Patchy Distribution of GTPases of Immunity-Associated Proteins (GIMAP) within Cnidarians and Dinoflagellates Suggests a Complex Evolutionary History

Jenny C. Coelho[1,2], Ethan D. Calhoun[1], Grant N. Calhoun[1], and Angela Z. Poole 🆔[1,*]

[1]Department of Biology, Berry College, Mt. Berry, Georgia, USA
[2]Department of Biology, University of North Carolina at Chapel Hill, North Carolina, USA

*Corresponding author: E-mail: apoole@berry.edu.

Abstract

GTPases of Immunity-Associated Proteins (GIMAP) are a group of small GTP-binding proteins found in a variety of organisms, including vertebrates, invertebrates, and plants. These proteins are characterized by the highly conserved AIG1 domain, and in vertebrates, have been implicated in regulation of the immune system as well as apoptosis and autophagy, though their exact mechanism of action remains unclear. Recent work on cnidarian GIMAPs suggests a conserved role in immunity, apoptosis, and autophagy—three processes involved in coral bleaching, or the breakdown of cnidarian-dinoflagellate symbiosis. Therefore, to further understand the evolution of GIMAPs in this group of organisms, the purpose of this study was to characterize GIMAP or GIMAP-like sequences utilizing publicly available genomic and transcriptomic data in species across the cnidarian phylogeny. The results revealed a patchy distribution of GIMAPs in cnidarians, with three distinct types referred to as L-GIMAP, S-GIMAP, and GIMAP-like. Additionally, GIMAPs were present in most dinoflagellate species and formed seven well-supported clades. Overall, these results elucidate the distribution of GIMAPs within two distantly related eukaryotic groups and represent the first in-depth investigation on the evolution of these proteins within both protists and basal metazoans.

Key words: AIG1, corals, Symbiodiniaceae, phylogenetics, symbiosis.

Significance

Previous research indicated a role for cnidarian GIMAPs in immunity, apoptosis, and autophagy, three processes known to be impacted by their symbiotic relationship with dinoflagellates. Additionally, previous work revealed GIMAPs were present in two symbiotic corals, but lacking in a nonsymbiotic anemone, raising questions about their evolution within this phylum. Therefore, as these proteins could be important in regulating processes related to symbiosis, the goal of this research was to characterize GIMAP proteins in a diversity of cnidarians and determine if these proteins are also present in dinoflagellates. The results indicate a patchy distribution of three GIMAP types within cnidarians and widespread distribution of GIMAPs in dinoflagellates, indicating a complex evolutionary history and broader distribution within eukaryotes than previously noted.

Introduction

GIMAPs, or GTPases of Immunity-Associated Proteins, are a recently discovered group of small G proteins that in vertebrates, play a role in immunity, apoptosis, and autophagy (Cambot et al. 2002; Nitta et al. 2006; Pascall et al. 2018).

GIMAPs are characterized by the presence of the AIG1 domain, which functions in GTP binding (Nitta et al. 2006). Additionally, GIMAPs are traditionally defined as containing a coiled-coil motif, with a proposed function in protein–protein interactions (Lu et al. 2020; Limoges et al. 2021).

Although the exact mechanism of action is still unclear for GIMAPs, it is known that GTPase activity is stimulated by dimerization, and that only some GIMAPs possess catalytic capabilities (Schwefel et al. 2013). It has been shown that GIMAPs form oligomer scaffolds, which may contribute to their ability to alter cellular activities. Specifically, it was proposed that GIMAPs that lack GTPase activity exist in the GTP-bound state, which favors scaffold formation, whereas those with catalytic activity disrupt scaffold formation (Schwefel et al. 2010, 2013; Ciucci and Bosselut 2014).

Although each vertebrate GIMAP plays a specific role in the cell, most are thought to function in regulation of the adaptive immune system, particularly in T-cell development and homeostasis (Ciucci and Bosselut 2014; Limoges et al. 2021). Some GIMAPs interact with Bcl-2 family members, which function to induce or inhibit apoptosis, providing a mechanism by which GIMAPs control the survival or death of immune cells. Specifically, mouse GIMAP4 has been shown to interact with Bax, a pro-apoptotic Bcl-2 member, whereas GIMAP3 and GIMAP5 interact with the anti-apoptotic Bcl-2 and Bcl-xL proteins (Nitta et al. 2006). GIMAPs have also been shown to play a role in autophagy of immune cells. Specifically, GIMAP6 localizes to autophagosomes in starved T-cells, and was shown to be important in maintaining appropriate peripheral T-cell levels (Pascall et al. 2013, 2018). Overall, vertebrate GIMAPs are involved in several cell death related processes to regulate immune cell development and homeostasis.

Although previously thought to be limited to vertebrates and plants, GIMAPs have also been identified in some invertebrates including cnidarians, molluscs, cephalochordates, and hemichordates (Weiss et al. 2013; McDowell et al. 2016; Lu et al. 2020). Additionally, there is growing evidence for a conservation of function in immunity, apoptosis, and autophagy for cnidarians GIMAPs. Specifically, massive upregulation of three GIMAP transcripts in the coral *Acropora millepora* was observed in response to treatment with muramyl dipeptide, a component of the bacterial cell wall, suggesting a role for these proteins in immunity (Weiss et al. 2013). Furthermore, a study in the anemone *Exaiptasia pallida* documented an overall downregulation of GIMAPs in response to induction of apoptosis and autophagy, indicating further conservation of function to their vertebrate counterparts (Bailey et al. 2020). Immunity, apoptosis, and autophagy are three processes linked to both cnidarian-dinoflagellate symbiosis and coral disease (Dunn et al. 2007; Hanes and Kempf 2013; Fuess et al. 2017). Since the breakdown of cnidarian-dinoflagellate symbiosis, or coral bleaching, and coral disease are currently major threats to the reef ecosystem (Precht et al. 2016; Hughes et al. 2018), further research into cnidarian GIMAPs is warranted.

Cnidarian GIMAPs are also of interest from an evolutionary perspective, as it was previously shown that GIMAPs were present in two closely related symbiotic coral species, but absent from the nonsymbiotic anemone *Nematostella vectensis* (Weiss et al. 2013). This raises the question of whether symbiotic status can explain the distribution of GIMAPs in cnidarians. To fully understand GIMAPs in the holobiont, it was also of interest to explore GIMAP distribution within the dinoflagellate symbionts. Although GIMAPs have not formally been characterized in protists, proteins containing the AIG1 domain have been shown to play a role in virulence of the protistan parasite *Entamoeba histolytica* via regulation of adhesion to host cells (Nakada-Tsukui et al. 2018). Therefore, investigation of whether dinoflagellates contain GIMAPs or AIG1 proteins is warranted as they may play a similar role to those in *E. histolytica* in the context of cnidarian-dinoflagellate symbiosis. Furthermore, GIMAP distribution remains completely unexplored in protists, making this a novel area of research.

The goal of this study was to identify GIMAP or GIMAP-like sequences in a diversity of cnidarian and dinoflagellate species and perform phylogenetic analysis to determine their relationship to known GIMAPs. This information provides a more holistic understanding of GIMAP evolution across eukaryotes and on the relationship between symbiotic status and GIMAP distribution. The results indicate a patchy distribution of GIMAPs in Cnidaria with three distinct types, whereas GIMAPs are present in almost all dinoflagellate species surveyed. There were no clear patterns observed between GIMAP distribution and symbiotic status for either cnidarians or dinoflagellates, suggesting other factors influence GIMAP evolution.

