

Current Biology

The integrin-mediated adhesive complex in the ancestor of animals, fungi, and amoebae

Highlights

- Integrins alpha and beta are present throughout the Amoebozoa supergroup
- Amoebozoan integrins are highly divergent, many with interesting catalytic domains
- The origin of the integrin-mediated adhesion complex predates Amorphea
- Differential loss of these protein complexes appears to be rampant among Amoebozoa

Authors

Seungho Kang, Alexander K. Tice, Courtney W. Stairs, Robert E. Jones, Daniel J.G. Lahr, Matthew W. Brown

Correspondence

matthew.brown@msstate.edu

In brief

The integrin-mediated adhesive complex mediates a variety of cellular processes in multicellular eukaryotes. Here, Kang et al. describe a wide variety of integrin homologs in Amoebozoa. These findings push back the evolutionary history of these important cell adhesion protein complexes, predating the origin of Obazoa.integrin, Amoebozoa, protein domain architecture, amoeba, reductive evolution



Article

The integrin-mediated adhesive complex in the ancestor of animals, fungi, and amoebae

Seungho Kang,^{1,2,7} Alexander K. Tice,^{1,2,7} Courtney W. Stairs,^{3,4} Robert E. Jones,^{1,2} Daniel J.G. Lahr,⁵ and Matthew W. Brown^{1,2,6,8,*}

¹Department of Biological Sciences, Mississippi State University, Starkville, MS, USA

²Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Starkville, MS, USA

³Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

⁴Department of Biology, Lund University, Lund, Sweden

⁵Department of Zoology, University of São Paulo, São Paulo, Brazil

⁶Twitter: @socialprotist

⁷These authors contributed equally

⁸Lead contact

*Correspondence: matthew.brown@msstate.edu

<https://doi.org/10.1016/j.cub.2021.04.076>

SUMMARY

Integrins are transmembrane receptors that activate signal transduction pathways upon extracellular matrix binding. The integrin-mediated adhesive complex (IMAC) mediates various cell physiological processes. Although the IMAC was thought to be specific to animals, in the past ten years these complexes were discovered in other lineages of Obazoa, the group containing animals, fungi, and several microbial eukaryotes. Very recently, many genomes and transcriptomes from Amoebozoa (the eukaryotic supergroup sister to Obazoa), other obazoans, orphan protist lineages, and the eukaryotes' closest prokaryotic relatives, have become available. To increase the resolution of where and when IMAC proteins exist and have emerged, we surveyed these newly available genomes and transcriptomes for the presence of IMAC proteins. Our results highlight that many of these proteins appear to have evolved earlier in eukaryote evolution than previously thought and that co-option of this apparently ancient protein complex was key to the emergence of animal-type multicellularity. The role of the IMACs in amoebozoans is unknown, but they play critical adhesive roles in at least some unicellular organisms.

INTRODUCTION

Integrins are transmembrane signaling and adhesive heterodimers, made up of α -integrin (ITGA) and β -integrin (ITGB) protein subunits.¹ The overall architecture of both integrin proteins includes a large extracellular N-terminal ligand-binding head domain and C-terminal stalk (also termed “legs”). Integrin activation is dependent on the binding of a ligand such as extracellular matrix proteins (collagen, fibronectin, and laminin)² and the binding of divalent cations (primarily Ca^{2+} , Mg^{2+} , or Mn^{2+}) on unique, but well-conserved, cation-binding motifs.^{3–6} In animals, there are universally five cation-binding sites on the ITGA and three on the ITGB paralogs.^{3,4,6,7} Integrins are important signal transduction molecules across the cell membrane and associate with a set of intracellular proteins that act on the actin cytoskeleton, together known as the integrin-mediated adhesive complex (IMAC) (Figure 1C; Figure S1C).⁸ The intracellular component of the IMAC includes adhesive proteins that are (1) integrin-bound proteins that can bind to actin directly (talin, α -actinin, and filamin); (2) integrin-bound proteins that can bind indirectly to the cytoskeleton (integrin-linked kinase [ILK], particularly interesting new cysteine-histidine-rich protein [PINCH], and paxillin); and (3) non-integrin-binding proteins such as vinculin. The IMAC is best known for its critical role in cellular

development, migration, proliferation, survival, and metabolism.^{1,9,10} They function by relaying messages through outside/in and inside/out signaling between the extracellular matrix and the internal cytoskeleton.²

