Comparison of NF-κB from the protists Capsaspora owczarzaki and Acanthoeca spectabilis reveals extensive evolutionary diversification of this transcription factor

Leah M. Williams, Sainetra Sridhar, Jason Samaroo, Jada Peart, Ebubechi K. Adindu, Anvitha Addanki, Christopher J. DiRusso, BBS22 Molecular Biology Laboratory*, Pablo J. Aguirre Carrión, Nahomie Rodriguez-Sastre, Trevor Siggers & Thomas D. Gilmore

We provide a functional characterization of transcription factor NF-κB in protists and provide information about the evolution and diversification of this biologically important protein. We characterized NF-κB in two protists using phylogenetic, cellular, and biochemical techniques. NF-κB of the holozoan Capsaspora owczarzaki (Co) has an N-terminal DNA-binding domain and a C-terminal Ankyrin repeat (ANK) domain, and its DNA-binding specificity is more similar to metazoan NF-κB proteins than to Rel proteins. Removal of the ANK domain allows Co-NF-κB to enter the nucleus, bind DNA, and activate transcription. However, C-terminal processing of Co-NF-κB is not induced by IκB kinases in human cells. Overexpressed Co-NF-κB localizes to the cytoplasm in Co cells. Co-NF-κB mRNA and DNA-binding levels differ across three Capsaspora life stages. RNA-sequencing and GO analyses identify possible gene targets of Co-NF-κB. Three NF-κB-like proteins from the choanoflagellate Acanthoeca spectabilis (As) contain conserved Rel Homology domain sequences, but lack C-terminal ANK repeats. All three As-NF-κB proteins constitutively enter the nucleus of cells, but differ in their DNA-binding abilities, transcriptional activation activities, and dimerization properties. These results provide a basis for understanding the evolutionary origins of this key transcription factor and could have implications for the origins of regulated immunity in higher taxa.

* A list of authors and their affiliations appears at the end of the paper. Email: gilmore@bu.edu
Transcription factor NF-κB (Nuclear Factor-κB) has been extensively studied for its roles in development and immunity in animals from sponges to humans. More recently, it has been discovered that certain single-celled eukaryotes, namely *Capsaspora owczarzaki* and some choanoflagellates, also contain genes encoding NF-κB-like proteins. To further our understanding of the origins and diversity of NF-κB, we have carried out a functional characterization of this important transcription factor in single-celled protists.

Protists comprise a large group of diverse eukaryotes that are either unicellular or multicellular with poorly differentiated tissue. Two protists that have been studied reasonably well are the monotypic genus *Capsaspora owczarzaki* and the taxonomic class of choanoflagellates, both of which are in the monophyletic Filozoa clade within the Opisthokonta.

*Capsaspora* is a close relative of animals (i.e., basal to sponges) and is classified in the Filastera, which is an independent group within the clade Filozoa that is sister to choanoflagellates and metazoa. The genome of *Capsaspora* encodes many proteins involved in metazoan multicellular processes such as integrins, protein tyrosine kinases, and transcription factors, including NF-κB. *Capsaspora* was originally discovered as an ameba-like symbiont in the hemolymph of the fresh-water snail *Biomphalaria glabrata*.

**Results**

Proteins in the NF-κB superfamily have a conserved N-terminal Rel Homology Domain (RHD), which contains sequences required for dimerization, DNA binding, and nuclear localization. All NF-κB proteins bind to a collection of related DNA sites known as κB sites, and NF-κB proteins bind DNA as either homodimers or heterodimers. In animals from insects to humans, there are two subclasses of NF-κB proteins that differ in amino acid sequence relatedness, C-terminal domain sequences, and DNA-binding site preferences. One subclass includes the traditional NF-κB proteins (p52/p100, p50/p105, Relish) that contain C-terminal inhibitory sequences known as Ankyrin (ANK) repeats, whereas the second class consists of the Rel proteins (RelA, RelB, cRel, Dorsal, Dif) that contain C-terminal transactivation domains. Among early-branching organisms—including cnidarians, poriferaeans, and some protists—only traditional NF-κB-like proteins have been found.

In most multicellular animals, the activity of NF-κB proteins is regulated by subcellular localization, wherein an inactive NF-κB dimer is sequestered in the cytoplasm due to interaction with inhibitory IkB sequences (including the C-terminal ANK repeats of NF-κB proteins). Many upstream signals, including the binding of various ligands to conserved receptors (e.g., Toll-like Receptors (TLRs), Interleukin-1 receptors (IL-1Rs), and tumor necrosis factor receptors (TNFRs)), initiate a signal transduction pathway culminating in activation and nuclear translocation of NF-κB. In the non-canonical pathway, the translocation of NF-κB from the cytoplasm to the nucleus is commenced by the phosphorylation of serine residues C-terminal to the ANK repeats, which leads to removal of the C-terminal ANK repeats by a proteasomal processing that begins at the C terminus and stops within a glycine-rich region (GRR) near the end of the RHD. Removal of the ANK repeat sequence allows NF-κB to translocate to the nucleus, bind DNA, and activate the transcription of target genes for a given biological outcome.

Herein, we have characterized molecular functions of transcription factor NF-κB from two unicellular protists using phylogenetic, cellular, and biochemical techniques. We find that the *Capsaspora* NF-κB protein (Co-NF-κB) requires removal of C-terminal ANK repeats to enter the nucleus, bind DNA, and activate transcription. Furthermore, the multiple NF-κB proteins of a single choanoflagellate (*Acanthoeca spectabilis*) can form heterodimers, suggesting that some choanoflagellates contain subclasses of interacting NF-κB proteins, as are found in vertebrates and flies. Overall, these results provide a functional characterization of NF-κB in a taxon other than Animalia.
these choanoflagellate NF-κBs arose by gene duplications within a given species because the multiple NF-κBs from a given species cluster closely to each other (Supplementary Fig. 1b and ref. 3). Nevertheless, other choanoflagellates have multiple NF-κBs that cluster separately with individual NF-κBs of other choanoflagellates3, suggesting that different classes of NF-κBs exist in certain choanoflagellates.

Unlike most early-branching metazoans, choanoflagellates NF-κB transcripts primarily encode RHD sequences, with no C-terminal GRRs or ANK repeats. However, some choanoflagellate NF-κBs contain extended N termini with homology to sequences that are not normally associated with NF-κBs in vertebrates (Fig. 1, pink bar).

DNA binding, nuclear translocation, and transactivation by Co-NF-κB. To investigate properties of Co-NF-κB in cells, we created pDNA-FLAG vectors for full-length Co-NF-κB and two truncation mutants, one (FLAG-Co-RHD) containing the N-terminal RHD sequences including the NLS and the GRR, and a second (FLAG-Co-Cterm) consisting of the C-terminal ANK repeat sequences and downstream residues (Fig. 2a). As a control, we used the active, naturally truncated sea anemone Nemastostella vectensis (Nv) FLAG-tagged Nv-NF-κB protein that we have characterized previously26 (Fig. 2a). As shown by anti-FLAG Western blotting, each plasmid expressed a protein of the appropriate size when transfected into HEK 293T cells (Fig. 2b, Supplementary Table 1).
For some sponge and cnidarian NF-κB, removal of C-terminal ANK repeat sequences is required for nuclear localization when expressed in vertebrate cells22–24. Therefore, we transfected each of ourFLAG expression plasmids into DF-1 chicken fibroblast cells and performed indirect immunofluorescence using anti-FLAG antiserum (Fig. 2c). Full-lengthCo-NF-κB and Co-Cterm were both located primarily in the cytoplasm of >94% of cells (Supplementary Table 2). In contrast, the Co-RHD and control Nv-NF-κB proteins were both primarily nuclear, as evidenced by co-localization with the Hoechst-stained nuclei. Thus, the removal of the ANK repeats allows Co-NF-κB to enter the nucleus, consistent with what is seen with other metazoan RHD-ANK bipartite NF-κB proteins.