Results

Cnidarian GIMAPs Have a Patchy Distribution and Form Three Distinct Clades on the Phylogenetic Tree

114 GIMAP sequences were identified in 29 different anthozoan species, including those belonging to orders Actiniaria, Corallimorpharia, and Scleractinia (table 1 and supplementary file S1, Supplementary Material online). However, no GIMAPs were found in species within class Hydrozoa, Cubozoa, or Scyphozoa (table 1). Even within Class Anthozoa, a patchy distribution of GIMAPs was observed, with some species containing many GIMAPs and others, such as those in Subclass Octocorallia, containing none.

The BLAST searches demonstrated that some anthozoan GIMAPs seemed more similar to human sequences, whereas others were more similar to the *A. millepora* sequences, leading to the hypothesis that there were multiple types of cnidarian GIMAPs. To investigate this hypothesis, maximum likelihood phylogenetic analysis was performed. The results revealed that there were three groups of cnidarian GIMAPs that separate into distinct clades on the phylogenetic tree with moderate to high bootstrap support (fig. 1). The first group, termed L-GIMAPs (L for long length) included

Table 1

Distribution of GIMAP Proteins in Anthozoans

Subclass	Order	Genus	L-GIMAP	S-GIMAP	GIMAP-Like
Hexacorallia	Actiniaria	Actinia		✓	
		Anthopleura		✓	
		Aulactinia		✓	
		Calliactus		✓	
		Exaiptasia		✓	
		Nemanthus			
		Nematostella			
		Telmatactis		✓	
	Scleractinia	Acropora	Some	Some	
		Agaricia	✓	✓	
		Astreopora	✓		
		Dendrophyllia	✓		
		Eguchipsammia			✓
		Fungia	✓		✓
		Galaxea	✓		✓
		Goniastrea	✓		✓
		Orbicella	✓		
		Pocillopora			✓
		Porites	✓		Some
		Rhizotrochus	✓		
		Siderastrea	✓		✓
		Stylophora	✓		✓
	Corallimorpharia	Amplexidiscus	✓		
		Discosoma	✓		
Octocorallia	Alcyonacea	Gorgonia			
	Pennatulacea	Renilla			

sequences from corallimorpharians and scleractinians, but no actinarian sequences. The second group, termed S-GIMAPs (S for short length), was sister to the L-GIMAPs and contained sequences from actinarians and limited scleractinians including the genera *Acropora* and *Agaricia*. Notably, the GIMAPs previously characterized in both *A. millepora* (Weiss et al. 2013) and *Exaiptasia pallida* (Bailey et al. 2020) fell within this clade. Lastly, the third group contained sequences from select scleractinian species, and was termed GIMAP-like proteins.

The distribution of these three GIMAP types did not show a clear pattern in regards to the anthozoan phylogeny or symbiotic status. For example, Actiniaria contained only S-GIMAP sequences. However, consistent with previous research (Weiss et al. 2013), *N. vectensis* lacked GIMAP sequences as did another anemone, *Nemanthus* sp. However, for *Nemanthus* sp., transcriptome data were utilized and therefore this conclusion is not definitive. Out of the anemone species that did contain S-GIMAPs, about half were symbiotic and half nonsymbiotic, indicating no relationship between symbiotic status and distribution for S-GIMAPs. Both corallimorph species searched contained multiple L-GIMAPs. Lastly, GIMAP distribution was much less consistent within Scleractinia. Of the three GIMAP types identified on the tree, some species searched contained only one GIMAP type, whereas others contained two, and this was not always consistent within a family or genus. For example, the genera *Acropora* and *Agaricia* contained species with both L- and S-GIMAPs. However, within *Acropora*, where sequence data for several species are readily available, the story was even more complex. For the two genomes searched within this genus, *Acropora digitifera* possessed both L- and S-GIMAPs, whereas *Acropora tenuis* had only L-GIMAPs. Another example includes the family Pocilloporidae, where two species, *Stylophora pistillata* and *Pocillopora damicornis* were searched. *S. pistillata* contained L-GIMAPs and GIMAP-like proteins, whereas *P. damicornis* was one of the few species that contained only GIMAP-like proteins. When considering symbiotic status, the large majority of the corals searched were symbiotic, but nonsymbiotic species did contain either L-GIMAPs or GIMAP-like proteins. Overall, these data indicate that the distribution of all three GIMAP types within cnidarians is patchy and does not completely mirror the anthozoan phylogeny or symbiotic status.

Presence of the Coiled-Coil Motif Correlates to Cnidarian GIMAP Groups

Interestingly, when looking at the phylogeny, a pattern emerges in regards to the COILS results. The output revealed
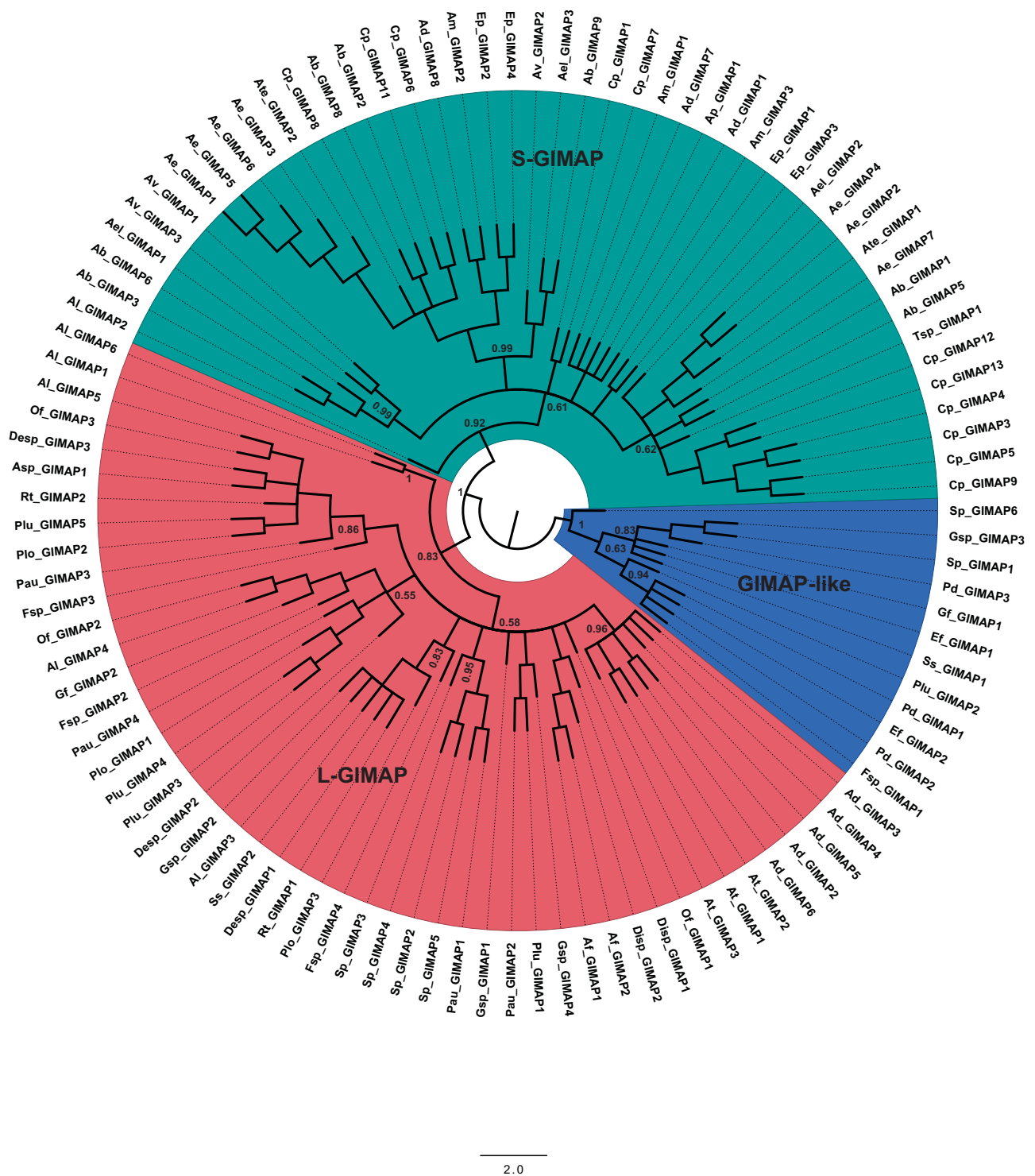
Fɪɢ. 1.—Maximum likelihood phylogenetic analysis of cnidarian GIMAP proteins. Phylogenetic analysis was performed using PhyML with 500 bootstrap replicates. The three GIMAP types are indicated with color blocking and the bootstrap value proportions for major nodes on the tree are indicated.

