK gene family and co-opted signaling network

Yoko Shigeno-Nakazawa1,†, Takuma Kasai2, Sewon Ki3, Elina Kostyanovskaya1, Jana Pawlak1,‡, Junya Yamagishi4,§, Noriaki Okimoto4, Makoto Taiji4, Mariko Okada3, Jody Westbrook5,¶, Yoko Satta6, Takanori Kigawa2 & Akira Imamoto1

CRK and CRKL adapter proteins play essential roles in development and cancer through their SRC homology 2 and 3 (SH2 and SH3) domains. To gain insight into the origin of their shared functions, we have investigated their evolutionary history. We propose a term, crk/crkl ancestral (crka), for orthologs in invertebrates before the divergence of CRK and CRKL in the vertebrate ancestor. We have isolated two orthologs expressed in the choanoflagellate Monosiga brevicollis, a unicellular relative to the metazoans. Consistent with its highly-conserved three-dimensional structure, the SH2 domain of M. brevicollis crka1 can bind to the mammalian CRK/CRKL SH2 binding consensus phospho-YxxP, and to the SRC substrate/focal adhesion protein BCAR1 (p130CAS) in the presence of activated SRC. These results demonstrate an ancient origin of the CRK/CRKL SH2-target recognition specificity. Although BCAR1 orthologs exist only in metazoans as identified by an N-terminal SH3 domain, YxxP motifs, and a C-terminal FAT-like domain, some pre-metazoan transmembrane proteins include several YxxP repeats in their cytosolic region, suggesting that they are remotely related to the BCAR1 substrate domain. Since the tyrosine kinase SRC also has a pre-metazoan origin, co-option of BCAR1-related sequences may have rewired the crka-dependent network to mediate adhesion signals in the metazoan ancestor.

The oncogene v-CRK was originally identified in CT10 avian sarcoma virus. Despite the fact that CRK does not encode a tyrosine kinase, v-CRK can increase phosphotyrosine contents in the cell1. The name CRK was given as an abbreviation for Chicken Tumor 10 Regulator of Kinases. In humans, the CRK gene family has two paralogous members, CRK and CRKL (CRK-Like), on chromosomes 17p13 and 22q11, respectively. CRKL is likely associated with DiGeorge syndrome2,3, whereas it is yet unclear whether CRK has a direct involvement in congenital disorders. Overexpression of CRK or CRKL has been linked to a subset of some cancer types such as ovarian and non-small cell lung cancer1. Both CRK and CRKL are broadly expressed, and are required for normal development in mice2,5. CRK and CRKL share some functions while also having distinct properties5–9. One can hypothesize that while functional and structural differences likely resulted from co-option upon their divergence, their shared functions were inherited from their common ancestral gene during evolution.

Both CRK and CRKL encode adapter proteins consisting of an N-terminal SRC homology 2 (SH2) domain followed by two SH3 domains, SH3n and SH3c1. The CRK gene generates a full-length product (isoform a; also known as CRK-II) and a shorter product (isoform b or CRK-I lacking the SH3c domain similar to v-CRK). CRKL generates only a full-length form in humans. The SH2 domain binds to protein tyrosine kinase (PTK) substrates such as BCAR1 as well as receptor tyrosine kinases (RTK) such as PDGFRA and receptor-associated scaffold proteins such as GAB and DOK1. The SH3n domain binds to proline-rich motifs in C3G (RAPGEF1), DOCK180 (DOCK1), and SOS, guanine nucleotide exchange factors (GEFs) for the small G-proteins RAP1, RAC1, and...
the substrate domain25. Importantly, phosphorylation of multiple tyrosine residues, rather than a single important FAT-like domains are necessary for its localization to focal adhesions as well as for tyrosine phosphorylation of hologs in respectively, in particular in the 5′ end sequences extending into the SH2 domain (Supplementary Figure S1a).

In the next section, we termed these transcripts

Identification of CRK/CRKL Orthologs. A previous analysis of the M. brevicollis genome (release 1.0) predicted that two genetic loci, MONBRDRAFT_25438 and 25437, in the converging orientation (green bars). The yellow bars indicate the coding sequences of the two mRNA transcripts crka1 and crka2 (red bars). The pink bars indicate the predicted SH2 and SH3 domains. The gray bars indicate the exons incorrectly predicted in the provisional models XM_001745882 and XM_001745746. Note that M. brevicollis crka1 mRNA had a splicing pattern different from XM_001745882 at the 5′ end, and that M. brevicollis crka2 mRNA was generated without splicing unlike the predicted model XM_001745746. The pattern of exons and introns in crka1 is supported by the presence of the GT-AG consensus splice site at the boundaries of each intron. Note that M. brevicollis crka2 does not have an intact SH3c domain, since lack of splicing results in a long intervening sequence that divides conserved sequences and also introduces a stop codon (hence, “SH3c?”; see Supplementary Figs S1 and S6 for more information).

RAS, respectively1. These GEFs become activated upon their translocation to the membrane10, and loss of Crk or Crkl impairs activation of corresponding small G-proteins5,6,11,12. The protein tyrosine kinase ABL (and the leukemogenic BCR-ABL) is known to bind the CRK/CRKL SH3n domain1. The tyrosine kinase ABL phosphorylates the CRK/CRKL internal YxxP motif, which inhibits their adapter functions by isolating the SH2 domain11,13. Unlike SH3n, however, SH3c is a non-canonical SH3 domain that does not bind proline-rich targets4. Through these domains, CRK and CRKL can relay PTK-activated signals to downstream signaling mediators, and regulate cell proliferation, migration, and adhesion in response to growth factors as well as cell-matrix interactions4.

Some components of the metazoan PTK network initially evolved in unicellular ancestors. The choanoflagellate Monosiga brevicollis as well as the filasterea Capsaspora owczarzaki may have an elaborate network of PTK-activated signaling, evidenced by an expansive repertoire of tyrosine-specific protein kinases as well as proteins with phosphotyrosine-binding domains such as SH2 and PTB (Phospho-Tyrosine-Binding) domains4–18. Inventions of SH2 or SH3 domains individually, however, appear to predate the emergence of PTKs in these pre-metazoan species17,18.