To assess the DNA-binding activity of Co-NF-κB proteins, whole-cell extracts from 293T cells transfected with each FLAG construct were analyzed in an electrophoretic mobility shift assay (EMSA) using a consensus κ-site probe. Extracts containing overexpressed Nv-NF-κB and Co-RHD showed a prominent bound κ-site complex, whereas extracts containing full-lengthCo-NF-κB and Co-Cterm showed essentially no κ-site-binding activity (Fig. 2d).

We also assessed the ability of Co-NF-κB proteins to activate transcription in HEK 293 cells using a κ-site luciferase reporter plasmid (Fig. 2e). Co-RHD and Nv-NF-κB activated transcription well above empty vector control levels (i.e., Co-RHD was ~60-fold above the vector alone control). In contrast, full-length and Cterm Co-NF-κB proteins showed little to no ability to activate transcription. Thus, the ability to activate transcription of a κ-site gene locus appears to be a property of sequences within the N-terminal half of Co-NF-κB.

To assess the inherent ability of Co-NF-κB proteins to activate transcription, we measured their abilities to activate transcription in a lacZ reporter gene assay in yeast using a GAL4-lacZ reporter (Fig. 2f). When fused to the DNA-binding domain of GAL4, the Co-RHD protein activated transcription strongly in yeast, nearly 400-fold above the GAL4 (aa 1–147) alone negative control. The potent transactivating ability of the GAL4-Co-RHD fusion in yeast suggests that transactivation is an intrinsic property of the RHD sequences. The Co-Cterm sequences did not activate transcription in yeast, and the full-lengthCo-NF-κB protein had greatly reduced transactivation ability compared to the Co-RHD domain sequences in yeast (Fig. 2f), indicating that C-terminal sequences can directly block transactivation by the Co-RHD sequences.

For some sponge and cnidarian NF-κB, removal of C-terminal ANK repeat sequences is required for nuclear localization when expressed in vertebrate cells22–24. Therefore, we transfected each of ourFLAG expression plasmids into DF-1 chicken fibroblast cells and performed indirect immunofluorescence using anti-FLAG antiserum (Fig. 2c). Full-lengthCo-NF-κB and Co-Cterm were both located primarily in the cytoplasm of >94% of cells (Supplementary Table 2). In contrast, the Co-RHD and control Nv-NF-κB proteins were both primarily nuclear, as evidenced by co-localization with the Hoechst-stained nuclei. Thus, the removal of the ANK repeats allows Co-NF-κB to enter the nucleus, consistent with what is seen with other metazoan RHD-ANK bipartite NF-κB proteins.

To assess the DNA-binding activity of Co-NF-κB proteins, whole-cell extracts from 293T cells transfected with each FLAG construct were analyzed in an electrophoretic mobility shift assay (EMSA) using a consensus κ-site probe. Extracts containing overexpressed Nv-NF-κB and Co-RHD showed a prominent bound κ-site complex, whereas extracts containing full-lengthCo-NF-κB and Co-Cterm showed essentially no κ-site-binding activity (Fig. 2d).

We also assessed the ability of Co-NF-κB proteins to activate transcription in HEK 293 cells using a κ-site luciferase reporter plasmid (Fig. 2e). Co-RHD and Nv-NF-κB activated transcription well above empty vector control levels (i.e., Co-RHD was ~60-fold above the vector alone control). In contrast, full-length and Cterm Co-NF-κB proteins showed little to no ability to activate transcription. Thus, the ability to activate transcription of a κ-site gene locus appears to be a property of sequences within the N-terminal half of Co-NF-κB.

To assess the inherent ability of Co-NF-κB proteins to activate transcription, we measured their abilities to activate transcription in a lacZ reporter gene assay in yeast using a GAL4-lacZ reporter (Fig. 2f). When fused to the DNA-binding domain of GAL4, the Co-RHD protein activated transcription strongly in yeast, nearly 400-fold above the GAL4 (aa 1–147) alone negative control. The potent transactivating ability of the GAL4-Co-RHD fusion in yeast suggests that transactivation is an intrinsic property of the RHD sequences. The Co-Cterm sequences did not activate transcription in yeast, and the full-lengthCo-NF-κB protein had greatly reduced transactivation ability compared to the Co-RHD domain sequences in yeast (Fig. 2f), indicating that C-terminal sequences can directly block transactivation by the Co-RHD sequences.

For some sponge and cnidarian NF-κB, removal of C-terminal ANK repeat sequences is required for nuclear localization when expressed in vertebrate cells22–24. Therefore, we transfected each of ourFLAG expression plasmids into DF-1 chicken fibroblast cells and performed indirect immunofluorescence using anti-FLAG antiserum (Fig. 2c). Full-lengthCo-NF-κB and Co-Cterm were both located primarily in the cytoplasm of >94% of cells (Supplementary Table 2). In contrast, the Co-RHD and control Nv-NF-κB proteins were both primarily nuclear, as evidenced by co-localization with the Hoechst-stained nuclei. Thus, the removal of the ANK repeats allows Co-NF-κB to enter the nucleus, consistent with what is seen with other metazoan RHD-ANK bipartite NF-κB proteins.

To assess the DNA-binding activity of Co-NF-κB proteins, whole-cell extracts from 293T cells transfected with each FLAG construct were analyzed in an electrophoretic mobility shift assay (EMSA) using a consensus κ-site probe. Extracts containing overexpressed Nv-NF-κB and Co-RHD showed a prominent bound κ-site complex, whereas extracts containing full-lengthCo-NF-κB and Co-Cterm showed essentially no κ-site-binding activity (Fig. 2d).

We also assessed the ability of Co-NF-κB proteins to activate transcription in HEK 293 cells using a κ-site luciferase reporter plasmid (Fig. 2e). Co-RHD and Nv-NF-κB activated transcription well above empty vector control levels (i.e., Co-RHD was ~60-fold above the vector alone control). In contrast, full-length and Cterm Co-NF-κB proteins showed little to no ability to activate transcription. Thus, the ability to activate transcription of a κ-site gene locus appears to be a property of sequences within the N-terminal half of Co-NF-κB.

To assess the inherent ability of Co-NF-κB proteins to activate transcription, we measured their abilities to activate transcription in a lacZ reporter gene assay in yeast using a GAL4-lacZ reporter (Fig. 2f). When fused to the DNA-binding domain of GAL4, the Co-RHD protein activated transcription strongly in yeast, nearly 400-fold above the GAL4 (aa 1–147) alone negative control. The potent transactivating ability of the GAL4-Co-RHD fusion in yeast suggests that transactivation is an intrinsic property of the RHD sequences. The Co-Cterm sequences did not activate transcription in yeast, and the full-lengthCo-NF-κB protein had greatly reduced transactivation ability compared to the Co-RHD domain sequences in yeast (Fig. 2f), indicating that C-terminal sequences can directly block transactivation by the Co-RHD sequences.
κB protein from Aiptasia (Ap) was used, which has been shown to undergo IKK-induced processing in HEK 293T cells. Although IKK induced processing of Ap-NF-κB, none of the IKKs induced processing of Co-NF-κB (Fig. 3a and Supplementary Fig. 3a), beyond the small amount of constitutive processing of Co-NF-κB that occurs even in the absence of IKK (Figs. 2b and 3a). Of note, the lower Co-NF-κB band seen in these Western blots was roughly the same size as the predicted RHD (Fig. 2b), and incubation of transfected cells with the proteasome inhibitor MG132 reduced the appearance of the lower band, suggesting that the lower band arises by proteasomal processing of full-length Co-NF-κB in 293T cells (Supplementary Fig. 3b).