that for the 114 cnidarian sequences, 57 (50%) had a probability of 0.5 or greater of containing a coiled-coil region (supplementary file S1, Supplementary Material online). Additionally, the presence of the coiled-coil domain was

strongly correlated with GIMAP type (Cramer's V test, 0.56). The majority of the sequences that did not contain the coiled-coil region are either within the S-GIMAPs or GIMAP-like proteins. Specifically, within the three groups on the tree, 28% of
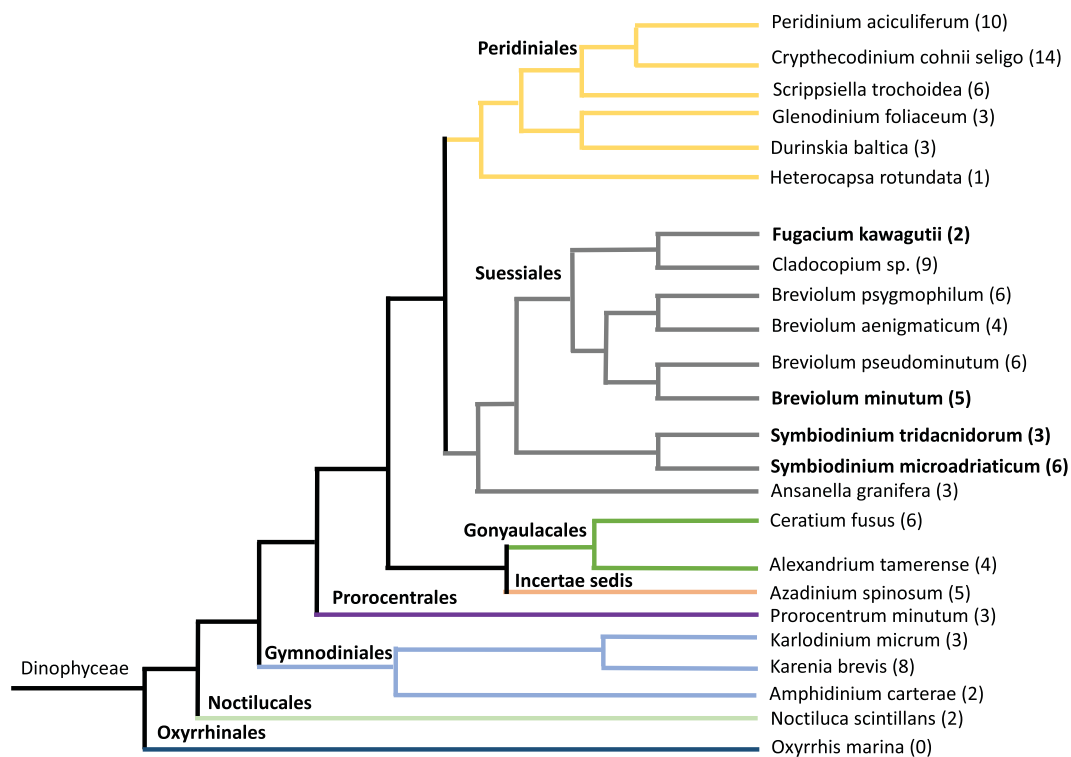
Fig. 2.—Dinoflagellate species tree with number of GIMAPs obtained through BLAST searches. Relationships between species used in this study was based off of previous dinoflagellate phylogenies (Janouškovec et al. 2017; Price and Bhattacharya 2017; LaJeunesse et al. 2018). For each, the number of GIMAPs present is indicated in parentheses along with whether these data were obtained from a genome (bolded) or transcriptome (nonbolded).

S-GIMAPs (13 out of 46), 25% of GIMAP-like proteins (3 out of 12), and 85% (41 out of 48) of L-GIMAPs contained coiled-coils. For the majority of these sequences, the probability value was in the range of 0.8–1, indicating strong evidence for the presence of this structural feature (supplementary file S1, Supplementary Material online). The majority of the L-GIMAP sequences that do not have the coiled-coil region are below 400aa (with the exception of Plo_GIMAP2 and Rt_GIMAP2) indicating that the genome assemblies may be incorrect and/or the part of the sequence containing the coiled coil may be missing.

GIMAP-Like Sequences Are Present in Most Dinoflagellate Species

GIMAP-like sequences were identified in 23 Dinophyceae species, which included all organisms searched, except *Oxyrrhis marina* (fig. 2). Four of the datasets searched were genomes, all within the family Symbiodiniaceae, including *Breviolum minutum*, *Fugacium kawagutii*, *Symbiodinium microadriaticum*, and *Symbiodinium tridacnidorum*. The BLAST searches revealed that these genomes contained multiple GIMAPs, specifically five for *B. minutum*, two for *F. kawagutii*, six for *S. microadriaticum*, and three in *S. tridacnidorum*. The remaining dinoflagellate species searched only had transcriptomes available and most contained a similar number of GIMAPs to

those found in the Symbiodiniaceae genomes, ranging from 0 in *O. marina* to 14 in *Crypthecodinium cohnii seligo* (fig. 2).

When assessed for the coiled-coil motif, a similar percentage of sequences contained this region compared with what was found for the cnidarian sequences. Out of the 114 dinoflagellate sequences, 69 (61%) had a probability of 0.5 or greater of containing a coiled-coil domain (supplementary file S1, Supplementary Material online). Most of these were close to one, indicating high confidence and many sequences contained more than one region with a probable coiled coil.