Historically, the absence of integrin proteins in other complex multicellular lineages of eukaryotes (i.e., plants and fungi) and the closest protistan relatives of Metazoa (i.e., choanoflagellates) led to the hypothesis that integrins were an animal-specific evolutionary innovation. However, over the past decade, investigations into the genomic content of other close protistan relatives of animals have led to the discovery of integrins and other IMAC components outside of Metazoa among other members of the Obazoa clade (Opisthokonta, Breviatea, and Apusomonadida¹¹), having been identified in the opisthokonts *Capsaspora owczarzaki*,¹² *Pigoraptor* spp.,¹³ *Syssomonas multififormis*,¹³ *Corallochytrium limacisporum*,¹⁴ *Creolimax fragrantissima*,¹⁵ and *Sphaeroforma arctica*;¹⁶ the apusomonad *Thecamonas trahens*;¹² and the breviate *Pygsuia biforma*.¹¹ Thus, the IMAC complex and other associated proteins have a more ancient evolutionary origin than previously expected, well outside of the animals. However, are these complexes exclusive to Obazoa, or have they appeared even deeper in evolutionary history?

Amoebozoa is the eukaryotic supergroup that is sister to Obazoa, altogether grouped in the Amorphea.¹⁷ Despite the relatively



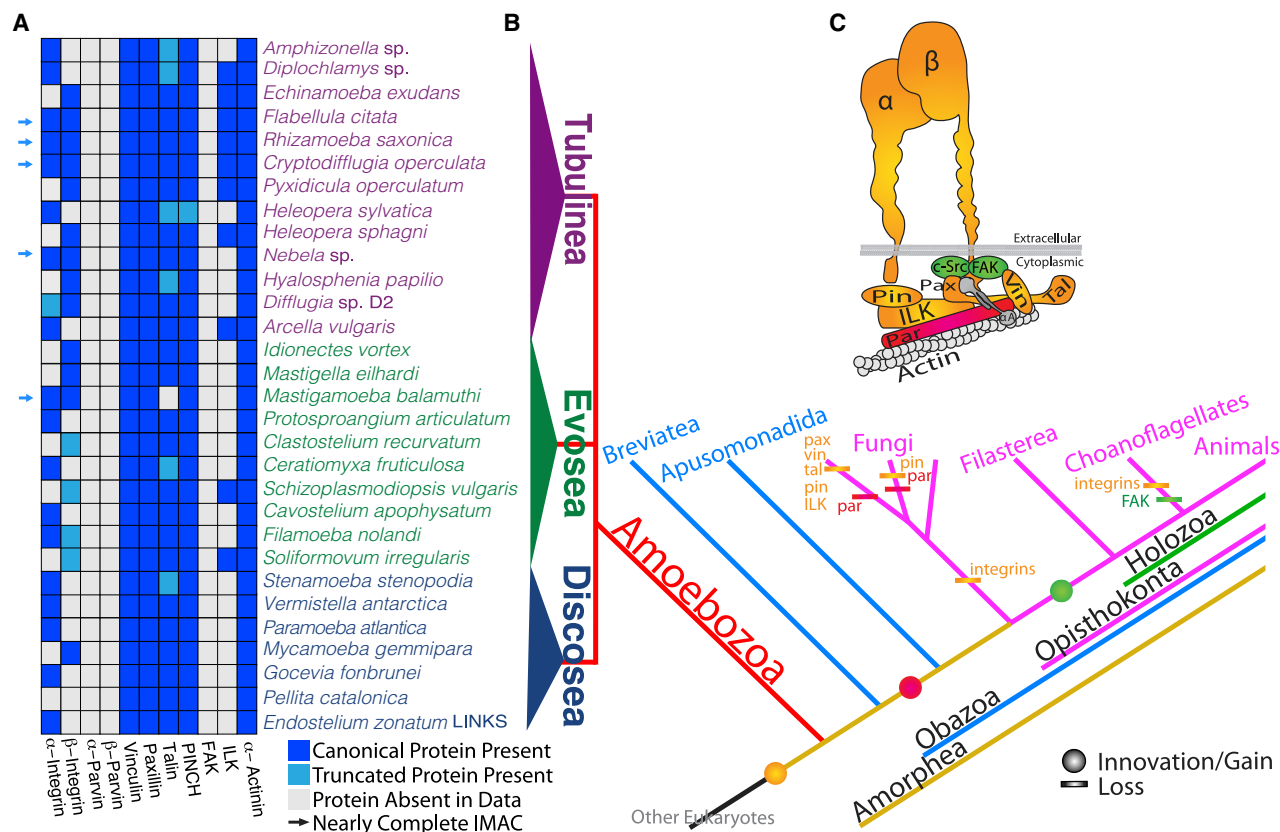


Figure 1. Repertoire of IMAC in amoebozoan species that have either integrin alpha or beta

(A) Names of IMAC proteins are listed below the map, and names of amoebozoan taxa are listed on side of the map. Arrows indicates amoebozoan species that have a nearly complete set of IMAC proteins. Dark blue indicates the presence of a canonical IMAC protein, white indicates the absence of an IMAC protein, and light blue indicates a truncated form of an IMAC protein.