FAK, SRC, ABL, and their paralogous members play essential roles in integrin-induced signal transduction in mammalian cells9,18–21. Among known PTK families, only the SRC subgroup (TEC, CSK, SRC and ABL) have orthologs in M. brevicollis15. An FAK ortholog was predicted in Capsaspora, but not in choanoflagellates16. Integrins and focal adhesion proteins such as paxillin and vinculin were possibly evolved earlier than these kinases21. Since communications between the cell and extracellular matrix are vital in multicellular organisms, one important question is how intracellular signaling components have become an integral part of cell-matrix adhesions in the signaling network during pre and post-metazoan evolution.

Among numerous PTK substrates in vertebrate cells, BCA1 protein, widely known as p130CAS (CRK-associated substrate), was first discovered as a major γ-CRK and SRC binding protein22. BCA1 is a focal adhesion protein and has been linked to cell migration, survival, and mechanosensing24–27. BCA1 has an N-terminal SH3 domain, followed by the substrate domain, Ser-rich domain, and C-terminal FAT-like domain23,26. Both the SH3 and FAT-like domains are necessary for its localization to focal adhesions as well as for tyrosine phosphorylation of the substrate domain26. Src phosphorylates ten or more tyrosine residues out of the 15 YxxP motifs that exist in the substrate domain26. Importantly, phosphorylation of multiple tyrosine residues, rather than a single important one, is required for cell migration25–28. Perhaps as expected, Bca1 is required for Src-induced cellular transformation29, while FAK and ABL also participate in phosphorylation of Bca113.

Evolution of the signaling network requires co-evolution of interactions after inventions of the primal network components. Our studies demonstrate that the CRK/CRKL ancestral gene evolved before the divergence of the metazoan ancestor, and that the SH2 domain of the CRK/CRKL ancestral protein already had the ability to recognize YxxP motifs, an SH2 binding consensus initially discovered with human CRK30. While the extant pre-metazoan species do not have a Bca1 ortholog, we have identified several transmembrane proteins that include many YxxP repeats similar to the metazoan Bca1 substrate domain. These results suggest a paradigm in which co-option of crk/crkl SH2 binding proteins may have rewired the network connections of highly conserved crk/crkl orthologs in the metazoan ancestor. Since Bca1 is a focal adhesion protein, our findings present an important implication in the evolution of integrin-based adhesion complex at the dawn of the metazoans.
In addition, *M. brevicollis crka2* differs in the region corresponding to the SH3c domain due to lack of splicing (Fig. 1; Supplementary Figure S1b). We have extended our search for orthologous sequences to include several additional species (Supplementary Table S1). To obtain more reliable information than predicted sequences, we have isolated mRNA transcripts or EST plasmids homologous to *CRK/CRKL* from some of these species. Our sequencing results indicate that the predicted sequences posted at NCBI and other public databases frequently include substantial prediction errors. In addition, we have identified a *crk/crkl* ortholog in the genome of the placozoa *Trichoplax adhaerens*. Details are provided in Supplementary Figures S2–S5 as well as in the Methods section.

**Phylogenetic Analysis of CRK and CRKL.** Using 28 nucleotide sequences from 19 pre-metazoan and metazoan species, we have generated a confidence score-based multiple sequence alignment (Supplementary Figure S6). All sequences show high levels of nucleotide identities/similarities in the three main regions that correspond to the one SH2 and two SH3 domains (Fig. 2; Supplementary Figure S6a). Notably, the reading frame in the alignment was perfectly conserved in the SH2 and SH3n domains across all species. While mammalian CRK SH2 domain includes a proline-rich protruded DE loop, we note that there is a conserved exon-exon junction at the end of the DE loop in vertebrate CRK SH2 domains, with the exception of *L. erineacea crk* (Supplementary Fig. S6b). Interestingly, the DE loop is proline-rich only in mammalian CRK SH2 domains. Contrary to CRK, the CRKL SH2 domain is encoded by a single exon in all vertebrates. The SH3n domain is highly homologous among all species, while the SH3c domain is the most divergent of the three SRC homology domains (Supplementary Figure S6c,d). It is noteworthy that most species including *S. rosetta crka*, *M. brevicollis crka1*, and *C. owczarzaki crka* have a highly conserved exon-exon junction near the start of the SH3c domain (Supplementary Figure S6d). As *M. brevicollis crka2* is not spliced and lacks this boundary, it is poorly aligned to the SH3c domain of the other species.

These results indicated that our nucleotide sequence alignments were reliable and offered important information such as conserved exon-exon junctions. We therefore proceeded to build phylogenetic trees from the nucleotide alignment in order to further analyze CRK, CRKL, and their orthologs (Fig. 2; Supplementary Figure S7). Figure 2 shows a best-scored maximum likelihood (ML) tree generated by RAxML along % identities of each SRC homology domains and the presence or absence of an internal YxxP motif (Fig. 2). The tree topology is in general agreement with known evolutionary relationships of the phyla and subphyla. The YxxP motif, a negative regulatory site as a key feature of mammalian Crk, is conserved not only in all vertebrates including *P. marinus*.
Structure of the SH2 Domain of *Monosiga brevicollis* crka1. To gain insight into functional evolution, we have determined the solution structure of the SH2 domain of *M. brevicollis* crka1 (Fig. 3; Supplementary Table S2). This domain consists of two α-helices separated by three anti-parallel β sheets at the middle of the domain (Fig. 3b). Two highly conserved key basic residues (R15 and R33) are located in a pocket on one side of the anti-parallel β sheets (Fig. 3a–c). As they are known to provide essential interactions with the phosphatase group of the ligand protein, it is likely that the crka1 SH2 domain is a functional phosphotyrosine binding domain. Interestingly, another basic residue (K56) faces the phosphotyrosine binding pocket in the *M. brevicollis* crka1 SH2 domain. This lysine residue is conserved exclusively in choanoflagellates (Supplementary Figure S6b). In mammalian CRK and CRKL SH2 domains, a hydrophobic pocket interacts with a proline residue at the +3 position from the phosphotyrosine previously reported to associate with mammalian CRK or CRKL SH2 domains25,34–39 (Fig. 4). We have identified four hydrophobic residues, Y55, I66, Y81, and L86, in a putative proline-binding pocket on the other side of the anti-parallel β sheets conserved in the sequence alignment (Fig. 3a). Overall, the *M. brevicollis* crka1 SH2 domain superimposes closely with those of human CRK32,39 and CRKL9 (Fig. 3d). Our results therefore suggest that the *M. brevicollis* crka1 SH2 domain is capable of interacting with phospho-YxxP similar to its mammalian counterparts.