To determine whether Co-NF-κB could be processed by an IKK-dependent mechanism, we created a mutant in which we replaced C-terminal sequences Co-NF-κB (downstream of the ANK repeats) with C-terminal sequences of the sea anemone Aiptasia (Ap)–NF-κB that contain conserved serines which can facilitate IKK-induced processing of Ap-NF-κB24. We termed this mutant Co-NF-κB-Ser, and also created the analogous protein (Co-NF-κB-ALA) in which the relevant serines were replaced by non-phosphorylatable alanines. Co-expression of Co-NF-κB-Ser with constitutively active human IKKβ (IKKβ SS/EE) resulted in increased amounts of the lower band, which was not seen with Co-NF-κB-ALA (Fig. 3b). Thus, the Co-NF-κB protein (consisting of the RHD, GRR, and ANK repeats) can undergo IKK-induced processing if supplied with a C terminus containing IKK target serine residues. However, the native Co-NF-κB protein does not appear to be susceptible to IKK-induced processing, which is consistent with the lack of any sequences encoding an IKK-like protein in the genome of Capsaspora.

Exogenously expressed full-length and truncated versions of Co-NF-κB localize primarily to the cytoplasm in Capsaspora cells. We were next interested in examining the subcellular localization of NF-κB in Capsaspora cells. For these experiments, we transfected Capsaspora cells with our FLAG-tagged Co-NF-κB constructs (Co-NF-κB, mutant Co-RHD, and mutant Co-Cterm, see Fig. 2a) and then performed anti-FLAG indirect immunofluorescence. Consistent with results seen in DF-1 chicken cells (Fig. 2d), FLAG-Co-NF-κB and FLAG-Co-Cterm localized to the cytoplasm of Capsaspora cells (Fig. 4, top and bottom rows). Surprisingly, FLAG-Co-RHD—which is nuclear in DF-1 cells—appeared to be primarily cytoplasmic in Capsaspora cells, as judged by its exclusion from the Hoechst-stained nuclei (Fig. 4, middle row).

Co-NF-κB mRNA levels and DNA-binding activity vary coordinately across different life stages and the identification of putative Co-NF-κB target genes. Capsaspora has been shown to have three different life stages: aggregative, filopodia, and cystic, and RNA-Seq of each life stage has been reported16. We were interested in determining whether Co-NF-κB was active at different levels in these three life stages and whether we could use protein-binding microarray (PBM) data and the previous RNA-Seq data to identify genes whose expression may be controlled by NF-κB.
To determine the overall DNA binding-site specificity of Co-NF-κB, we analyzed a bacterially expressed Co-NF-κB RHD-only protein on a PBM containing 2592 κB-like sites and 1159 random background sequences (for array probe sequences, see ref. 24). By comparison of the z-scores for binding to DNA sites, the PBM-based DNA-binding profile of Co-NF-κB is most similar to κB proteins from the sea anemone N. vectensis and human p52, and it is distinct from human cRel and RelA (Fig. 5a; (Co-NF-κB vs.: Nv-NF-κB, R² = 0.23; Hu-p52, R² = 0.27; Hu-RelA, R² = 0.01; Hu-cRel, R² = 0.08), which is consistent with the sequence and structural data indicating that Co-NF-κB is more like NF-κB proteins than Rel proteins. Based on these PBM data, we...
generated a DNA-binding motif logo for Co-NF-κB (Fig. 5b), which is quite similar to the binding motif of the p50:p50 homodimer\textsuperscript{27}.

By examining previous mRNA expression data\textsuperscript{16} for NF-κB mRNA, we found that NF-κB mRNA was expressed at the lowest level in the aggregative stage and at 2.3- and 5-fold higher levels in the filopodic and cystic stages, respectively (Fig. 5c). We next generated cultures of Capsaspora at each life stage (Fig. 5c), made protein extracts, and performed an EMSA using the κB-site probe that we showed can be bound by Co-NF-κB expressed in HEK 293T cells (see Fig. 2c) and by bacterially expressed Co-RHD in our PBM assays (Supplementary Data 1). Consistent with the mRNA expression data, the κB-site probe was bound progressively stronger in lysates from aggregative, filopodic, and cystic stage cells (Fig. 5d). To confirm that the EMSA band included Co-NF-κB, we incubated our protein extracts with 10- and 25-fold excesses of unlabeled κB-site probe, and we saw a substantial decrease in binding to the putative NF-κB band. In contrast, incubation with a 25-fold excess of unlabeled IRF-site probe did not decrease the putative Co-NF-κB complex, indicating that the binding activity in the extracts was specific for the κB-site probe (Fig. 5d).

We next sought to identify genes that might be influenced by the expression of NF-κB in these life stages. For this analysis, we examined existing RNA-Seq data\textsuperscript{16}, which contains mRNA expression data for 8674 genes of Capsaspora during its three life stages. We first narrowed our gene list to those genes that are differentially expressed in a manner similar to NF-κB mRNA levels and DNA-binding activity during each life stage (i.e., genes that are progressively increased in expression in aggregative, filopodic, and cystic stages). From this exercise, we identified 1348 mRNAs that were expressed at low levels in the aggregative stage, and progressively higher levels in the filopodic and cystic stages (Supplementary Data 2).

Of the 1348 genes that we identified, 389 genes were annotated (which is consistent with approximately 25% of Capsaspora’s total predicted protein-encoding genes being annotated\textsuperscript{4,16}), and 305 of these 389 genes had human homologs (Supplementary Data 2). We then performed GO analysis to identify biological processes overrepresented in these 305 genes. The GO analysis showed that this set of 305 genes was enriched for genes predicted to be involved in several biological processes (Supplementary Fig. 4), including 16 Capsaspora genes that encode proteins associated with developmental and immune system processes (Fig. 5e), which are biological processes regulated by NF-κB in many multicellular organisms and suggested to be regulated by NF-κB in several early-branching organisms\textsuperscript{22–25,28–30}. Although 16 genes may seem low, the total database for human GO analysis of immune system and developmental processes genes is approximately 2200 genes, but the number of annotated homologs that exist in Capsaspora in these two GO categories is only 66 genes (Supplementary Table 3). Thus, about 25% (16/66) of the Capsaspora genes in the GO category for the developmental and immune processes subset are among the 305 homologous genes coordinately expressed with NF-κB. Other broad GO categories overrepresented in these 305 genes included Signaling, Metabolic Process, and Locomotion (Supplementary Fig. 4).

We then looked for κB sites within 500 base pairs (bp) upstream of the transcription start sites (TSS) for each of the 1348 genes with expression profiles that were similar to NF-κB using the Co-NF-κB DNA-binding motif (Fig. 5b) generated from the PBM analysis. The upstream regions of 235 of these 1348 genes contained 1-4 κB sites within 500 bp of the TSS, with the majority of these genes containing 1 κB site (Supplementary Data 3). Two (SRMS and SLK) of the 16 genes that are coordinately regulated with Co-NF-κB and encode protein homologs associated with GO developmental and immune system processes also contain a κB site within the 500 bp upstream of their TSS (Fig. 5e). Of note, one gene (S\textsuperscript{2}-AMP-activated protein kinase catalytic subunit alpha-2) contains four upstream κB sites and six genes have three κB sites within 500 bp upstream of their TSS (Supplementary Data 3).

Choanoflagellate NF-κBs can form heterodimers and have different abilities to bind DNA and activate transcription. Richter et al\textsuperscript{5} identified transcripts encoding RHD-containing NF-κB-related proteins in 12 choanoflagellate species. We chose to characterize the NF-κB proteins from Acanthocoea spectabilis (A) because it has three NF-κB-like proteins, which separated into multiple branches when phylogenetically compared to other choanoflagellate NF-κBs (Supplementary Fig. 1b). The three 43-NF-κB proteins contain ostensibly complete DNA-binding sequences, which are similar to other NF-κB proteins, and a putative NLS\textsuperscript{5}. These three 43-NF-κB proteins contain few residues C-terminal to the RHD (and no GRRs or ANK repeats). As a first step in characterizing these proteins, we created pcDNA-FLAG vectors for 43-NF-κB1, 43-NF-κB2, and 43-NF-κB3 (Fig. 6a) and transfected them into HEK 293T cells. As judged by anti-FLAG Western blotting, each plasmid directed the expression of an appropriately sized FLAG-tagged 43-NF-κB protein (Fig. 6b, Supplementary Table 1).