(B) Phylogenetic distribution of IMAC. The cartoon model of IMAC color corresponds with the species tree. Circle, innovation/gain; bar, loss. Abbreviations are as follows: vin, vinculin; pax, paxillin; tal, talin; par, parvin; pin, particularly interesting new cysteine-histidine-rich protein; FAK, focal adhesive kinase; ILK, integrin-linked kinase; α A, alpha-actinin; α , α -integrin; and β , β -integrin.

(C) Cartoon schematic of the membrane docked IMAC, which associates with intracellular actin (modified from Brown et al.¹¹).

See also Figure S1.

close evolutionary proximity of Amoebozoa to the animals and fungi, genomic-level data and relevant research into the supergroup remain sparse. The inferred genomic complement present has yet to be determined in the last common ancestor of Amoebozoa using broad taxonomic sampling. Recent work revealed that crucial components of the IMAC—namely ITGA and ITGB—were missing in Amoebozoa (e.g., *Dictyostelium discoideum* and *Acanthamoeba castellanii*).¹⁸ Was this result simply due to a lack of data, or is it an evolutionary trend? Here, we examine the gene repertoires across the Amoebozoa supergroup to investigate the evolutionary history of the adhesive complex associated with the integrin proteins. The absence of the integrin proteins reported previously may simply be due to the great paucity of taxonomically broad genomic data from this clade, which is a limiting factor in our knowledge of IMAC evolution.

To test whether the IMAC is truly absent in Amoebozoa or rather unobserved due to a lack of data from a broad taxonomic sampling, we took a comparative genomic/transcriptomic approach utilizing data recently used to resolve the tree of Amoebozoa.¹⁹ Here, we identified numerous IMAC proteins—including

integrins, both ITGA and ITGB—across the breadth of Amoebozoa and obazoans (*Ministeria vibrans* [Opisthokonta¹⁷] and *Lenisia limosa* [Breviatea²⁰]) (Figure 1B). These findings suggest that the IMAC was already present in the last common ancestor of Amorphea (Figures 1A and 1B).