While the overall domain structure appeared to be highly conserved, we have noted an unusual residue, aspartic acid D67, at the i+1 position of the β turn in the EF loop in the *M. brevicollis* crka1 SH2 domain, where other species have a highly conserved glycine residue (Supplementary Figure S8c). Curiously, the program PROCHECK mapped D67 in a “disallowed” region in a Ramachandran plot unless the residue was manually converted to a glycine (Supplementary Figure S8b). To verify the local structure around the D67 residue, we determined the 3JHNHa coupling constant by HNHA experiment (Supplementary Table S3). The coupling constant is close to the predicted value of type I or II β turns, rather than other standard β turns. However, we could not firmly determine a β turn classification, since the signal intensity is too low to determine the coupling constant of the i+2 residue, T68, and its backbone dihedral angles. Nevertheless, the 3JHNHa analysis indicates that while the *M. brevicollis* crka1 SH2 domain has an unconventional amino acid in the EF loop, the β turn structure itself is conserved.

Binding of Phosphotyrosyl Peptides to *Monosiga brevicollis* crka1 SH2 Domain. The structural analysis above has predicted that the *M. brevicollis* crka1 SH2 domain may recognize the YxxP binding consensus of mammalian CRK and CRKL SH2 domains. To address this issue further, we have used isothermal titration calorimetry (ITC) to determine physical binding of the *M. brevicollis* crka1 SH2 domain with selected 21 phosphotyrosyl peptides previously reported to associate with mammalian CRK or CRKL SH2 domains25,34–39 (Fig. 4). As expected, most peptides tested show specific-binding to either or both CRK and CRKL SH2 domains. Significant two phosphotyrosyl peptides from *M. musculus* Dok7 (Y406) and Bcar1 (p130Cas, Y238) bind the crka1 SH2 domain with Kd values ranging 8–10 μM (Fig. 4; Supplementary Figure S9). These results confirm the prediction from our structural analysis above and suggest that the binding preference of CRK/CRLK SH2 domains is remarkably conserved throughout evolution.

*Monosiga brevicollis* crka1 binds to mammalian Bcar1 and Rapgef1 in a Heterologous System. As the biophysical and biochemical experiments above suggest that Bcar1 may bind to *M. brevicollis* crka1, we constructed a heterologous system to test their association in the presence or absence of activated Src in human embryonic kidney 293 cell line (Fig. 5). Although the amount of association could not be directly compared due to the differential effects of crka1 and CRK on the protein expression of Src and Cas, HA-tagged Bcar1 was co-purified efficiently with *M. brevicollis* crka1 when a constitutively active Src was co-expressed (Fig. 5a). Interestingly, *M. brevicollis* crka1 associated with a broader range of proteins besides HA-tagged Bcar1 as compared to human CRK (Fig. 5b). In addition, we found that *M. brevicollis* crka1 physically associated with the major CRK/CRLK SH3 binding protein RAPGEF1 (C3G) in 293 cells, although the level of association was much
Figure 3. Solution Structure of the *M. brevicollis* crka1 SH2 Domain. (a) The panel shows an amino-acid sequence alignment based on the solution structures of the SH2 domain of *M. brevicollis* crka1 (Mb-crka1), human CRKL and CRK (Hs-CRKL and Hs-CRK). Secondary structure elements of Mb-crka1 are indicated below the alignment. The basic residues surrounding the phosphotyrosine binding site and the hydrophobic residues located at the +3 residue binding pocket are indicated as gray and black shades, respectively. An asterisk indicates the i+1 position of β turn between βE and βF. (b,c) Panels b and c show ribbon and surface models of the mean structure of the Mb-crka1 SH2 domain, respectively. The surface model also shows the electrostatic potential. The basic and acidic residues are shown in blue and red, respectively. The phosphotyrosine binding site (pY) and the +3 residue binding site (+3) are indicated by purple letters. Consistent with previous structural studies of the mammalian CRK SH2 domain, two highly conserved basic residues (R15 and R33) lie in the pY binding pocket and interact with the phosphate group of phosphotyrosine in ligand proteins. In Mb-crka1 SH2, another basic residue (K56) also faces the pY binding pocket. This residue is conserved among the choanozoa crka proteins, but not in the other species (Supplementary Figure S6b). Another pocket lies on the other side of the three anti-parallel beta sheets and appears to be structurally sufficient to accommodate a proline residue identified in mammalian CRK/CRKL SH2 binding motif (the +3 position from a tyrosine residue). (d) A wire model of Mb-crka1 SH2 domain (black) is shown in stereo views, overlaid with human CRK and CRKL SH2 domains based on PDB IDs 2EO3 and 2EYV (red and blue, respectively). While human CRK has an extruded DE loop (see also Panel a), the other parts of the SH2 backbones are well-aligned.
lower than that of human CRK (Fig. 5c). These results therefore demonstrate that M. brevicollis crka1 has biochemical properties similar to that of human CRK, despite their large distance in evolutionary time.

Search for Bcar1 Orthologs. Our results above have hinted that possible co-evolution of Bcar1 may offer insight into the crka SH2-mediated network during the pre-metazoan-metazoan evolution. To address this hypothesis, we first searched bcar1 orthologs in several species. Our tblastn search identified bcar1 orthologs in the hemichordate Saccoglossus kowalevskii and two Cnidarians, Hydra vulgaris and Nematostella vectensis based on high similarities in the N-terminal SH3 domain, Ser-rich domain, and C-terminal FAT-like domain (Table 1; Supplementary Figure S10a). We have also identified a bcar1 ortholog in the sponge Amphimedon queenslandica using similar strategies, although high similarities were confined only to the SH3 and FAT-like domains. Although the placozoa Trichoplax adhaerens showed a high degree of similarity in the SH3 domain, the other domains showed low overall sequence similarities (Table 1; Supplementary Figure S10a and Table S4). Despite relatively low similarity scores, the presence of several YxxP motifs, ranging 6–25 repeats depending on the species, is a highly conserved feature of the substrate domain (15 motifs in Homo sapiens and Mus musculus; 6 in A. queenslandica and T. adhaerens). However, similar BLAST search strategies did not identify a bcar1 ortholog in the choanozoa or C. owczarzaki genome (Table 1). These findings suggest that the major CRK/CRKL binding partner BCAR1 likely evolved during or before the emergence of the common metazoan ancestor, although it may have been lost in the extant pre-metazoan genomes.