Dinoflagellate GIMAP Proteins Separate into Seven Distinct Clades on the Phylogenetic Tree

The results of Bayesian analysis indicate that there are several distinct clades of dinoflagellate GIMAP sequences, similar to what was found for cnidarians. Specifically, the tree contained seven well-supported groups, with additional sequences either on their own, in small clades (defined as less than six sequences), or in clades that had low support (fig. 3). The largest of these groups, termed A, contains all sequences from thecates, including all Symbiodiniaceae genera used in this study, with the exception of one *Karenia brevis* sequence. Group B is quite different in its composition in that it contains sequences mostly from more basal species, but lacks any from Gonyaulacales, Suessiales, and Peridiniales. Several of the
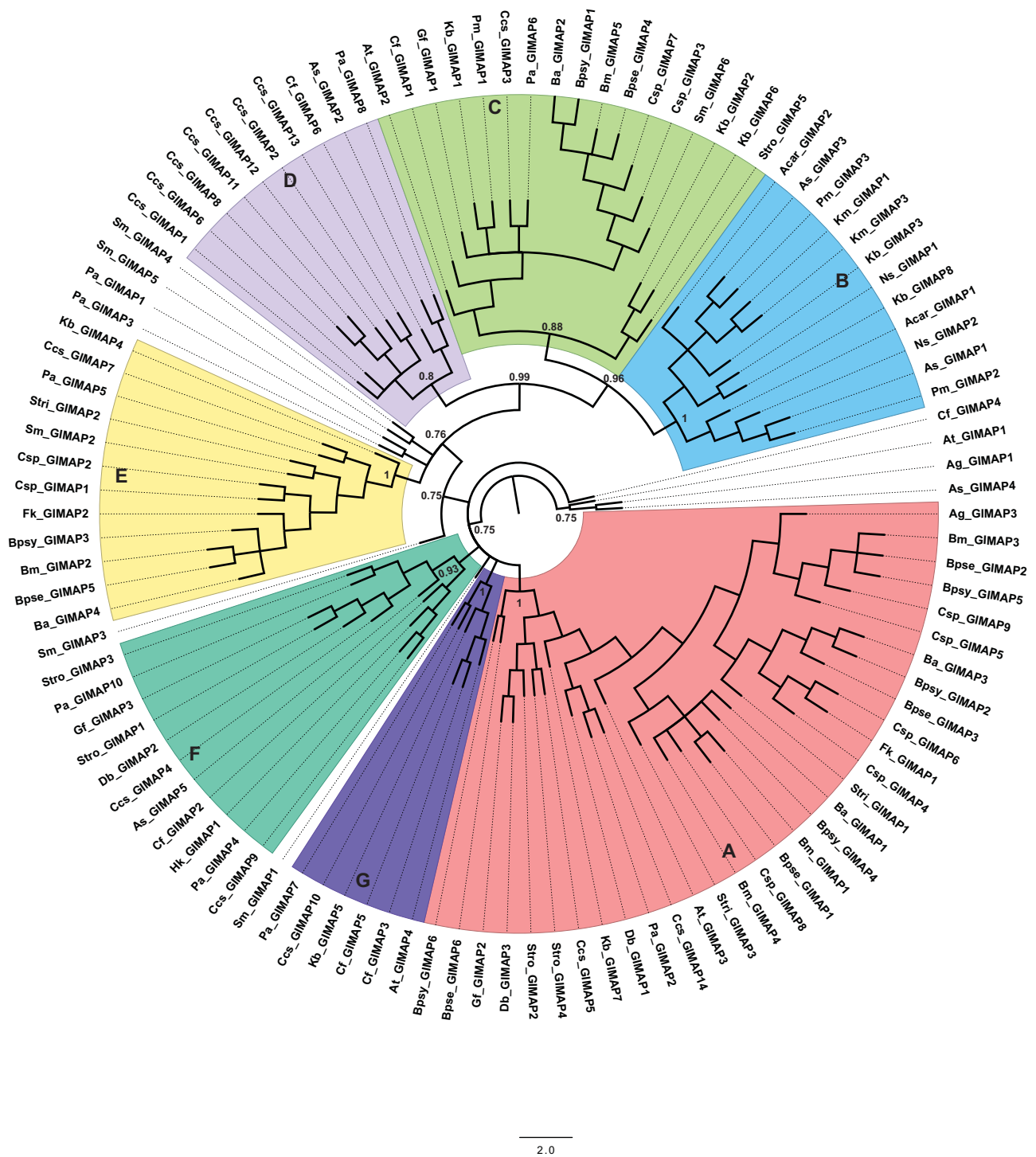
Fig. 3.—Bayesian analysis of dinoflagellate GIMAP proteins. Phylogenetic analysis was performed with Mr Bayes and the seven major groups of sequences (A–G) are highlighted with unique colors. Posterior probabilities for major nodes on the tree are indicated.

more basal dinoflagellate species used, including *Noctiluca scintillans* and *Karlodinium micrum* only have sequences within this clade of the tree. Groups C and E are similar to A in that they contain all thecate sequences with the

exception of one from *K. brevis*. Group D includes a lineage-specific expansion of *C. cohnii seligo* sequences with one sequence each from three other species. Group F contains almost all sequences from the Peridiniales species, with

one sequence each from *Azadinium spinosum* and *Ceratium fusus*. Lastly, Group G lacks Symbiodiniaceae sequences, but contains species from groups across the dinoflagellate phylogeny.

Similar to the patchy distribution observed in cnidarians, the Symbiodiniaceae genomes searched in this study revealed variation in the number and types of GIMAPs present. For example, group A contains a sequence from *S. tridacnidorum*, but not *S. microadriaticum*, group C lacks sequences from *F. kawagutii* and *S. tridacnidorum*, and lastly, *S. microadriaticum* contains several sequences that are distinct from all others on the tree, including Sm_GIMAP1, 3, 4, and 5. Therefore, variation in the number and type of GIMAP can also exist at the genus level for dinoflagellates.

The presence of the coiled-coil domain was also correlated with dinoflagellate GIMAP type, but this relationship was not as strong as for cnidarians (Cramer's V, 0.40). For group A, 30 out of 32 (94%) contained the coiled-coil domain, and the two sequences that did not (Ba_GIMAP3 and Bpse_GIMAP6) were significantly shorter than the others suggesting they may be incomplete. However, the distribution of the coiled-coil motif was patchy within the other groups. Specifically, in group B, 9 out of 12 (75%) sequences contained the coiled-coil domain; in group C, 13 out of 17 sequences (76%); group D, 5 out of 10 (50%); group E, 1 out of 11 (9%); group F, 4 out of 11 (36%); and group G, 2 out of 6 (33%).

Multiphyla Tree Indicates Cnidarian and Dinoflagellate Sequences Are Largely Distinct from Previously Characterized GIMAPs

Phylogenetic analysis of cnidarian and dinoflagellate GIMAP sequences with other previously characterized GIMAPs from vertebrates, invertebrates, plants, and protists revealed several interesting findings. First, for the cnidarian sequences, the representative L-GIMAP, S-GIMAP, and GIMAP-like sequences were recovered in the same groups as shown in the cnidarian tree, providing further evidence for these three distinct types (fig. 4). Each of these groups fell in a distinct location in the tree, with the S-GIMAP as an independent clade within a large group that was mostly unresolved. The GIMAP-like sequence clade was also within this same large group, and interestingly grouped with a sequence from the dinoflagellate *Alexandrium tamerense* with strong support. Lastly, the L-GIMAPs were located sister to a group of sequences from the cephalochordate *Branchiostoma floridae*, though with low support, indicating minimal confidence in this grouping. Overall, the only cnidarian GIMAP group that showed similarity to other animal sequences was the L-GIMAPs.

All of the major groups identified on the dinoflagellate tree were also recovered, but the relationships between these groups were slightly different (fig. 4). For example, in the dinoflagellate tree, B and C grouped together with strong support. However, in the multiphyla tree, group C was more

closely related to group E, though groups B, C, D, and E were still in a larger clade together with moderate support in both trees. There were also interesting results recovered for the relationship between other protist and dinoflagellate sequences. Most notable was that a sequence from the haptophyte *Chrysochromulina* was located in a clade with sequences belonging to dinoflagellate group C indicating that they likely share an evolutionary origin. Additionally, previously characterized *E. histolytica* sequences fell within a clade with group F dinoflagellate sequences with moderate support. However, aside from groups C and F, the other dinoflagellate groups showed no clear relationship to known GIMAP proteins.