RESULTS

Integrin alpha

We identified ITGA homologs in 23 of 113 amoebozoan genomes and transcriptomes examined (Figures 1A and 2). Of those 23 amoebozoans, eight have more than one ITGA paralog (Figure 2). Amoebozoan ITGAs vary greatly in size, ranging from ~500 to 1,200 amino acids in length (Table S1), whereas metazoan ITGA proteins are typically 700–800 amino acids in length. Metazoan ITGAs, which we refer to herein as “canonical ITGA proteins,” encode a β -propeller made of seven blades (IPR013519)⁵ (InterProScan version 5.27.66;²¹ see STAR Methods). Many of the amoebozoan ITGA proteins we identified are predicted to have a diverse number of β -propellers blades



Figure 2. Domain architecture of ITGA from Amoebozoa

The domain architecture of ITGA for amoebozoan species is shown at the right side and phylogenetic tree is shown at the left side of the figure. Keys to each domain's color and shape are represented at the bottom of the figure. Phylogenetic tree of amoebozoan ITGA, 160 sites phylogeny of amoebozoan ITGA rooted with Obazoa ITGA. The tree was built using IQ-TREE v.1.5.5 under the LG+C60+F+G model of protein evolution. Numbers at nodes are maximum likelihood bootstrap (MLBS) values derived from 1,000 ultrafast MLBS replicates. Any values that are below 50% are not shown. The color of branches for amoebozoan sequences are as used in Figure 3 of Kang et al.¹⁹ When domains are overlapped because of different software algorithms, these are represented in transparent colors. See also Figures S4 and S5 and Tables S1, S2, S4, and S5.

from a few (3–5) to canonical (6–7) to many (more than 8) and cation-binding sites (1–5 sites). The classification of an authentic ITGA protein outside Metazoa would be a membrane-docked protein with an ITGA head composed of several β -propeller blades (IPR013519) (Figure 2), some with FG-GAP domains (IPR013517) only, and some with a cation-binding motif flanked by FG and GAP motifs (Figure 3A; Figure S2Aa; a canonical set of motifs found in Metazoa are shown for reference in Figure S3). However, we do not discard the possibility that these represent truncated ITGA proteins because they occur in metazoans^{22,23} and in the opisthokont *Creolimax fragrantissima*.¹⁵ All amoebozoan ITGA proteins have at least three β -propeller blade domains (Figure 2), with an FG-GAP domain and FG-GAP/cation-binding motif (Figure 3A). The positions of the cation-binding motif/FG-GAP domains are seemingly distributed randomly (Table S1)

and not necessarily in the last 3–4 blades as in canonical animal ITGA. We noticed the presence of a C-terminal transmembrane anchor (IPR021157) domain in three amoebozoan ITGAs. In animal ITGA proteins, there is also a GFFKR motif in the cytoplasmic tail that is important for ITGA and ITGB intercellular interactions and heterodimerization.²⁴ We identified an animal-like KXGFFXKR cytoplasmic tail motif in one amoebozoan (*Filamoeba nolandii*) (Table S1), which is also present in some non-metazoan obazoans (Figure S2Ac).

Phylogenetically, we find amoebozoan ITGAs are present in all three major clades, but interestingly these proteins are not in many amoebozoan taxa that have sequenced genomes (e.g., several dictyostelid amoebae,²⁵ *Physarum polycephalum*,²⁶ *Entamoeba* spp.,²⁷ *Acanthamoeba* spp.,²⁸ and *Protostelium fungivorum*²⁹). However, very importantly, we were able to