Pre-metazoan crka SH2 Binding Partners. The bcar1 substrate domain is not easily identified by BLAST using the whole domain sequence, even in metazoan species. Therefore, we searched RefSeq protein databases instead with the two mouse peptide sequences that showed significant affinities to M. brevicollis crka1 SH2 domain (Table 1). This search yielded BCAR1 as the only protein with “multiple hits” in the human and Mus musculus RefSeq protein databases (Supplementary Table S5). Using this strategy, we have identified three putative proteins that contain multiple YxxP motifs in the choanoflagellate Salpingoeca rosetta (Supplementary Table S5). Upon closer examinations, we have found that PTSG_05573, 12435, and 12436 encode putative transmembrane proteins that include several YxxP motifs in the cytoplasmic region (Fig. 6). These YxxP motifs, such as in PTSG_05573, aligned to the metazoan bcar1 substrate domain (Supplementary Figure S10b). Unlike metazoan bcar1 orthologs, however, these proteins did not have an SH3 domain or other bcar1-like domains (Table 1). To

Figure 4. ITC Reveals Physical Binding of Phospho-YxxP Peptides to Monosiga crka1 SH2 Domain. Values indicate the binding affinity as a dissociation constant (Kd). Kd values are arbitrarily categorized as high affinity (Kd < 3 μM, red), moderate affinity (Kd between 3–20 μM, yellow), or low affinity (Kd > 20 μM, blue). Kd values are not shown in cells shaded gray due to poor curve fitting (PCF). See titration plots in Supplementary Figure S9.
provide evidence that these YxxP motifs may serve as pre-metazoan crka SH2 binding sites, we performed ITC assays with M. brevicollis crka1 SH2 and two phosphopeptide sequences, EM-pYDVP-RS and DM-pYDVP-RN, which overlap 5 out of the 10 YxxP motifs in PTSG_05573. We found that these peptides bound to the M. brevicollis SH2 domain (Table 1). Since biochemical and structural properties are remarkably conserved in the SH2 domain, we designed a heterologous system to co-express RAPGEF1 with GFP-tagged M. brevicollis crka1 or Hs CRK in HEK293 cells as above. The amount of Bcar1 co-precipitation could not be directly compared between M. brevicollis crka1 and Hs CRK, since Hs CRK appeared to reduce protein levels of transfected HA-Mm Bcar1 and Src (reproducible in three experiments; see Western blots of cell lysates for HA and Src). M. brevicollis crka1 binds to multiple proteins phosphorylated by active Src. Phosphotyrosine contents were probed by an anti-phosphotyrosine antibody in the same immunoprecipitates used in panel a. While Bcar1 was the major protein that bound to M. brevicollis crka1 and Hs CRK (lanes 2 and 4), additional tyrosine-phosphorylated proteins of unknown identities (*) also associated with M. brevicollis crka1 (lane 2). We noted that GFP-fusion proteins and GFP itself were also phosphorylated by active Src (lanes 2, 4, and 6). (a) Monosiga brevicollis crka1 can associate with the CRK SH3 binding protein RAPGEF1 (also known as C3G), albeit at a lower level than that of human CRK (lanes 2 and 4). A heterologous system was designed to co-express RAPGEF1 with GFP-tagged M. brevicollis crka1 or Hs CRK in HEK293 cells as above.

Figure 5. Association of Monosiga brevicollis crka1 Protein with Mammalian Bcar1 and Rapgef1. M. brevicollis crka1 associated with an HA-tagged mouse Bcar1 (HA-Mm Bcar1) when an activated mouse Src (Y529F) was co-expressed in a heterologous system (lane 2). Human embryonic kidney 293 cells (HEK293) were transfected with combinations of plasmids as indicated. M. brevicollis crka1 and human CRK were fused to an enhanced GFP at their N-terminus to enable immunoprecipitation (IP) with anti-GFP antibodies. Lanes 1 and 3 were samples from control groups without active Src, showing little co-precipitation of HA-Bcar1. Lane 4 was a positive control with human CRK instead of M. brevicollis crka1. Lanes 5 and 6 were negative controls in which GFP without human CRK or M. brevicollis crka1 SH2 domain was expressed. The amount of Bcar1 co-precipitation could not be directly compared between M. brevicollis crka1 and Hs CRK, since Hs CRK appeared to reduce protein levels of transfected HA-Mm Bcar1 and Src (reproducible in three experiments; see Western blots of cell lysates for HA and Src). M. brevicollis crka1 binds to multiple proteins phosphorylated by active Src. Phosphotyrosine contents were probed by an anti-phosphotyrosine antibody in the same immunoprecipitates used in panel a. While Bcar1 was the major protein that bound to M. brevicollis crka1 and Hs CRK (lanes 2 and 4), additional tyrosine-phosphorylated proteins of unknown identities (*) also associated with M. brevicollis crka1 (lane 2). We noted that GFP-fusion proteins and GFP itself were also phosphorylated by active Src (lanes 2, 4, and 6). (a) Monosiga brevicollis crka1 can associate with the CRK SH3 binding protein RAPGEF1 (also known as C3G), albeit at a lower level than that of human CRK (lanes 2 and 4). A heterologous system was designed to co-express RAPGEF1 with GFP-tagged M. brevicollis crka1 or Hs CRK in HEK293 cells as above.

provide evidence that these YxxP motifs may serve as pre-metazoan crka SH2 binding sites, we performed ITC assays with M. brevicollis crka1 SH2 and two phosphopeptide sequences, EM-pYDVP-RS and DM-pYDVP-RN, which overlap 5 out of the 10 YxxP motifs in PTSG_05573. We found that these peptides bound to the M. brevicollis SH2 domain (Table 1). Since biochemical and structural properties are remarkably conserved in the SH2
domain between human CRK/CRKL and Monosiga crka1, one may speculate that S. rosetta crka likely shares such conserved protein properties, which may permit binding to PTSG_05573. In addition, the multiplicity of YxxP motifs likely boost the protein-protein association, despite a low affinity at each site. BLASTP did not identify proteins containing the two-peptide sequences in Monosiga brevicollis or Capsaspora owczarzaki. BLASTP could also not identify orthologs of these S. rosetta genes in the other pre-metazoan species.