To determine the subcellular localization properties of the three 43-NF-κBs, we performed indirect immunofluorescence using DF-1 chicken cells transfected with each FLAG-tagged
vector. All three As-NF-kB proteins co-localized with Hoechst-stained nuclei in DF-1 cells (Fig. 6c).

We then performed a κB-site EMSA on whole-cell extracts from HEK 293T cells overexpressing each As-NF-kB, using Co-RHD as a positive control. As-NF-kB2 and 3 bound the κB-site probe to nearly the same extent as Co-RHD, but As-NF-kB1 only weakly bound the probe (Fig. 6d). We also assessed the transactivating ability of each As-NF-kB protein in a κB-site reporter assay in 293 cells, using the strongly activating Co-RHD protein as a positive control. Both As-NF-kB2 and 3 were able to activate transcription of the luciferase reporter above vector control levels (~1.6- and 2.7-fold, respectively), but As-NF-kB1 did not (Fig. 6e). We also determined the intrinsic transactivating ability of each As-NF-kB protein using a GAL4 DNA-binding domain-fusion protein reporter assay in yeast cells. In this assay, all three As-NF-kBs activated transcription above the GAL4 alone vector control levels, although As-NF-kB1 and 3 activated to a much lesser extent than As-NF-kB2 (Fig. 6f). From these data, we hypothesized that the homodimeric version of As-NF-kB1 was not capable of binding DNA and therefore, would not be expected to activate transcription of a κB-site reporter, but likely contained some intrinsic ability to activate transcription (based on its ability to activate transcription as a GAL4-fusion protein in yeast). In contrast, homodimeric As-NF-kB2 and 3 could both activate transcription in human cell-based κB-site and yeast GAL4-fusion protein reporter assays.
Since As-NF-kB1 did not substantially bind DNA or activate transcription when expressed alone in 293 cells, we hypothesized that As-NF-kB1 normally only acts as a heterodimer with the other As-NF-kBs. To determine whether As-NF-kB1 could interact with itself or As-NF-kB2 or 3, we performed a series of co-immunoprecipitation experiments (Fig. 6g). To do this, we subcloned As-NF-kB1, 2, and 3 into MYC-tagged vectors, co-transfected each with FLAG-As-NF-kB1 (as well as FLAG-As-NF-kB2 and 3) into 293T cells, and first immunoprecipitated FLAG-As-NF-kB1 from cell extracts using anti-FLAG beads. We then performed anti-FIAG and anti-MYC Western blotting on the immunoprecipitates to determine whether these NF-kBs could interact. MYC-As-NF-kB2 and MYC-As-NF-kB3 were both co-immunoprecipitated with FLAG-As-NF-kB1, as well as with FLAG-tagged versions of themselves (Fig. 6g, IP). In contrast, MYC-As-NF-kB1 was not co-immunoprecipitated with the FLAG-tagged version of itself. The MYC-As-NF-kB proteins were not seen in anti-FLAG immunoprecipitates when they were co-transfected with the empty vector control (Fig. 6g, IP).

From these data, it appears that all three As-NF-kBs can enter the nucleus when expressed in vertebrate cells, but they bind DNA and activate transcription to varying degrees. Of note, the transcriptional activating abilities of the three As-NF-kB proteins, both from a kB-site reporter in 293 cells and from a GAL4-site reporter in yeast, are not nearly as strong as the As-NF-kB DNA and activate transcription to varying degrees. Of note, the transfected each with FLAG-As-kB-site probe (GGGAATTCCC) and each of the indicated lysates from b. The NF-kB complexes and free probe are indicated by arrows. Raw image is in Supplementary Fig. 13. e A kB-site luciferase reporter gene assay was performed with expression vectors for the indicated proteins or the empty vector control (-) in HEK 293 cells. Luciferase activity is relative (Rel.) to that seen with the empty vector control (1.0). Values are averages of n = 3 biological replicates per sample each performed in triplicate, and are reported with standard error. Raw data are in Supplementary Data 5.

f A GAL4-site Loz2 reporter gene assay was performed with the indicated GAL4-fusion proteins or the GAL4 alone vector control (-) in yeast Y190 cells. Values are average β-gal units of n = 8 biological replicates. Raw data are in Supplementary Data 5. g Co-immunoprecipitation (IP) assays of MYC-tagged As-NF-kB1, As-NF-kB2, As-NF-kB3. In each IP assay, MYC-As-NF-kBs were co-transfected in HEK 293T cells with the pcDNA-FLAG vector control (-), or FLAG-As-NF-kB1, 2 or 3 as indicated. An IP using anti-FLAG beads was performed, and pulled down proteins were then analyzed by Western blotting with anti-FLAG (top) or anti-MYC (middle) antisera. An anti-MYC Western blot of 5% of the whole-cell (WC) lysates used in the pulldowns was also performed (bottom). Raw image is in Supplementary Fig. 13.

Discussion
In this manuscript, we have functionally characterized and compared NF-kB proteins from two protists. Our results demonstrate that functional DNA-binding and transcriptional-activating NF-kB proteins exist in both of these protists, but the overall structures, activities, and regulation of these proteins vary considerably, both among protists and when compared to animal NF-kBs.

In the vertebrate NF-kB proteins p100 and p105, C-terminal ANK repeats inhibit the DNA-binding activity of the RHD. IKK-induced proteasome-mediated processing of the ANK repeats is terminated within the GRR, and this processing activates the DNA-binding activity of these mammalian NF-kB proteins. In the Drosophila Relish protein, the C-terminal ANK repeats also inhibit RHD DNA binding, but Relish has no GRR and the C-terminal ANK repeats are removed by an internal site-specific proteolytic cleavage event, which does not involve the proteasome. Thus, the presence of ANK repeats and a GRR in Co-NF-kB suggests that proteasomal processing would lead to nuclear translocation and activation of its DNA-binding activity. Indeed, removal of the C-terminal residues of Co-NF-kB does allow it to enter the nucleus of vertebrate cells and unleashes its DNA-binding and transactivation activities (Fig. 2c-h). Moreover, the proteasome inhibitor MG132 blocks the basal processing of Co-NF-kB that is seen in transfected 293T cells (Supplementary Fig. 3b). However, this basal processing of Co-NF-kB in 293T cells is not enhanced by co-expression of IKK (unless C-terminal IKK target sequences are added, see Fig. 3b) and there are no obvious IKK target serines in Co-NF-kB nor are there any IKK homologs in the Capsaspora genome. If Co-NF-kB does undergo a signal-induced processing in Capsaspora cells, it is unlikely to be dependent on an IKK-like kinase.

Using proteomic analyses, Sebé-Pedrós et al. showed that Co-NF-kB protein levels and the phosphorylation of two C-terminal serines are low in aggregative stage cells and higher in cystic stage cells. This pattern of protein expression and phosphorylation of Co-NF-kB parallels its mRNA expression (Fig. 5c) and DNA-binding activity (Fig. 5d). Thus, phosphorylation of the two C-terminal serines identified by phosphoproteomic analysis may be involved in activation of the DNA-binding activity of Co-NF-kB or may simply reflect increased amounts of Co-NF-kB in aggregative vs cystic cells. The presence of increased levels of C-terminal Co-NF-kB residues in cystic cells, as judged by proteomic analysis, indicates that these sequences are intact in this stage, which has high NF-kB DNA-binding activity. In Drosophila, the C-terminal sequences of Relish are not degraded after signal-induced cleavage. Whether the C-terminal sequences of Co-NF-kB identified by proteomics are part of the full-length protein or have been cleaved but are still intact in native Capsaspora cells is not known. Co-NF-kB-specific antisera would be required to address this question.

In contrast to Capsaspora, the choanoflagellate NF-kB proteins lack C-terminal ANK repeats and GRRs, and all three As-NF-kB proteins are constitutively in the nucleus when overexpressed in vertebrate cells (Fig. 6c). We have not been able to identify an IkB-like protein in the A. spectabilis transcriptome. Thus, it is unclear whether choanoflagellate NF-kB is regulated by an ANK-dependent mechanism, or whether, for example, choanoflagellate proteins are constitutively nuclear in their native settings.