Discussion

A Patchy Distribution of GIMAPs Is Not Unique to Cnidarians

The results of this study demonstrated that GIMAPs have a patchy distribution in cnidarians, where hydrozoans, scyphozoans, and cubozoans lack GIMAPs, actinarians have S-GIMAPs, corallimorpharians have L-GIMAPs, and scleractinians show the most variation with one or more of L-GIMAPs, S-GIMAPs, and GIMAP-like proteins (table 1). This patchy distribution is not surprising given the overall distribution of GIMAPs in metazoans, which is marked by independent loss-events and lineage-specific expansions (Weiss et al. 2013; McDowell et al. 2016). Previous studies provide evidence for the presence of GIMAPs in corals, molluscs, cephalochordates, and hemichordates, but they have not been detected in sponges, placozoans, urochordates, echinoderms, and the model invertebrates *Caenorhabditis elegans* and *Drosophila melanogaster* (Weiss et al. 2013; McDowell et al. 2016). A patchy distribution of GIMAPs at the phylum level is also not unique to cnidarians. In a comprehensive study of AIG1 proteins in molluscs, the number of GIMAPs (defined by AIG+coiled-coil) varied greatly across species, ranging from 0 in two cephalopod species to 64 in the snail *Biomphalaria glabrata* (Lu et al. 2020). Likewise, within vertebrates, a recent study showed that within the Aves/reptile and mammalian groups, there were species that lacked GIMAP genes (Balla et al. 2020). Together, these data demonstrate that GIMAP gene evolution across metazoans is complex and future research will increase our understanding of the functional consequence of this patchy distribution.

Distribution of Cnidarian GIMAPs Was Not Related to Symbiotic Status

The patchy distribution observed in this study did not support the original hypothesis that the distribution of GIMAPs in cnidarians could be explained by variation in symbiotic status. Within the dataset, there were examples of both symbiotic and nonsymbiotic species that contained and lacked all three
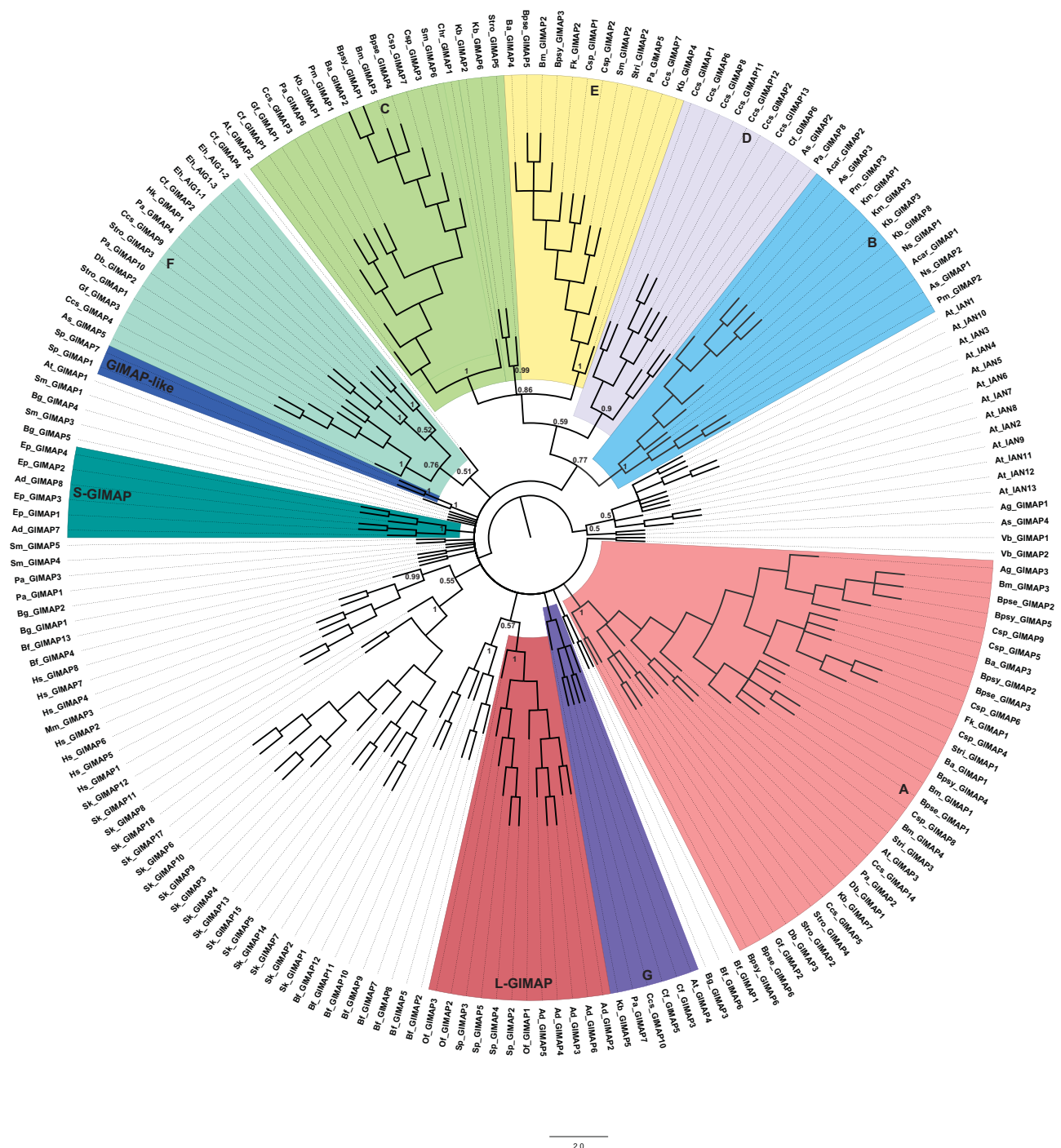
Fig. 4.—Bayesian analysis of cnidarian and dinoflagellate sequences with other known GIMAPs. Phylogenetic analysis was performed with Mr Bayes and the major groupings of cnidarian and dinoflagellate GIMAPs are indicated with the same colors used in previous figures. Posterior probabilities for major nodes on the tree are indicated.

GIMAP types (table 1 and supplementary table S1, Supplementary Material online). However, the general lack of GIMAPs within medusozoans matches results previously observed for immunity related genes. Specifically, medusozoans were found to lack prototypical pattern recognition

receptors including Nod-like receptors (NLRs), Toll-like receptors (TLRs), and RIG-like receptors that were present in most anthozoan species and had less complete NF-kB and complement pathways (Emery et al. 2021). For TLRs, it was hypothesized that this increased complexity in anthozoans,

particularly scleractinians, could be attributed to regulating the microbiome composition (Poole and Weis 2014; Emery et al. 2021). Similarly, increased NLR diversity was proposed to increase specificity in the immune response (Emery et al. 2021). Therefore, since current evidence suggests cnidarian GIMAPs play a role in immunity and the related processes of apoptosis and autophagy, it can be hypothesized their presence could also contribute to a greater diversity and specificity in the immune response.