| Species                  | Gene or locus / RefSeq / WGS scaffold | E-values from tblastn search with Hydra bcar1 domains |
|--------------------------|--------------------------------------|------------------------------------------------------|
|                          |                                      | SH3 Domain | Substrate Domain (YxxP repeats) | Serine-rich Domain | FAT-like Domain |
| Homo sapiens             | BCAR1                                 | 7.00E-17   | NS* (15)                      | NS                | 2.00E-16       |
| Mus musculus             | bcr1                                  | 7.00E-16   | NS (15)                       | NS                | 5.00E-17       |
| Sacoglossus kowalevski   | bcar1-like/XM_006811703**/ACQM01029203.1 | 1.00E-14   | NS (13)                       | 3.00E-08          | 1.00E-13       |
| Nematostella vectensis   | bcar1/ABA01009884.1                    | 3.00E-14   | NS (10)                       | 9.00E-22          | 8.00E-20       |
| Amphimedon queenslandica | LOC105313492/XM_011406973/ACUQ01005063.1 (not annotated) | 6.00E-04   | NS (6)                        | NS                | 8.00E-14       |
| Trichoplax adhaerens     | TRIADRAFT_52500/XM_002108193/ABGP01000099.1 (not annotated) | 8.00E-10   | NS (6)                        | NS                | NS            |
| Salpingoea rosetta       | PTSG_05573/XP_004993441§§            | NS         | NS (10)§§                    | NS                | NS            |
|                          | PTSG_12435/XP_004993252§§            | NS         | NS (9)§§                     | NS                | NS            |
|                          | PTSG_12436/XP_004993253§§            | NS         | NS (5)§§                     | NS                | NS            |
| Monosiga brevicollis     | NS                                    | NS         | NS                            | NS                | NS            |
| Capsaspora owczarzaki    | NS***                                 | NS         | NS                            | NS                | NS            |

Table 1. TBLASTN Reveals bcar1 Homologs in Basal Metazoans. *The number in parentheses indicates the number of YxxP repeats. NS: Not Significant, with an arbitrary cut off at an E-value of 0.001. **Our EST plasmid sequences do not agree with the provisional model at the 5’end including the potential translation start. However, the difference does not extend into the SH3 domain. ***Although ACFS020000105.1 is identified by tblastn with Hydra bcar1 SH3 sequence at 6E-04, it is an NCK1 homolog (XM_004348736.1). Thus it is listed as NS in this table. §§Not identified by tblastn with Hydra bcar1. See Supplementary Table S5.

Figure 6. Predicted Pre-metazoan Proteins Include Several YxxP Motifs. The diagrams illustrate two groups of predicted/uncharacterized proteins: 1) three S. rosetta proteins identified by BLASTP with the Bcar1 and Dok7 peptide sequences that bind to the M. brevicollis crka1 SH2 domain (Table 1; Supplementary Table S5 and 2) two M. brevicollis proteins identified by a profile hidden Markov model constructed from the metazoan bcar1 alignment shown in Supplementary Figure S10a without the SH3 domain. Despite the fact that they are identified by two independent methods, all 5 proteins are transmembrane proteins with several YxxP motifs in the predicted cytosolic region. Closed boxes across the membrane indicate predicted transmembrane regions. PTSG_05573 has an extracellular DUF011 domain that has not been assigned to known functions. PTSG_12435 and 12436 have a C2 domain (Ca2+–dependent phospholipid binding domain) after the N-terminal signal peptide. In PTSG_12436, repeats of EGF-like domains likely mediate protein-protein interactions. In M. brevicollis A9V AL3, the extracellular region includes three thrombospondin type 1 repeats (TSP1) that may mediate cell-cell interactions or ligand-receptor binding. In M. brevicollis A9V 449, the extracellular region includes a cysteine-rich growth factor receptor (GFR) domain found frequently in mammalian growth factor receptors such as EGFR, PDGFR and IGF1R, as well as a DUF domain. PTSG_05573, A9V AL3, and A9V 449 have SH2-like domains (SH2SF: SH2 domain superfamily) near the C-terminus. A9V AL3 and A9V 449 are UniProt accession numbers, while PTSG_05573, 12435 and 12436 are current gene symbols/IDs.
To identify sequences distantly related to the metazoan bcar1 in pre-metazoans, we have constructed profile hidden Markov models (profile HMMs) using the bcar1 protein alignment of 10 metazoan species shown in Supplementary Figure S10a. Since the SH3 domain is much more highly conserved than other regions, our preliminary search with an HMM constructed from the whole alignment was overly represented by SH3 domain-containing proteins, none of which included regions similar to the bcar1 substrate domain. Therefore, alternative profile HMMs were constructed from the alignment without the SH3 domain. With this strategy, we identified two proteins, A9VAL3 (XP_001749805, MONBRDRAFT_11892) and A9V449 (XP_001747564, MONBRDRAFT_27039) in the M. brevicollis UniProt database at E-values of 0.00047 and 0.0032, respectively. Although there were a few hits in S. rosetta and C. owczarzaki, they had E-values above 0.05 and contained no YxxP motifs.

The five proteins identified in S. rosetta and M. brevicollis have notable features (Fig. 6). All of them are transmembrane proteins. In addition to several YxxP repeats, three proteins have a cytosolic region with one or two C-terminal SH2-like domains (SH2 domain superfamily). In the extracellular region, A9VAL3 has three thrombospondin type 1 repeats, whereas A9V449 has a cysteine-rich growth factor receptor-like domain and DUF5011 domain (domain of unknown function). PTSG_05573 has a DUF5011 in its extracellular region, but it does not have other domain-like, pattern-identifiable features. Although chained HMMs may introduce a higher risk of false findings, we constructed another profile HMM from the five proteins. The second HMM identified several transmembrane proteins, subsets of which included YxxP repeats in M. brevicollis, S. rosetta, and Capsaspora (Supplementary Table S6). These proteins may represent highly divergent groups of pre-metazoan-specific transmembrane proteins which harbor sequences remotely related to the metazoan bcar1 substrate domain.

**Discussion**

Based on the refined sequence information we obtained, our analysis suggests that the divergence of vertebrate crk and crkl genes may have taken place at or before the divergence of the vertebrate ancestor. While P. marinus contains only a CRKL ortholog, all other vertebrates have two distinct genes identified as CRK and CRKL (Fig. 2). Susumu Ohno originally proposed a revolutionary idea that whole genome duplications (WGDs) had been an important driver responsible for the complexities of metazoan species. The 2R hypothesis assumes that two rounds of WGDs (2R) occurred around the speciation of the vertebrate ancestor. Recent studies of the most primitive vertebrates, the cyclostomes P. marinus and hagfish, suggest that 2R occurred before the divergence of the vertebrates. WGDs generate paralogs called 'ohnologs' evidenced by chromosomal synteny, while many such paralogs undergo sequence degeneration followed by gene loss. Chromosomes 17q13 and 22q11, to which CRK and CRKL are localized in humans, show extensive synteny (Supplementary Table S7). Our phylogenetic analysis is consistent with the hypothesis that the last of the 2R occurred before the divergence of the vertebrate ancestor.