Constitutively nuclear localization of NF-kB proteins has also been seen in other settings. That is, we have previously shown that in both the sea anemone Aiptasia and one sponge most NF-kB staining is nuclear, and these NF-kB proteins are mostly processed in their native settings. Thus, we have argued previously that NF-kB proteins in these early-branching animals may be
constitutively in an active state, perhaps due to continual stimulation by upstream activating ligands or pathogens. Of note, most NF-κB p100 is also in its processed form in mouse spleen tissue\footnote{\textsuperscript{3,8,9}}. Among the 21 choanoflagellates for which there is sufficient transcriptomic/genomic information, only 12 have been found to have NF-κB genes\textsuperscript{3,8,9,10,12,20}. In seven of these 12 choanoflagellates there are multiple NF-κB genes\textsuperscript{3}. Thus, it appears that there have been gains and losses of NF-κB genes among the choanoflagellates. We note that NF-κB has also been lost in other organisms including C. elegans and ctenophores\textsuperscript{3}. The apparent absence of NF-κB in some choanoflagellates and its expansion in others (e.g., A. spectabilis) suggest that NF-κB has a specialized, rather than a general and required, biological function in choanoflagellates.

The presence of three A. spectabilisNF-κB-like proteins that can form heterodimers is the only example of an organism in a lineage predateing flies that has multiple NF-κB proteins which are capable of forming heterodimers. Thus, expansion of NF-κB genes has occurred at least two times in evolution, i.e., at least once in the metazoan lineage and once within choanoagellates. Furthermore, since each As-NF-κB homodimer has a different ability to form heterodimers, bind DNA, and activate transcription, it appears that there are subclasses of NF-κB dimers in A. spectabilis and likely in other choanoflagellates that have multiple NF-κBs. It is interesting to note that As-NF-κB1, 2, and 3 phylogenetically separate and cluster most closely with NF-κBs from certain other choanoagellate species that contain multiple NF-κBs (Supplementary Fig. 1b and ref. \textsuperscript{3}). For example, Savillea parva contains three NF-κBs, each of which clusters with a separate As-NF-κB (Supplementary Fig. 1b and ref. \textsuperscript{3}). Thus, we hypothesize that choanoagellates, like vertebrates, have evolved a mechanism for differential transcriptional control of genes through the use of combinatorial NF-κB dimer formation.

In both NF-κB-site reporter assays in human 293 cells and in GAL4-site reporter assays in yeast, Co-RHD is more strongly activating than the individual As-NF-κB proteins (i.e., compare data from Fig. 2e, f to Fig. 6c, \textit{f}). These two reporter assays are, of course, measuring different activities—NF-κB site-based activation vs. intrinsic activation when fused to the GAL4 DNA-binding domain—and in non-native cell types (human and yeast cells, respectively). As-NF-κB1 does not activate from the NF-κB-site reporter (Fig. 6e) because it does not bind to DNA (Fig. 6d); however, it can activate if brought to DNA by the GAL4 DNA-binding domain (Fig. 6f). In contrast As-NF-κB-3 (competent for DNA-binding [Fig. 6d]) can activate the NF-κB-site reporter (Fig. 6e), but when bound to GAL4 As-NF-κB-3 does not have substantial transactivation ability (Fig. 6f). These results suggest that heterodimers (e.g., between As-NF-κB-1 and As-NF-κB-3) are relevant dimers in vivo in their native organism. It will be interesting to assess the NF-κB-site transactivating abilities of As-NF-κB heterodimers, especially given that heterodimers are the most relevant dimer complexes in mammals\textsuperscript{12}.

The differential mRNA expression and DNA-binding activity of NF-κB among different life stages of Capasaspora suggest that Co-NF-κB has life stage-specific functions. It is notable that the DNA-binding activity of NF-κB in these different life stages correlates with differences in the levels of NF-κB mRNA\textsuperscript{16} and protein\textsuperscript{17}, rather than as differences in induced DNA-binding activity. In most metazoans, the activity of NF-κB is regulated at the post-transcriptional level, whereas in Aiptasia and corals, we have found that the levels of NF-κB mRNA, protein, and DNA-binding activity appear to be coordinately regulated, suggesting that the regulation of NF-κB occurs at the transcriptional level. For example, in Aiptasia, thermal bleaching causes transcriptional upregulation of NF-κB, which also results in increased protein expression of nuclear, DNA binding-active NF-κB\textsuperscript{24}. Similarly, treatment of the coral Orbicella faveolata with lipopolysaccharide results in increased expression of NF-κB pathway genes, rather than increased post-translational activation of NF-κB\textsuperscript{22}. Since there has yet to be an IκB-like homolog identified in choanoagellates, it is possible that the activity of the choanoagellate NF-κB proteins is fully regulated by transcriptional control of their genes. Thus, it appears that induced activity of NF-κB in several early-branching organisms is the result of transcriptional upregulation of NF-κB mRNA, rather than induced proteolysis, which occurs in most mammalian and fly systems.

Of the nearly 1350 genes whose expression correlated with NF-κB expression across different Capasaspora life stages, almost 20% contained predicted Co-NF-κB binding sites within 500 base pairs upstream of their TSS (Supplementary Data 3). While 20% is almost certainly an overestimate of Co-NF-κB direct or indirect gene targets, there are also likely additional NF-κB binding sites that could affect target gene expression. For example, ATAC-seq data suggested that the regulatory sites in the Capasaspora genome are present in first introns,\textsuperscript{5,9} 3' UTRs, as well as the proximal intergenic regions\textsuperscript{33}.

The list of ~1350 genes that are coordinately expressed with NF-κB mRNA most certainly contains genes that are controlled by other transcription factors or are activated by signaling or developmental events that do not involve NF-κB. Within this gene set, there are enriched GO terms other than immune system and development such as multi-organism process, interspecies interaction between organisms, and multicellular organism process. One could speculate that aggregation in Capasaspora and the correlative decrease in NF-κB in aggre-gated cells reflect a need to suppress collective immunity to form a symbiotic group. Alternatively, given that Capasaspora is normally a symbiont in the hemolymph of the snail B. glabrata\textsuperscript{9,14,15}, NF-κB may play a role in maintaining symbiosis, which has been suggested as one function of NF-κB in other organisms\textsuperscript{34}. In any event, it is clear that NF-κB may play several roles, and perhaps different roles, in each life stage of Capasaspora.

It is not clear what type of pathway controls activation of NF-κB in protists. In early-branching organisms, many common upstream NF-κB-activating components are missing, few in number, or lack critical domains (Supplementary Fig. 5). For example, Capasaspora does not contain homologs to TLRs, ILR-1 or TNFRs\textsuperscript{8}. However, choanoagellates do contain both full-length TLR-like and TIR-only homologs, as well as cGAS and STING homologs\textsuperscript{3}, which mediate innate immune responses in many organisms.

The differences in NF-κB that we describe here between Capasaspora and choanoagellates, which are members of the same group (Opisthokont) of protists, suggest that the diversification of NF-κB among all protists is considerable. The continued study of the evolution of NF-κB and other basally derived transcription factors will likely lead to an understanding of where and how these factors originated, as well as the basal biological functions they control.

**Methods**

**Phylogenetic analyses.** For Supplementary Fig. 1, the RHD sequences of NF-κB from C. owczarzaki and choanoagellates were compared to the NF-κB-like sequences present in several other organisms. Details on databases and sequence acquisition can be found in Supplementary Data 4. For the phylogenetic analysis in Supplementary Fig 1b, sequences were aligned by Clustal Omega\textsuperscript{25}. A maximum likelihood phylogenetic tree was created using www.phylogeny.fr in one-click mode\textsuperscript{36}, rooted to A. fl-9,14,15, NF-κB and the three As-NF-κB homologs, as well as cGAS and STING homologs\textsuperscript{3}, which mediate innate immune responses in many organisms.