ITGA Motifs

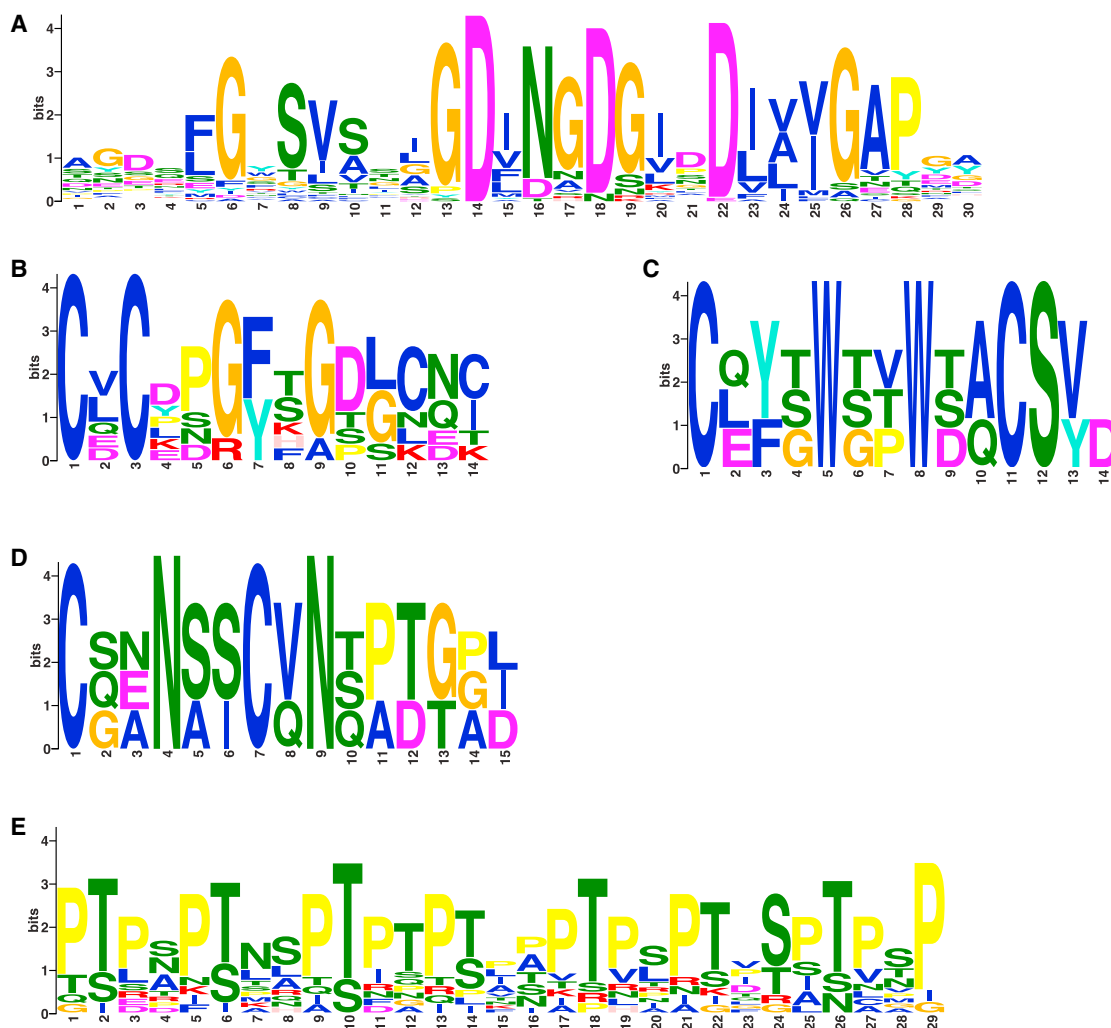


Figure 3. Most significantly enriched motifs in 23 amoebozoan ITGA homologs

(A) Consensus cation motif and FG-GAP motifs found from ITGA homologs in 23 amoebozoan taxa.

(B–D) These consensus motifs are exclusively enriched in three amoebozoans type II ITGA homologs (*Arcella vulgaris* c23786_g1_i1, *Nolandella* sp. AFSM_TR12097_c0_g1_i1, *Cryptodiffugia operculatum* paralog2_c8521_g1_i1): (B) EGF-like motif, (C) Thrombospondin repeat-1 motif, and (D) carbohydrate WSC motif.

(E) Proline-rich extensin motif found in 5 of the amoebozoan ITGA homologs.