Evolutionary Origin of Cnidarian GIMAPs Remains Largely Unresolved by Multiphyla Tree

The multiphyla analysis did not clearly resolve the origin of each cnidarian GIMAP type, with all three falling in different locations on the tree. L-GIMAPs were located in a clade with sequences from the cephalochordate *B. floridae*, but the support value was low, providing little confidence for this grouping. However, based on BLAST searches, L-GIMAPs did show the greatest similarity to human GIMAPs out of the three types (data not shown), and this combined with their placement on the tree and the presence of the coiled-coil motif suggests they may be ancestral GIMAPs. The evolutionary origin of the S-GIMAPs and GIMAP-like proteins is less clear and will require further investigation. S-GIMAPs formed an independent clade on the multiphyla GIMAP tree and were only found in actinarians and two genera of scleractinians. Therefore, it seems likely that these proteins were subsequently lost from corallimorphs and many scleractinian lineages. Lastly, GIMAP-like proteins formed a clade with the dinoflagellate sequence At_GIMAP1 with strong support, which was an unexpected result. However, further investigation will need to be done to provide convincing evidence for their relation to dinoflagellate sequences. Further availability of sequenced genomes within this phylum will help to confirm and explain the patterns observed in this study.

Dinoflagellate GIMAPs Show Widespread Distribution and Unique Evolutionary History

The BLAST searches in dinoflagellate genomes and transcriptomes revealed that all species with the exception of *O. marina* had GIMAP proteins. This is consistent with findings of a previous study that showed a small percentage of *O. marina* transcripts matched previously sequenced dinoflagellate ESTs, suggesting this species is highly divergent (Lowe et al. 2011). However, *O. marina* only had a transcriptome resource available, and therefore it is possible that GIMAPs are present in the genome, but not expressed in the transcriptome sample. The widespread occurrence of GIMAPs in dinoflagellates also indicated that engaging in symbiotic relationships, in the capacity of host or symbiont, did not explain GIMAP distribution given that these proteins were found in both free-living and symbiotic species. As the presence of

AIG1 proteins in protists has only been previously reported in the parasite *E. histolytica*, the fact that GIMAPs are widespread within dinoflagellates is a novel finding.

It is clear from phylogenetic analysis that most dinoflagellate GIMAPs are distinct from known GIMAP proteins in animals, plants, and protists and that there are seven clades, or groups, of dinoflagellate GIMAPs. Although most of the dinoflagellate data used in this study was obtained from transcriptomes, there were trends in the data regarding which dinoflagellate species were contained in each group. For example, the diversity of GIMAPs present in basal groups of dinoflagellates (i.e., *Nocticula* and *Karlodinium*) was more limited than that of more derived groups, as they only contained sequences within group B. There were also several examples of lineage-specific expansions including Symbiodiniaceae sequences within group A and *C. cohnii seligo* within group D. Lastly, as was the case for cnidarians, there were differences in GIMAP repertoire observed at lower taxonomic levels, specifically within the genus *Symbiodinium*. *S. microadriaticum* has a greater overall number and diversity of GIMAPs than *S. tridacnidorum*, but is notably missing a sequence in group A. This difference is not entirely surprising as a recent study indicated that divergence in the genomes of different *Symbiodinium* species can be of the same magnitude as differences between genera of Symbiodiniaceae (González-Pech et al. 2021). Overall, however, without any knowledge of the function of dinoflagellate GIMAPs, it is difficult to speculate on the significance of any of these patterns. Therefore, future functional work that investigates dinoflagellate proteins from multiple groups could help to explain the results obtained in this study.

One of the most interesting findings from the multiphyla GIMAP analysis was that group C dinoflagellate sequences formed a clade with strong bootstrap support with a sequence from the haptophyte *Chrysochromulina*, indicating that they likely share an evolutionary origin. This is a surprising result given the fact that haptophytes and dinoflagellates are not closely related protist groups. Dinoflagellates are within the alveolates, which is part of the SAR supergroup, whereas haptophytes are part of Haptista, which is currently placed as a sister group to the SAR (Burki et al. 2016). The similarity between the dinoflagellate and haptophyte sequences raised the question of whether this GIMAP type is present in other protists. No evidence of other similar sequences in other protist groups was recovered (data not shown), indicating this GIMAP type is not widespread within this group. Interestingly, similar results were observed in phylogenetic analysis on xanthorhodopsins, light-driven proton pumps, in which sequences from haptophytes and dinoflagellates were positioned as sister groups (Hovde et al. 2015). Since horizontal gene transfer is common in dinoflagellates (Wisecaver and Hackett 2011), the authors proposed this as one explanation for this observation, and is one that warrants further investigation for this GIMAP type as well.

Another interesting finding from the phylogenetic tree was that group F dinoflagellate GIMAPs formed a clade with previously characterized *E. histolytica* sequences. This group only contained sequences from members of the Peridiniales along with one from *A. spinosum* and *C. fusus*, but notably lacks any from Symbiodiniaceae. Therefore, it seems unlikely that Symbiodiniaceae GIMAPs directly mirror the functions of AIG1 proteins from *E. histolytica* in cell adhesion and virulence but does not rule out the possibility that GIMAPs from other groups could mediate host-symbiont interactions.

Considerations for GIMAP Prediction and Annotation

One question that was raised during this work and in previous studies (Lu et al. 2020) was the requirements to annotate a protein as a GIMAP. Due to the fact that cnidarian and dino-flagellate sequences are highly divergent compared with vertebrate GIMAPs, there were cases in which it was difficult to decide whether a result from the BLAST searches should truly be considered a GIMAP. Additionally, another issue was whether to require a coiled-coil motif. Lu et al. (2020) used this as a requirement of a GIMAP whereas we have taken a more relaxed approach in this study for several reasons. First, the prediction programs are all based on vertebrate sequences, from which cnidarian and dinoflagellate sequences are quite divergent. Additionally, the tool utilized in this study, COILS, is based on the coiled-coil domains in non-GIMAP proteins, including myosins, tropomyosins, intermediate filaments, desmosomal proteins, and kinesins (Lupas et al. 1991) and therefore may not work as well for GIMAPs. Lastly, and potentially most important is that although GIMAPs are frequently defined in the literature as containing the AIG1 and coiled-coil domain, there is also documentation from rats and mice that some GIMAPs lack this structural feature (Nitta et al. 2006; Rutledge et al. 2009). In running the human/mouse GIMAPs used in this study through the COILS prediction tool only six out of eight vertebrate GIMAPs contained the coiled-coil (lacking in GIMAP1 and 2, data not shown). Therefore, we argue that the lack of a coiled-coil domain, as determined by predictive software, should not rule out identification as a GIMAP.

Another important point to emphasize is that the function of the coiled-coil in GIMAPs has yet to be determined and therefore, it is hard to define whether it should be a requirement for classification. Although this is an area that requires further investigation, it has been proposed that like the closely related septin and dynamin proteins, the coiled-coil motif may mediate protein interactions, specifically in the formation of homo- and heterodimers (Okamoto et al. 1999; Sheffield et al. 2003; Low and Macara 2006; Limoges et al. 2021). This function seems even more likely given the fact that vertebrate GIMAPs have been shown to form homo- and heterodimers and that dimer formation can alter their activity (Schwefel et al. 2013). Further information regarding the

function of the coiled-coil motif in GIMAPs could help to determine whether this should be required to define a protein as a GIMAP. In the context of this study, it could also help to explain the functional significance of the presence of the coiled-coil motif in only some cnidarian and dinoflagellate GIMAP groups.

Conclusions

The results of this study illustrate the complexity of GIMAP evolution within metazoans, specifically cnidarians, and document the presence of GIMAP proteins in dinoflagellates for the first time. We provide evidence for three distinct groups of GIMAPs in cnidarians that each have a patchy distribution, generating questions about the evolutionary origin of each and the occurrence of multiple loss events across many taxa. Similarly, the phylogenetic analyses of dinoflagellate GIMAPs revealed seven distinct clades. Some of these showed similarity to known GIMAPs and AIG1 proteins, whereas others formed independent groups with no clear relationship to known GIMAPs. These results pave the way for performing functional work on both cnidarian and dinoflagellate GIMAPs groups to determine whether they have retained a conserved function with plant and vertebrate GIMAPs and to explore potential functional diversification.