The differences in NF-κB that we describe here between Capasaspora and choanoagellates, which are members of the same group (Opisthokont) of protists, suggest that the diversification of NF-κB among all protists is considerable. The continued study of the evolution of NF-κB and other basally derived transcription factors will likely lead to an understanding of where and how these factors originated, as well as the basal biological functions they control.

**Plasmids.** Human cell codon-optimized versions of Co-NF-κB and the three As-NF-κB were synthesized by GenScript and were provided as cDNA subclones in pUC57-Simple; their sequences are shown in Supplementary Figs. 6–9. Plasmids of the following types were used in these experiments: (1) pcDNA-based plasmids for the expression of epitope-tagged (FLAG, HA, MYC) proteins in tissue culture and Capasaspora cells; (2) pGB9-based plasmids for the expression of GAL4 (aa 1–147) fusion proteins in yeast; (3) p3x-KB-luc reporter plasmid that has three NF-κB-binding sites upstream of the luciferase reporter gene for transcriptional reporter gene assays in HEK 293 cells. Details of primers used for plasmid constructions
transferred to fresh 1.5-ml microcentrifuge tubes and stored at –80 °C. Protein concentration was determined using the Bio-Rad Protein Assay dye reagent (Bio-Rad, #2663000).

Whole-cell lysates of Capsaspora cells from each life stage (see below) were prepared for EMSAs. On the day of lysis, cells for each cell stage from five, 25 cm² tissue culture flasks (ThermoScientific Nunclon) were gently washed once with ice-cold PBS and were then removed with a rubber scraper into 1 ml of PBS for each plate and transferred into five, 1.5-ml microcentrifuge tubes at 3000 rpm for 5 min at 4 °C and then combined and resuspended by pipetting in 250 μl of AT Lysis Buffer (see above). Samples were then passaged five times through a 27.5-gauge needle to ensure complete lysis. NaCl was then added to a final concentration of 150 mM and protein lysates were clarified by centrifugation at 13,000 rpm for 25 min at 4 °C. Supernatants were then transferred to fresh 1.5-ml microcentrifuge tubes and stored at –80 °C. Protein concentration was determined using the Bio-Rad Protein Assay dye reagent (Bio-Rad).

**Western blotting and electrophoretic mobility shift assays (EMSA)** Western blotting was performed essentially as described in detail elsewhere. Briefly, extracts (~50 μg) of transfected HEK 293T cells were resuspended in 2X SDS sample buffer (0.125 M Tris-HCl [pH 8.0], 4.6% w/v SDS, 20% w/v glycerol, 10% v/v β-mercaptoethanol, 0.2% w/v bromophenol blue) and heated at 90 °C. Proteins were then separated on 7.5% or 10% SDS-polyacrylamide gels. Proteins were then transferred to nitrocellulose at 4 °C for 2 h followed by 170 mA overnight. The membrane was blocked in TBST (10 mM Tris–HCl [pH 7.4], 150 mM NaCl, 0.1% v/v Tween-20) containing 5% powdered non-fat milk (Carnation) for 1 h at room temperature. Filters were incubated at 4 °C with anti-FLAG primary antiserum (1:1000, Cell Signaling Technology, #2368) or anti-HA primary antiserum (1:500, Santa Cruz Biotechnology, #sc-805) diluted in TBST containing 5% non-fat powdered milk. After washing, filters were incubated with anti-rabbit horseradish peroxidase-linked secondary antiserum (1: 4000, Cell Signaling Technology, #7074). Immunoreactive proteins were detected with SuperSignal West Dura Extended Duration Substrate (ThermoFisher, #34075) and imaged on a Sapphire Biomolecular Imager (Azure Biosystems).

We have recently described our protocol for EMSA in detail. DNA-binding reactions were performed in DNA-binding reaction buffer (10 mM HEPES [pH 7.8], 50 mM KCl, 1 mM DTT, 1 mM EDTA, 4% w/v glycerol, 40 μg/ml poly-dI-dC, 2X BSA) containing 200,000 cpm of a 32P-labeled 27-mer double-stranded probe (GGGAATTCCTC, see Supplementary Table 5) and whole-cell extracts from 293T (50 μg total protein) or Capsaspora cells (70 μg total protein) prepared in AT Lysis Buffer. Samples were incubated at 30 °C for 20 min. Samples were then separated on 5% non-denaturing polyacrylamide gels. After drying, gels were exposed to a phosphor screen, and then imaged on a Sapphire Biomolecular Imager (Azure Biosystems).

**Reporter gene assays.** NF-kB-site luciferase reporter assays were performed in HEK 293 cells as these assays do not require the SV-40 T-antigen, which can also affect expression from some reporter plasmids. Co-transfections with 0.5 μg of 3x κB-site luciferase reporter plasmid and 2 μg of pDNA-FLAG expression plasmids were performed in triplicate for each expression plasmid in 35-mm plates using 15 μg PEI as described above for HEK 293T cells. Two days after transfection, cells were clarified by centrifugation at top-speed for 2 min in a microcentrifuge, and the amount of β-galactosidase activity was measured on lysates that contained equal amounts of protein using the Luciferase Assay System (Promega, #E397A). Luciferase activity for each triplicate were averaged and were normalized to the empty vector control (set as 1.0). Three independent experiments were performed (with triplicate samples), and values are reported as the averages of the three experiments plus standard error (SE), determined as follows:

\[
SE = \sigma / \sqrt{n}
\]

where σ is the sample standard deviation, and n is the number of independent experiments. For GAL4-fusion protein reporter gene assays, yeast strain Y190, which contains an integrated GAL4-σ162 reporter locus, were used. Y190 cells were transformed with pGBl9-based plasmids and stable transformants were selected on plates containing yeast media but lacking tryptophan. Independent colonies were picked and grown overnight in 3 ml of liquid media lacking tryptophan. The density of the culture was measured by visible light spectrometry at OD600, as an indication of cell number. For reporter gene activity, 1.0 ml of cells was pelleted and washed in 1 ml of 100 mM Tris [pH 7.6], 0.5% Triton X-100. To each sample, 450 μl of Z/ONPG solution (60 mM NaH2PO4, 40 mM KCl, 1 mM MgSO4, 0.3% w/v mercaетодithanol, 0.8 mg/ml ONPG [orthonitrophenyl-β-D-galactopyranoside]; Sigma, 1 mM DTT, 0.005% SDS) was added. Samples were incubated at 30 °C until yellow color appeared or for a maximum of 30 min for the negative control. The reaction was terminated by the addition of 225 μl of 1M Na2CO3. Samples were then centrifuged through a 27-gauge needle, and the amount of β-galactosidase activity was measured at OD420.
as follows:

\[ V = \frac{1000 \times OD_{485}}{V \times t \times OD_{340}} \]  

(2)

where \( V \) = volume of cells used (1.0 ml) and \( t \) = time (min) until the reaction was terminated.

Values for samples (number of samples indicated in figure legends) were averaged and are reported as average units with SE (determined as in Eq. 1 above).

Co-immunoprecipitations. HEK 293T cells in 60-mm dishes were co-transfected with 2.5 μg of MYC-As-NFκB1, 2, or 3 and 2.5 μg of either a pDNA-FLAG empty vector, FLAG-As-NFκB1, 2, or 3 (Fig. 6g), or were co-transfected with 2.5 μg of MYC-As-NFκB1, 2, or 3 and 2.5 μg of either pDNA-FLAG or FLAG-Co-2-RSHD (Fig. 2g), using PEI as described above. Lysates were prepared 48 h later in AT Lysis Buffer and were incubated with 50 μl of a PBS-washed anti-FLAG bead 50% slurry (Sigma, #A2220) overnight at 4°C with gentle rocking. The next day, the beads were washed three times with PBS. The pellet was then boiled in 2× SDS sample buffer, the supernatant was electrophoresed on a 7.5% SDS-polyacrylamide gel, and proteins were transferred to nitrocellulose. The membrane was then probed with a rabbit anti-MYC antisera (1:1000, Cell Signaling Technology, #71D10), followed by anti-rabbit horseradish peroxidase-linked secondary antisera (1: 4000, Cell Signaling Technology, #2368). Complexes were detected with SuperSignal West Dura Extended Duration substrate and an image was obtained on a Sapphire Biomolecular Imager, as described above. The membrane was stripped and probed with mouse anti-FLAG antisera (1:1000, Cell Signaling, #8164) and goat anti-mouse horseradish peroxidase secondary antibody (1:4000, Cell Signaling, #7076). The filter was then imaged as described above.