One notable difference between vertebrate CRK and CRKL genes is the fact that the SH2 domain is split over two exons with a junction after the DE loop in CRK orthologs, while it is encoded by a single exon in CRKL orthologs. It is possible that either gain of a new intron in CRK or loss of a previously existed intron in CRKL occurred after the WGD. When we consider these possibilities with highly divergent sequences in the DE loop, it seems more plausible that CRK gained a new intron that may have included alternative splice donor sequences. If a splicing slippage occurred before the divergence of vertebrate species after the WGD, we would anticipate more conserved sequences in the DE loop. Thus, the highly divergent CRK DE loop may be explained by the possibility that splicing slippages occurred independently in multiple vertebrate lineages.

Outside the vertebrates, all primitive species have one or more crka genes. It cannot be determined if these orthologs are more closely related to either CRK or CRKL. We have noted that Ciona intestinalis has two additional model loci related to the crka gene. One is immediately adjacent to crka on chromosome 12 (XM_002130761), and the other is on chromosome 5 (XM_009860485). While these two are also annotated as crk-like at NCBI, our alignment and phylogenetic analysis produce long-branch attraction artifacts if they were included (not shown). In addition, a closely related urochordate, Ciona savignyi, has only one ortholog that corresponds to crka in the Ensembl database (ENSCSAVG00000010659). Hence, we excluded the two additional loci from the current study, and called them crkb and crkc, respectively. It is possible that the crkb and crkc loci accumulated extensive sequence degeneration after they were duplicated in the C. intestinalis lineage. It is noteworthy that C. intestinalis chromosome 12 contains several genes that are orthologous to human chromosomes 17 or 22 and to Xenopus tropicalis genomic scaffolds to which crk and crkl are localized (Supplementary Table S7). It has been proposed that the 2R occurred after the split of C. intestinalis. Therefore, together with the topology of our phylogenetic trees, we speculate that C. intestinalis chromosome 12 retains a block of the common ancestor chromosome from which a WGD may have generated vertebrate crk and crkl.

We have demonstrated that the common ancestry of CRK and CRKL can be traced back to a pre-metazoan origin. As adapter proteins, pre-metazoan crka proteins likely mediate protein–protein interactions through their SRC homology domains in similar ways to their metazoan counterparts (Fig. 7). In basal metazoan and pre-metazoan species, we identified orthologs for a few major CRK/CRKL SH3n-binding proteins, some of which have proline rich motifs with which SH3n may associate (Supplementary Table S8). On the other hand, we have little evidence of orthologs for known CRK/CRKL SH2-binding partners in pre-metazoan species (Supplementary Table S8).

Our current study has offered evidence that the focal adhesion protein BCA1 is likely evolved during or before speciation of metazoans, while previous studies suggested that some focal adhesion proteins such as paxillin (pax) were present in the pre-metazoan species. Paxillin is phosphorylated by FAK, ABL and SRC in response to integrin-mediated cell adhesion and growth factor-mediated signaling in mammalian cells. Like Bca1, paxillin is essential for mouse development. Mammalian paxillin proteins have two YxxP motifs (Y31 and Y118) as primary phosphorylation sites at the N-terminus, which bind to the SH2 domain of CRK and CRKL when phosphorylated. However, conserved YxxP motifs are not found in paxillin orthologs in M. brevicollis.
**Figure 7. Proposed Model of the CRK/CRKL Signaling Network in Metazoans and Pre-metazoans.** In mammalian cells, CRK and CRKL have been assigned to pathways induced by growth factors or extracellular matrix (ECM) proteins. The open arrows indicate the physical association of CRK-CRKL orthologs to their binding partners, while the closed arrows indicate the flow of signals when pathways are activated. Receptor tyrosine kinases (RTK) and intracellular tyrosine kinases such as PDGFRα, MUSK, FAK (PTK2), SRC, and ABL phosphorylate their substrates, which the CRK/CRKL SH2 domain may bind. Many such substrates localize to the plasma membrane as well as to adhesion structures such as focal adhesions in vertebrate cells. Although choanozoa and filasterea have many TKs, integrin-based signaling complexes or known growth factor RTKs are absent. Pre-metazoan CRK/CRKL orthologs likely respond to transmembrane signals that do not exist in metazoan cells. Since SRC and ABL orthologs exist in pre-metazoan species, they may phosphorylate crka SH2 binding proteins. On the other hand, orthologs of major CRK/CRKL SH3n binding proteins such as ABL, DOCK1, and SOS exist in choanoflagellates and Capsaspora (see Supplementary Table S7).

Methods

**Identification of Sequences Closely Related to Mammalian CRK and CRKL.** CRK and/or CRKL homologs have been annotated in RefSeq databases for several species, although some sequences are still listed as “provisional” at NCBI (http://www.ncbi.nlm.nih.gov/; Supplementary Table S1). The annotated sequences available at NCBI were also compared with the Ensembl databases (http://www.ensembl.org/). In these sequences, we identified a significant difference between the Xenopus tropicalis crkl homolog sequences available at NCBI and Ensembl. We chose the sequence from Ensembl as it was more in line with the criteria described below. In the species in which genes are not fully annotated, tblastn was used to identify candidate nucleotide sequences (http://www.ncbi.nlm.nih.gov/
amphioxus.icob.sinica.edu.tw/ for the cephalochordate B. floridae; personal communication with Chris Lowe for the hemichordate S. kowalevskii; http://skatebase.org/ for the little skate L. erinacea; http://cnidarians.bu.edu/stellabase/ for the sea anemone N. vectensis; and http://blast.ncbi.nlm.nih.gov/Blast.cgi for other species). When available, genomic information was also used to confirm RNA sequences. In species in which database information provided no or partial information, we sequenced EST plasmids to obtain full sequence information, or isolated cDNA using RT-PCR from RNA samples obtained from the organism.

Four independent cDNA clones were isolated from a pool of embryonic RNA preparations of the sea lamprey Petromyzon marinus. All four cDNAs are 90 bases longer in the coding region than the annotated/provisional P. marinus crkl sequence available at Ensembl, which would result in additional 30 amino acids between the SH2 and SH3n domains when translated. The missing bases in the annotated crkl sequence can be attributed to a stretch of nucleotide ambiguities that still exists in the corresponding region of in the genomic contig, likely including an additional small exon (Supplementary Figure S2). The four sequences we obtained were independent and unique clones from a single gene, as they differ at synonymous sites or one amino acid insert within the open reading frame.