Materials and Methods

Identification of Cnidarian GIMAP Sequences

To identify GIMAP sequences in cnidarians, a variety of genomes and transcriptomes were searched (supplementary table S1, Supplementary Material online). For these searches, human, mouse, and previously characterized *A. millepora* sequences (Bailey et al. 2020) were used as queries for BlastP or TBlastN searches of each resource with an e-value cutoff of 1×10^{-1}. A high e-value cutoff was used to account for the high sequence divergence between human and cnidarian sequences. If a TBlastN search was performed, the resulting nucleotide sequences were translated using the ExPASy translate tool (https://web.expasy.org/translate/, last accessed March 23, 2021, Gasteiger et al. 2003). The protein sequences obtained were then run through the NCBI conserved domain database (CDD) search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi, last accessed March 23, 2021, Marchler-Bauer et al. 2017) to identify presence of the conserved AIG1 domain found in all GIMAP proteins. All sequences that contained the AIG1 domain (cd01852/pfam04548) with an e-value cutoff of 1×10^{-4} were then used in a reciprocal BlastP seach to humans (taxid:9606) in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed March 23, 2021, Madden 2003) to ensure that the top hit was a GIMAP, further confirming its identity. Any sequence that did not have a GIMAP as the top hit was

removed. Lastly, all sequences were run through the program COILS (https://embnet.vital-it.ch/software/COILS_form.html, last accessed October 1, 2020, Lupas et al. 1991) to determine whether a coiled-coil motif was present. However, sequences were not removed from the analysis if they lacked the coiled-coil motif. Specifically, the program was run using the unweighted MTIKD matrix with all window widths (14, 21, and 28). To determine the strength of the correlation between GIMAP type and the presence of the coiled-coil domain, Cramer's V test was run using R version 4.2.1 (R Core Team 2018). The resulting sequences for each species were aligned in Geneious version 11.1.3 (http://www.geneious.com, last accessed May 21, 2021, Kearse et al. 2012), and duplicate sequences (defined as proteins with fewer than 5aa differences) were removed.

Identification of Dinoflagellate GIMAP Sequences

As GIMAPs could play an important role in regulating processes involved in symbiosis, we were also interested in whether these proteins were present in dinoflagellates. Therefore, dinoflagellate genomic and transcriptomic data were obtained from the Reef Genomics website (Liew et al. 2016), NCBI (https://ncbi.nlm.nih.gov, last accessed March 23, 2021), and imicrobe (Keeling et al. 2014) (supplementary table S2, Supplementary Material online). Since a limited number of dinoflagellate genomes have been sequenced, transcriptomes were selected to represent all major groups within this phylum. Mammalian and *A. millepora* GIMAP protein sequences (Weiss et al. 2013; Bailey et al. 2020) were used as queries in a BlastP or TBlastN search against these databases as previously described for cnidarians. All sequences recovered from the BLAST searches were analyzed as described for cnidarians to confirm their identity as a GIMAP, run through the COILS program, and to detect duplicate sequences in the dataset.

Sequence Alignment and Phylogenetic Tree Construction

With the data collected from BLAST searches, three phylogenetic analyses were performed. First, analysis of cnidarian and dinoflagellate sequences each was performed separately and then subsequently all sequences were combined with known GIMAP sequences from other organisms for a multiphyla analysis. For the cnidarian analysis, upon completing the BLAST searches, it became clear that there were at least two distinct protein types, which could be differentiated by similarity to mammalian or *A. millepora* GIMAPs and length. Therefore, phylogenetic analysis was used to better understand the different types of cnidarian GIMAPs and how these sequences were related to one another. To do so, all cnidarian GIMAP sequences were aligned using the MAFFT plugin in Geneious version 11.1.3 (http://www.geneious.com, last accessed May 21, 2021, Kearse et al. 2012) with the default

settings. The alignment was then manually trimmed to remove any regions that did not align well and sequences that did not span more than half the alignment length were removed. For the dinoflagellate analysis, all sequences obtained from the BLAST searches were aligned and trimmed as described above for cnidarian sequences.

Lastly, we were interested in how the newly characterized cnidarian and dinoflagellate GIMAPs were related to known GIMAP or AIG proteins from other organisms. For animals, sequences were acquired from the mollusc *B. glabrata*, the cephalochordate *B. floridae*, the hemichordate *Saccoglossus kowalevskii*, and humans. For *B. glabrata*, previously characterized GIMAPs were used. Specifically, one GIMAP sequence (defined by an AIG1 domain and a coiled-coil motif) was selected from each of the five groups denoted in a previously constructed phylogeny (Lu et al. 2020). For the cephalochordate *B. floridae* and the hemichordate *S. kowalevskii*, sequences were acquired from a previously constructed phylogeny of GIMAPs (Weiss et al. 2013). However, the GenBank accession numbers for many of these sequences indicated the records had been removed, and therefore these sequences were updated with the closest match to current accession numbers as determined by a BlastP search. To ensure that all GIMAP sequences were captured within these two species, BlastP searches were performed through NCBI with human and *A. millepora* GIMAPs as previously described for cnidarians. It was also of interest to include plant and other protist sequences. Previously characterized *Arabidopsis thaliana* (Liu et al. 2008) and *E. histolytica* (Nakada-Tsukui et al. 2018) sequences were obtained from NCBI. Additionally, sequences from two other protists, *Chrysochromulina* sp. and *Vitrella brassicaformis* were obtained when initially exploring and performing BLAST searches in NCBI with Symbiodiniaceae sequences. All sequences used in this analysis and corresponding accession numbers are presented in supplementary file S1, Supplementary Material online. These sequences were aligned with dinoflagellate and representative cnidarian sequences of the S-GIMAP, L-GIMAP, and GIMAP-like sequences using the MAFFT plugin in Geneious version 11.1.3 (http://www.geneious.com, last accessed May 21, 2021, Kearse et al. 2012) with the default settings. The alignment was then trimmed as previously described.

For phylogenetic analysis, either maximum likelihood or Bayesian analysis was performed. For the cnidarian analysis, the trimmed alignment was submitted to PhyML version 3.0, and a maximum likelihood analysis with 500 bootstrap replicates was conducted (http://www.atgc-montpellier.fr/phyml/, last accessed June 2, 2021, Guindon et al. 2010). The PhyML SMS automatic model selection was used (Lefort et al. 2017), with the AIC criterion, and consequently the WAG + G + I + F model was selected. For the dinoflagellate and multiphyla analyses, Bayesian analysis was performed due to poor resolution obtained by initial runs of PhyML (data not shown). The SMS model selection was used as described above to

determine the best model of protein evolution, which was LG + G + I for both the dinoflagellate and multiphyla trees. Mr Bayes (Huelsenbeck and Ronquist 2001) was subsequently run using the default settings in the Geneious plugin, which included 1,100,000 generations, a subsampling frequency of 200, a burn-in length of 100,000, and the model provided by SMS. All trees were imported into Mega11 (Tamura et al. 2021) to collapse nodes with less than 50% support followed by subsequent use of FigTree v1.4.4 for annotation (http://tree.bio.ed.ac.uk/software/figtree/, last accessed November 10, 2021).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Author Contributions

This project idea was conceived by J.C.C. and A.Z.P. BLAST and CDD searches, sequence alignments, and phylogenetic analysis were performed by J.C.C. and A.Z.P. COILS analysis was run by E.D.C. and G.N.C. The manuscript was written by J.C.C. and A.Z.P. and approved by all authors prior to submission.