Indirect immunofluorescence. For immunofluorescence of transfected DF-1 cells, the media was removed, and cells were washed three times with room temperature PBS. Cells on coverslips were fixed in 100% methanol (pre-chilled at −20°C for 10 min. Coverslips were allowed to dry at room temperature for about 10–20 min on a paper towel. When dry, fixed cells on coverslips were blocked with PBS containing 3% calf serum (CaS, Sigma, #C8056) for 30 min at room temperature. Cells were then incubated with dilute primary antisera (rabbit anti-FLAG, 1:90, Cell Signaling, #2368, or mouse anti-Myc, 1:100, Santa Cruz Biotechnology, #9E10) for 1 h at 37°C. Coverslips were then washed 4× with PBS/3%CaS. Coverslips were then incubated with secondary antibody diluted 1:100 (goat anti-rabbit conjugated to Alexa Fluor-488, Invitrogen, #A21202, or goat anti-mouse conjugated to Alexa Fluor-555, Invitrogen, #A21202) for 1 h at room temperature. Coverslips were then washed four times with PBS, and incubated with Hoechst 33421 (4:1000, Invitrogen, #H3570) for 15–20 min at room temperature. Coverslips were washed three times with PBS and then mounted on a glass slide with 25–40 μl of mounting medium (Slow Fade, Invitrogen, #S36937). Slides were then stored in the dark at either 4°C (short-term) or −20°C (long-term). Cells were imaged on a Nikon C2 Si confocal microscope at 405 nm (Hoechst), 488 nm (Alexa Fluor-488), and 561 nm (Alexa Fluor-555). Images were taken with a Ti-E Spectral imaging system.

Indirect immunofluorescence of transfected Capsaspora cells was carried out essentially as described above for DF-1 cells, except Capsaspora cells were fixed with 4% paraformaldehyde for 20 min (instead of methanol). Cells were then blocked with PBS/3%CaS for 30 min. As described for DF-1 cells, Capsaspora cells were then incubated sequentially with anti-FLAG primary antisera, Alexa Fluor-488-conjugated goat anti-rabbit secondary antisera, and Hoechst 33342 stain were then incubated sequentially with anti-FLAG primary antiserum, Alexa Fluor-488, and Hoechst 33342 stain were then incubated with 4% paraformaldehyde for 20 min (instead of methanol). Cells were then taken with a Ti-E Spectral imaging system.

RNA-seq analysis and annotation. RNA-seq data from each life stage of Capsaspora was obtained from Sebé-Pedrós et al. To identify potential target genes of NFκB, we sorted all RNA-seq data to identify genes that were differentially expressed in the same manner as NFκB for each life stage (lowest expression in aggregative, medium expression in filopodic, highest expression in cystic). We discarded genes that were not expressed in a given life stage (i.e., had an RPKM of 0). Of the 8674 total genes, 1348 genes were differentially expressed in a manner similar to Co-NFκB. Of these 1348 genes, 389 genes had annotated homologs, and 305 had human homologs (Supplementary Data 2). We then performed GO analysis (http://panthierdb.org) by entering the UniProt ID of each gene and selecting Homo sapiens. We then created and analyzed the Biological Processes that were present in this gene list. To identify genes with potential upstream NFκB binding sites among these 1348 genes, we extracted the 500 base pair sequence upstream of each gene. We then imported these 1348 upstream regions into MEME-FIMO and scanned the sequences using the PBM-generated Co-NFκB motif (Supplementary Data 3).

Statistics and reproducibility. Reporter gene assays in tissue culture cells (Figs. 2e and 6d) were performed with multiple independent samples as described in Methods and figure legends, and are reported with standard error (as determined in Eq. 1). The sample sizes were as follows: Fig. 2e, n = 3 (with triplicate samples); Fig. 6d, n = 3 (with triplicate samples); Fig. 2l, n = 34, except Co-NFκB, n = 8; and Fig. 6f, n = 8. For PBM analysis, the median fluorescence values for each unique DNA binding sequence were determined over eight replicate probe measurements. Log median fluorescence values were transformed to a 'z-score' using the mean (μ) and variance (σ) of the log median fluorescence values for the 1195 random background DNA sequences. The best fit line and R2 values for the scatter plots were calculated by excluding z-scores lower than 2 for each comparison. Other experiments, e.g., EMSAs, Western blots, were performed at least two times and had similar results to the ones reported here. Moreover, such experiments have internal controls for molecular weight markers and loading accuracy.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. All data generated and analyzed during this study are included in this published article (and its Supplementary information files). The raw images for Western blots and EMSAs can be found in Supplementary Figs. 10–14. The raw data for reporter assays can be found in Supplementary Data 5. Capsaspora RNA-seq data for different life stages were taken from ref. 13, https://doi.org/10.6084/m9.ghareg.458547.v2. Capsaspora proteomic and phosphoproteomic data from were taken from ref. 17. Accession number PRIDE PXD004567. Data for choanoflagellate NFκB proteins were taken from ref. 5, Accession at NCBI SRA under BioProject PRNA419411. Additional data relating to the study are available from the corresponding author upon reasonable request.

Received: 27 March 2021; Accepted: 24 November 2021; Published online: 16 December 2021

References

1. Gilmore, T. D. Introduction to NF-κB. players, pathways, perspectives. Oncogene 25, 6680–6684 (2006).
2. Ghosh, S. & Hayden, M. S. Celebrating 25 years of NF-κB research. ImmunoL Rev. 246, 6–13 (2012).
3. Williams, L. M. & Gilmore, T. D. Looking down on NF-κB. Mol. Cell. Biol. 40, e00104–2020 (2020).
4. Suga, H. et al. The Capsaspora genome reveals a complex unicellular prehistory of animals. Nat. Commun. 4, 1–9 (2013).
5. Richter, D., Foziouni, P., Eisen, B. M. & King, N. Gene family innovation, conservation and loss on the animal stem lineage. eLife 7, e34226 (2018).
6. Sebé-Pedrós, A., de Mendoza, A., Lang, B. F., Degnan, B. M. & Ruiz-Trillo, I. Unexpected repertoire of metazoan transcription factors in the unicellular choanoflagellate, Capsaspora owczarzaki. Mol. Biol. Evol. 28, 1241–1242 (2011).

7. Simpson, A. G. B. et al. Revisions to the classiﬁcation, nomenclature, and diversity of eukaryotes. J. Eukaryot. Microbiol. 66, 4–119 (2019).

8. Senftleben, U. et al. Activation by IKK of NF-κB. Curr. Biol. 27, PR829–PR830 (2017).

9. Ruíz-Trillo, I. et al. Capsaspora owczarzaki is an independent opisthokont lineage. Curr. Biol. 14, R946–7 (2004).

10. Ruíz-Trillo, I., Roger, A. J., Burger, G., Gray, M. W. & Lang, B. F. A phylogenomic investigation into the origin of metazoans. Mol. Biol. Evol. 25, 664–72 (2008).

11. Owczarzak, A., Stibbs, H. H. & Bayne, C. J. The destruction of Schistosoma mansoni mother sporocysts isolated from Biomphalaria glabrata. J. Invert. Pathol. 35, 26–33 (1980).

12. Sebé-Pedrós, A. et al. Regulated aggregative multicellularity in a close unicellular relative of metazoans. eLife 2, e01287 (2013).

13. Sebé-Pedrós, A. et al. High-throughput proteomics reveals the unicellular roots of animal phosphosignaling and cell differentiation. Dev. Cell 39, 186–197 (2016).