We have also used full-length EST plasmids to obtain transcripts of Crk/Crkl homologs in the urochordate Ciona intestinalis, which turns out to be identical to the published model sequence XM_002130671 for the coding sequence. We have also sequenced two distinct full-length EST plasmids for the crk/crkl homologs in the cephalochordate Branchiostoma floridae (Supplementary Figure S3). These two transcripts are similar but mapped to non-overlapping genomic scaffolds (Supplementary Table S1; mapping data not shown). Thus, they are transcribed from two independent loci, rather than transcripts produced by alternative splicing or polymorphisms from a single gene. The crk/crkl transcript we obtained from the hemichordate Saccoglossus kowalevskii is different from the sequence annotated as 'crk-like protein like' at NCBI (XM_006822248) by 147 bases corresponding to approximately half the SH2 domain (Supplementary Figure S4). Multiple sequence alignment indicates that our S. kowalevskii ortholog is more similar to the mammalian Crk/Crkl SH2 domain (Supplementary Figure S4).

Hence, the sequence we obtained is used in our phylogenetic analysis.

While we did not find annotated sequences for crk or crkl in the placozoa Trichoplax adhaerens46, we have found a genomic contig (ABGP01000065) that contains crk/crkl homologous sequences by tblastn over its WGS database with the sea anemone Nemotostella vectensis homolog as a query. From this genomic sequence, we have predicted a coding sequence that aligns well with other crk/crkl homologs (Supplementary Figure S5).

Total RNA samples were obtained from Nicole King for the choanoflagellate Monosiga brevicollis, and Marianne Bronner for the sea lamprey Petromyzon marinus. EST plasmids were obtained from Nori Satoh for the tunicate Ciona intestinalis and the amphioxus Branchiostoma floridae, and from Chris Lowe for the hemichordate Saccoglossus kowalevskii. We chose only the sequences that have all three characteristic domains found in mammalian Crk and Crkl: one SH2 followed by two SH3 domains. When multiple transcripts (isoforms) were identified from a single gene, we chose the longest transcript for phylogenetic analysis.

RT-PCR and Sequencing. Reverse transcriptase (Superscript II, Lifetechnologies) was used with oligo dT primers to generate cDNA pools from total RNA isolated from P. marinus or M. brevicollis per manufacturer’s protocol. Crk/Crkl homologs were then amplified by high-fidelity PCR with PrimeSTAR (TAKARA) using specific-primers designed for the Gateway cloning system (Invitrogen). Isolated cDNA fragments as well as full-length EST plasmids were sequenced for both upper and lower strands at the University of Chicago Sequencing Core Facility.

Phylogenetic Analysis. GUIDANCE2 server was used with MAFFT alignment options (\(--\)genpair \(\backslash\backslash\)-maxiterate 1000) to generate iterative nucleotide and protein alignments and to provide a confidence score for each aligned column. Poorly aligned columns below a confidence score of 0.93 were removed before tree building. All sequences scored higher than a default sequence cutoff score of 0.6. For maximum likelihood (ML) analysis, raxmlGUI version 1.5b1 was used to interface a parallel-threaded RAxML51,52. A preset of rapid bootstrap and best tree search was used with the general time reversible (GTR) substitution model + gamma or with a fixed cost matrix of Whelan and Goldman (WAG) + gamma for RAxML nucleotide or protein tree generations, respectively. For Bayesian analysis, an MPI version was compiled from the MrBayes source code (version 3.2.6), and was used to run a Metropolis-coupled Markov chain Monte Carlo (MCMC) analysis for 1,400,000 generations with 1 cold and 3 heated chains. A relative burn-in was set at 25% of total generations. Bayesian nucleotide and protein “consensus” trees were generated using GTR or WAG with 8 gamma categories, respectively. For both RAxML and MrBayes, 8 processor cores were assigned to run parallel computations. When multiple sequence alignments were performed without tree generations for bcarl orthologs and related sequences, MAFFT (v7.271) E-INS-i iterative algorithm was used with a gap extension penalty of 0 (--ep 0), suited to identify conserved motifs in long poorly aligned regions. To search bcarl-related sequences poorly conserved in pre-metazoan species, a Linux version of HMMER v3.1b2 (hmmer.org) was used to construct profile hidden Markov models (HMMs) from the protein alignment shown in Supplementary Figure 10a without SH3 domain or similar alignments of pre-metazoan transmembrane proteins. Searches using profile HMMs were conducted against UniProt protein databases downloaded from www.uniprot.org (SALR5, MONBE, and CAPO3 protein databases for S. rosetta, M. brevicollis, and C. owczarzaki, respectively). Geneious program suite (Biomatters) was used to visualize trees and alignments.

Protein Synthesis. The SH2 domains of human CRK, CRKL, and M. brevicollis crka1 were generated initially as polyhistidine-tagged proteins as described. Recombinant proteins were purified with HisTrap HP columns (GE Healthcare). His-tag was then cleaved from the protein with Tobacco Etch Virus protease and removed.
by the affinity column. After His-tag removal, the synthesized protein included additional seven amino acids (GSSGSGSG) at the N-terminus as a byproduct of the production process.

**NMR and Structure Analysis.** For structure determinations, the SH2 domain of *M. brevicollis* crka1 was synthesized as above, then labeled with $^{13}$C and $^{15}$N using the dialysis mode of the *Escherichia coli* cell-free protein synthesis system. The NMR sample containing 1.1 mM labeled proteins in the NMR buffer (20 mM $^{2H}_{11}$Tris–$\phi$AMBER. For the confirmation of 1.0 mM $^{15}$N labeled proteins. The electrostatic potential was calculated by the program GRASP. Graphic rendering was processed in the program MOLMOL.