Data Availability

The accession numbers and resources for all sequence data use in the analyses in this manuscript are available in supplementary file S1 and tables S1 and S2, Supplementary Material online. All sequence alignments used for phylogenetic analysis are included in the supplementary materials (supplementary figs. S1–S3, Supplementary Material online).

Literature Cited

Bailey GF, Coelho JC, Poole AZ. 2020. Differential expression of *Exaiptasia pallida* GIMAP genes upon induction of apoptosis and autophagy suggests a potential role in cnidarian symbiosis and disease. J Exp Biol. 223:jeb229906.

Balla KM, Rice MC, Gagnon JA, Elde NC. 2020. Linking virus discovery to immune responses visualized during zebrafish infections. Curr Biol. 30(11):2092–2103.e5.

Burki F, et al. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc R Soc B. 283(1823):20152802.

Cambot M, Aresta S, Kahn-Perlès B, de Gunzburg J, Roméo P-H. 2002. Human immune associated nucleotide 1: a member of a new guanosine triphosphatase family expressed in resting T and B cells. Blood 99(9):3293–3301.

Ciucci T, Bosselut R. 2014. GIMAP and T cells: a matter of life or death. Eur J Immunol. 44(2):348–351.

Dunn SR, Schnitzler CE, Weis VM. 2007. Apoptosis and autophagy as mechanisms of dinoflagellate symbiont release during cnidarian bleaching: every which way you lose. Proc Biol Sci. 274(1629):3079–3085.

Emery MA, Dimos BA, Mydlarz LD. 2021. Cnidarian pattern recognition receptor repertoires reflect both phylogeny and life history traits. Front Immunol. 12:689463.

Fuess LE, Pinzón CJ, Weil E, Grinshpon RD, Mydlarz LD. 2017. Life or death: disease-tolerant coral species activate autophagy following immune challenge. Proc Biol Sci. 284:20170771.

Gasteiger E, et al. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31(13):3784–3788.

González-Pech RA, et al. 2021. Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus *Symbiodinium*. BMC Biol. 19(1):73.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.

Hanes SD, Kempf SC. 2013. Host autophagic degradation and associated symbiont loss in response to heat stress in the symbiotic anemone, *Aiptasia pallida*. Invertebr Biol. 132(2):95–107.

Hovde BT, et al. 2015. Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent order Prymnesiales (Haptophyceae). PLoS Genet. 11(9):e1005469.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17(8):754–755.

Hughes TP, et al. 2018. Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. Science 359(6371):80–83.

Janouškovec J, et al. 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. Proc Natl Acad Sci U S A. 114(2):E171–E180.

Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.

Keeling PJ, et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12(6):e1001889.

LaJeunesse TC, et al. 2018. Systematic revision of symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. Curr Biol. 28(16):2570–2580.e6.

Lefort V, Longueville J-E, Gascuel O. 2017. SMS: smart model selection in PhyML. Mol Biol Evol. 34(9):2422–2424.

Liew YJ, Aranda M, Voolstra CR. 2016. Reefgenomics.Org – a repository for marine genomics data. Database (Oxford) 2016:baw152.

Limoges M-A, Cloutier M, Nandi M, Ilangumaran S, Ramanathan S. 2021. The GIMAP family proteins: an incomplete puzzle. Front Immunol. 12:679739.

Liu C, Wang T, Zhang W, Li X. 2008. Computational identification and analysis of immune-associated nucleotide gene family in *Arabidopsis thaliana*. J Plant Physiol. 165(7):777–787.

Low C, Macara IG. 2006. Structural analysis of septin 2, 6, and 7 complexes. J Biol Chem. 281(41):30697–30706.

Lowe CD, et al. 2011. The transcriptome of the novel dinoflagellate *Oxyrrhis marina* (Alveolata: Dinophyceae): response to salinity examined by 454 sequencing. BMC Genomics 12:519.

Lu L, Loker ES, Zhang S-M, Buddenborg SK, Bu L. 2020. Genome-wide discovery, and computational and transcriptional characterization of an AIG gene family in the freshwater snail *Biomphalaria glabrata*, a vector for *Schistosoma mansoni*. BMC Genomics 21(1):190.

Lupas A, Dyke MV, Stock J. 1991. Predicting coiled coils from protein sequences. Science 252(5009):1162–1164.

Madden T. 2003. The BLAST sequence analysis tool. National Center for Biotechnology Information (US). Available from: https://www.ncbi.nlm.nih.gov/books/NBK21097/; Date accessed January 23, 2020.

Marchler-Bauer A, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45(D1):D200–D203.

McDowell IC, Modak TH, Lane CE, Gomez-Chiarri M. 2016. Multi-species protein similarity clustering reveals novel expanded immune gene families in the eastern oyster *Crassostrea virginica*. Fish Shellfish Immunol. 53:13–23.

Nakada-Tsukui K, et al. 2018. AIG1 affects in vitro and in vivo virulence in clinical isolates of *Entamoeba histolytica*. PLoS Pathog. 14(3):e1006882.

Nitta T, et al. 2006. IAN family critically regulates survival and development of T lymphocytes. PLoS Biol. 4(4):e103.

Okamoto PM, Tripet B, Litowski J, Hodges RS, Vallee RB. 1999. Multiple distinct coiled-coils are involved in dynamin self-assembly. J Biol Chem. 274(15):10277–10286.

Pascall JC, et al. 2013. The immune system GTPase GIMAP6 interacts with the Atg8 homologue GABARAPL2 and is recruited to autophagosomes. PLoS One 8(10):e77782.

Pascall JC, et al. 2018. GIMAP6 is required for T cell maintenance and efficient autophagy in mice. PLoS One 13(5):e0196504.

Poole AZ, Weis VM. 2014. TIR-domain-containing protein repertoire of nine anthozoan species reveals coral-specific expansions and uncharacterized proteins. Dev Comp Immunol. 46(2):480–488.

Precht WF, Gintert BE, Robbart ML, Fura R, van Woesik R. 2016. Unprecedented disease-related coral mortality in southeastern Florida. Sci Rep. 6:31374.

Price DC, Bhattacharya D. 2017. Robust Dinoflagellata phylogeny inferred from public transcriptome databases. J Phycol. 53(3):725–729.

R Core Team. 2018. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Rutledge EA, et al. 2009. Sequence variation and expression of the GIMAP gene family in the BB rat. Exp Diabetes Res. 2009:835650.

Schwefel D, et al. 2010. Structural basis of oligomerization in septin-like GTPase of immunity-associated protein 2 (GIMAP2). Proc Natl Acad Sci U S A. 107(47):20299–20304.

Schwefel D, et al. 2013. Structural insights into the mechanism of GTPase activation in the GIMAP family. Structure 21(4):550–559.

Sheffield PJ, et al. 2003. Borg/septin interactions and the assembly of mammalian septin heterodimers, trimers, and filaments. J Biol Chem. 278(5):3483–3488.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis, version 11. Mol Biol Evol. 38(7):3022–3027.

Weiss Y, et al. 2013. The acute transcriptional response of the coral *Acropora millepora* to immune challenge: expression of GiMAP/IAN genes links the innate immune responses of corals with those of mammals and plants. BMC Genomics 14:400.

Wisecaver JH, Hackett JD. 2011. Dinoflagellate genome evolution. Annu Rev Microbiol. 65:369–387.

Associate editor: Sujal Phadke