14. Sleigh, S. et al. The genome of the choanoﬂagellate Monosiga brevicollis and the origin of metazoans. Nature 451, 783–788 (2008).

15. Fairclough, S. R. et al. Premetazoan genome evolution and the regulation of cell differentiation in the choanoﬂagellate Salpingoeca rosetta. Genome Biol. 14, R15 (2013).

16. Sun, S.-C. Non-canonical NF-κB signaling pathway. Cell Res. 21, 71–85 (2011).

17. Williams, L. M. et al. A conserved Toll-like receptor to NF-κB signaling pathway in the endangered coral Orbicella faveolata. Dev. Comp. Immunol. 79, 128–136 (2018).

18. Williams, L. M. et al. Transcription factor NF-κB in a basal metazoan, the sponge, has conserved and unique sequences, activation, and regulation. Dev. Comp. Immunol. 104, 103559 (2020).

19. Mansﬁeld, K. M. et al. Transcription factor NF-κB is modulated by symbiotic status in a sea anemone model of cnidarian bleeding. Sci. Rep. 7, 16025 (2017).

20. Wolenski, F. S., Bradtman, C. A., Finnerty, J. R. & Gilmore, T. D. NF-κB is required for cnidocyte development in the sea anemone Nemastola vectensis. Dev. Biol. 373, 205–215 (2013).

21. Wolenski, F. S. et al. Characterization of the core elements of the NF-κB signaling pathway of the sea anemone Nemastola vectensis. Mol. Cell. Biol. 31, 1076–1087 (2011).

22. Siggers, T. et al. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. Nat. Immunol. 13, 95–102 (2011).

23. Steward, R. Dorsal, an embryonic polarity gene in Drosophila, is homologous to the vertebrate proto-oncogene, c-rel. Science 238, 692–694 (1987).

24. Kappler, C. et al. Insect immunity. Two 17 bp repeats nesting a kappa B-related binding site specify activating interaction with the vertebrate proto-oncogene, c-rel. Science 238, 692–694 (1987).

25. Reichhart, J. M. et al. Expression and nuclear translocation of the rel/NF-κB related morphogen dorsal during the immune response of Drosophila. C. R. Acad. Sci. III 316, 1218–1224 (1993).

26. Stoven, S. et al. Caspase-mediated processing of the Drosophila NF-κB factor Relish. Proc. Natl. Acad. Sci. USA 100, 5991–5996 (2003).

27. Senftleben, U. et al. Activation by IKKα of a second, evolutionarily conserved, NF-κB signaling pathway. Science 293, 1495–1499 (2001).

28. Sebé-Pedrós, A. et al. The dynamic regulatory genome of Capsaspora and the origin of animal multicellularity. Cell 165, 1224–1237 (2016).

29. Mansfield, K. M. & Gilmore, T. D. Innate immunity and cnidian-Symbiodiniaceae mutualism. Dev. Comp. Immunol. 90, 199–209 (2019).

30. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539 (2011).

31. Deereper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, W465–W469 (2008).

32. Foster, D. Development of a spontaneously immortalized chicken embryo fibroblastic cell line. Virology 210, 305–311 (1995).

33. Aguierre Carrión, P. J., Williams, L. M. & Gilmore, T. D. Molecular and biochemical approaches to study the evolution of NF-κB signaling in basal metazoans. Franzoso G. E. Zazzeroni, eds. Humana, New York, NY. Mol. Biol. Cell. 2366, 67–91 (2021).

34. Siggers, T., Gilmore, T. D., Barron, B. & Povesse, A. Characterizing the DNA binding site speciﬁcity of NF-κB with protein binding microarrays. Mol. Biol. Cell. 1280, 609–630 (2015).

35. Dudley, A. M., Aach, J., Steffen, M. A. & Church, G. M. Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. Proc. Natl. Acad. Sci USA 99, 7554–7559 (2002).

36. Bailey, T. L. & Noble, W. S. Searching for statistically significant regulatory modules. Bioinformatics 2, i15–i25 (2003).

Acknowledgements
This research was supported by National Science Foundation grants IOS-1354935 and IOS-1937650 (to T.D.G.). L.M.W. was supported by an NSF Graduate Research Fellowship, a Warren-McLeod Graduate Fellowship in Marine Sciences, and a Charles Terner Award (Boston University). J.S., J.P., and E.K.A. were supported by NSF REU grant IOS-1659665 (T.D.G.). S.S. was supported by the Boston University Undergraduate Research Opportunities Program, and A.A. was supported by the BU GRoW program. Designated authors performed research as part of the undergraduate Molecular Biology Laboratory course BB522 (Spring 2019, 2020), and were supported by funds from the BU Biology Department.

Author contributions
Students in the Molecular Biology laboratory, C.J.D., and N.R.S. designed and created pDNA-FLAG and pGt9-based expression plasmids for Co-NF-kB and As-NF-kB proteins, performed indirect immunofluorescence in DF-1 cells, and most yeast reporter assays. S.S. created plasmids and did Western blots for Fig. 6g and Supplementary Fig. 3. J.S. performed the Co-NF-kB luciferase reporter assays (Fig. 2e). J.P. performed some yeast reporter assays and the co-immunoprecipitation experiments (Fig. 2f, g). E.A. performed preliminary experiments for Fig. 3. A.A. prepared the NF-kB comparisons (Supplementary Fig. 1a). C.J.D. helped with cloning, and performed experiments for Figs. 2c, g, and 6a. P.I.A.C. performed the EMSA on Fig. 2d. T.W. carried out the PB experiments and analyzed the PB data. L.M.W. and T.D.G. performed experiments for the remaining figures, oversaw the project, and wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02924-2.

Correspondence and requests for materials should be addressed to Thomas D. Gilmore.

Peer review information Communications Biology thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: George Inglis.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2021

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021
BB522 Molecular Biology Laboratory
Dana H. M. Alburi¹, Ludmila Anisimov², Aria Y. Armstrong¹, Sydney J. Badger², Elham Banaie¹, Joana A. Barbosa Teixeira², Madeleine T. Billingsley², Anoush Calikyan¹, Yinxing Chen², Aidan B. Coia¹, Daniel Cutillo¹, Breanna R. Dooling¹, Parth P. Doshi¹, Kyra R. Dubinsky¹, Berta Escude Velasco², Jabari R. Evans², Jasmine Gordon¹, Huibo Guan², Spiro N. Haliotis², Niccolas T. Hood¹, Yen-Chun Huang¹, Wenjing Jiang¹, Isabelle C. Kreber¹, Ekin B. Kurak¹, Cheng-Che Lee², Tanner M. Lehmam¹, Savina J. W. Lin¹, Edward Liu¹, Kevin Liu², Yen-Yu Liu², Alexandra L. Luther², Alexa A. Macgranaky-Quaye¹, Daniel J. Magat¹, Lauren E. Malsick², Parmida Masoudi¹, Parsida Masoudi¹, Chad R. H. Max¹, Ethan Z. McCaslin², Eleanor T. McGeary¹, Kathleen M. McLaughlin¹, Victoria S. A. Momyer², Lake D. Murphy², Sonny V. Nguyen¹, Kareemah Ni², Leon Novak², Roberto Nunes Campos E. Santos², Yemi D. Osayame², Jun Bai Park Chang², Harshal M. Patel², Tony V. Pham², Sheila M. Phillips¹, Jhonathan Perea Piedrahita¹, Tricia L. Post², Rebecca A. Prather¹, Pauline I. Reck², Jaime A. Rodriguez², Kirquenique A. Rolle², Joseph A. Salzo¹, Kathryn M. Satko², Davis G. Settipane², Kara J. Sevola², Mithil V. Shah², Viktoriya Skidanova², Georgia M. Snyder², Rebecca J. Sprague¹, Ryan A. Stagg¹, Danielle Tong¹, Andreas A. Towers¹, Nicholas W. Turgiss², Natalie S. Wheeler² & Ann S. Yung²

²Program in Biochemistry & Molecular Biology, Boston University, Boston, MA, USA.