**Isothermal Titration Calorimetry (ITC).** Synthesized phosphoryl peptides (95% or greater purity by HPLC) were purchased from Toray Research Center. Based on a pilot experiment with different peptide lengths, we found that 8-mer was the minimal length to obtain reliable ITC and NMR titration results (not shown). Isothermal titration calorimetry (ITC) was performed in MicroCal Auto-iTC200 (Malvern) at 25 °C as described in the manufacturer’s instruction manual. The SH2 domain and peptide were dissolved in 25 mM HEPES, pH7, 100 mM NaCl at a final concentration of 40 μM or 400 μM, respectively. Titration was carried out by injecting the peptide solution into the reaction cell holding the SH2 domain containing solution every 150 seconds. The results were analyzed by the software ORIGIN supplied by the manufacturer. For the confirmation of φ angle, HNHA experiment was performed with an NMR sample containing 1.0 mM $^{15}$N labeled proteins. The electrostatic potential was calculated by the program GRASP. Graphic rendering was processed in the program MOLMOL.

**Heterologous Expression and Detection of Protein Association.** The human Embryonic Kidney 293 cell line (CRL-1573, ATCC) was used to construct a heterologous expression system by plasmid transfection. *M. brevicollis* crka1 was subcloned into a 3′ in-frame cloning site of a CMV-driven humanized GFP vector to generate GFP-Mb crka1. GFP-Hs CRK plasmid was constructed by subcloning the cDNA into pEGFP C1 plasmid (Clontech/TAKARA). HA-Mm BcrA1 tagged with a hemagglutinin (HA)-epitope at the N-terminus and subcloned into a CMV expression plasmid. A constitutively active mouse Src in which the negative regulatory tyrosine Y529 was replaced with a phenylalanine residue, Src(Y529F), was subcloned into a Pol2 promoter-driven expression plasmid. Forty-eight hours after transfection, cell lysates were prepared and analyzed by 7.5–15% gradient SDS PAGE and Western blots, or were first subjected to immunoprecipitation followed by electrophoresis and Western blots. GFP-tagged proteins were precipitated by agarose beads chemically coupled with single chain anti-GFP V3-H (GFP-Trap_A, Chromotek). HA-tag, GFP, Src, and Rapgef1 (C3G) were probed in Western blots using anti-HA (sc-805, Santa Cruz), anti-GFP (sc-8334, Santa Cruz), anti-Src (05–184, EMD Millipore), and anti-C3G (sc-15359, Santa Cruz), respectively.

**Accession Numbers.** The mRNA (cDNA) sequences of *Monosiga brevicollis* crka1 and crka2, *Saccoglossus kowalevskii* crka, *Branchiostoma floridae* crka1 and crka2, and *Petromyzon marinus* crkl have been assigned GenBank accession numbers (KT795024-795329, respectively). In addition to these sequences, Supplementary material FASTA file includes a full-length *Ciona intestinalis* crka (cisma833d10) transcript sequence and the GENSCAN-predicted coding sequence for *Trichoplax adhaerens* crka. The NMR structure data for *Monosiga brevicollis* crka1 SH2 domain have been assigned Protein Data Bank (PDB) ID number 2RVE.

**References**

1. Feller, S. M. Crk family adaptors-signalling complex formation and biological roles. *Oncogene* 20, 6348–6371 (2001).
2. Gurus, D. L., Deusser, G., Pappasoua, V. E. & Imamoto, A. Dose-dependent interaction of Tbx1 and Crkl and locally aberrant RA signaling in a model of del22q11 syndrome. *Dev Cell* 10, 81–92 (2006).
3. Racedo, S. E. et al. Mouse and Human CRKL Is Dosage Sensitive for Cardiac Outflow Tract Formation. *Am J Hum Genet* 96, 235–244 (2015).
4. Birge, R. B., Kalodimos, C., Inagaki, F. & Tanaka, S. Crk and Crkl adaptor proteins: networks for physiological and pathological signaling. *Cell Commun Signal* 7, 13 (2009).
5. Park, T. J. & Curran, T. Crk and Crkl like play essential overlapping roles downstream of disabled-1 in the Reelin pathway. *J Neurosci* 28, 13551–13562 (2008).
6. Li, L., Gurus, D. L., Okura, M. & Imamoto, A. Translocation of Crkl to focal adhesions mediates integrin-induced migration downstream of Src family kinases. *Mol Cell Biol* 23, 2883–2892 (2003).
7. Moon, A. M. et al. CrkI deficiency disrupts Fg8 signaling in a mouse model of 22q11 deletion syndromes. *Dev Cell* 10, 71–80 (2006).
8. Seo, J. H. et al. A Specific Need for CRKL in p210BCR-ABL-Induced Transformation of Mouse Hematopoietic Progenitors. *Cancer Res* 70, 7325–7333 (2010).
9. Jankowski, W. et al. Domain organization differences explain Bcr-Abl’s preference for Crkl over CrkII. *Nat Chem Biol* 8, 590–596 (2012).
10. Hasegawa, H. et al. DOCK180, a major CRK-binding protein, alters cell morphology upon translocation to the cell membrane. *Mol Cell Biol* 16, 1770–1776 (1996).
11. Ballif, B. A. et al. Activation of a Dab1/Crkl/C3G/Rap1 pathway in Reelin-stimulated neurons. *Curr Biol* 14, 606–610 (2004).
12. Chen, K. et al. Interaction between Dab1 and CrkII is promoted by Reelin signaling. *J Cell Sci* 117, 4527–4536 (2004).
13. Colicelli, J. ABL tyrosine kinases: evolution of function, regulation, and specificity. *Sci Signal* 3, re6 (2010).
14. Pincus, D., Letunic, I., Bork, P. & Lim, W. A. Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. *Proc Natl Acad Sci USA* 105, 9680–9684 (2008).
Acknowledgements
The authors are grateful to Nicole King, Chris Lowe, Tatjana Sauka-Spengler, Marianne Bronner, and Nori Satoh for gifts of samples and reagents. We thank Tsuyoshi Momose and Anna Styrczula for critical reading of the manuscript. Low-throughput sequencing was performed at the Sequence Core Facility of the University of Chicago. This study was supported in part by Ben May Innovation Funds to A.I. and research funds from RIKEN to M.T., T.Ki, and M.O.

Author Contributions
A.I. conceived the project and directed experimental designs. A.I., E.K. and T.Kasai. wrote the manuscript text. Y.S.-N., J.P., E.K., J.W., S.K. and T. Kasai. isolated materials and reagents central to the manuscript, and conducted the experiments. J.Y. and N.O. aided computation. T. Kigawa, M.T., M.O. and Y.S. provided guidance in structural, computational and evolutionary analyses.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Shigeno-Nakazawa, Y. et al. A pre-metazoan origin of the CRK gene family and co-opted signaling network. Sci. Rep. 6, 34349; doi: 10.1038/srep34349 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016