See also [Figures S2](#) and [S3](#) and [Tables S1](#), [S2](#), [S4](#), and [S5](#).

confidently find ITGA in *Mastigamoeba balamuthi* (Figure 2), which has a complete genome available.^{30,31} To confirm the presence of these genes in the genome, we used a BLAST-based homology search to identify two ITGA genes of *M. balamuthi* to the genome data (GenBank: CBKX00000000) on NCBI (see [STAR Methods](#)). For ITGA, there are no introns in either paralog 1 (GenBank: CBKX010025823.1) or paralog 2 (GenBank: CBKX010005624.1).

We classified ITGA into two representative types based on their predicted complement of IPR domains, their cation-binding motifs, their similarity to metazoan ITGA, and novel domains that are previously unobserved in ITGA proteins. For communication

purposes, based on domain and motif architecture from InterProScan outputs, we classify non-metazoan ITGAs into two representative types: ITGA that are similar to a canonical ITGA (type I) and unique domain architecture proteins that are non-canonical ITGA-like proteins (type II) (Figure 2).

From the 23 amoebozoan taxa in which we identified an ITGA, 19 taxa have type I (Figure 2). One amoebozoan (*Amphizonella* sp. 9 paralog 2) has a type I ITGA with a domain architecture and cation-binding motifs that are virtually identical to the head region of metazoan ITGA (Figure 2; Table S1). The rest of the amoebozoan type I ITGAs have canonical β -propellers, but they have varying numbers of cation-binding motifs that differ

from the canonical ITGA proteins (Table S1). The cation-binding motifs follow the consensus pattern of D/E-h-D/N-x-D/N-G-h-x-D/E (where h is hydrophobic residue and x is any residue).³ Most of the cation-binding motifs of all but four amoebozoan type I ITGAs follow the consensus sequence observed in metazoan ITGAs (Figure 3A; Table S1). For reference, we present canonical metazoan motifs in Figure S2. Thirteen of the amoebozoan type I ITGAs have both a predicted signal peptide and transmembrane region (Figure 2). The rest of the amoebozoans type I ITGAs have only a predicted signal peptide region or a transmembrane region. Amoebozoa ITGA without a transmembrane region are potentially neofunctionalized cytoplasmic integrins. In a few cases, we did not observe the signal peptide or the transmembrane region. For example, in the genome of *Mastigamoeba balamuthi*, all ITGA paralogs lack signal peptides and one lacks a transmembrane region (paralog 1). This may be due to truncation in *in silico* gene prediction or the product of shedding (evolution from membrane docked to a secretory function).³²

Of 19 type I amoebozoan ITGA homologs, four ITGAs have an extracellular immunoglobulin (Ig)-fold-like (IPR013783) or cadherin-like (IPR015919) domain next to the ITGA head (Figure 2). In metazoan ITGA, there are three ITGA leg domains (IPR013649) located next to the ITGA head. These ITGA leg domains are structurally and functionally similar to Ig-fold-like β sandwich.⁵ A cadherin-like domain (IPR015919) and an Ig-fold-like domain (IPR013783) are part of an Ig-fold-like β -sandwich protein class, which fold in a similar structure.³³ The three amoebozoan ITGAs that contain either an Ig-fold-like or a cadherin-like domain also have canonical β -propellers (6–7 blades) (Figure 2), although two of the four type I ITGA have an extra cation motif (Table S1).

We designate any ITGA homolog with either adhesive or enzymatic domains flanking the head domain that have not previously been reported in an ITGA protein as a “type II” ITGA (Figure 2). From the 23 amoebozoan taxa in which we identified an ITGA, 12 taxa have type II (Figure 2), several of which have multiple paralogs (Table S1). Of these type II amoebozoan ITGA paralogs we discovered, two have a protein kinase domain (IPR000719) (Figure 2), five amoebozoan ITGAs have a proline-rich extension motif called “extensins” (PR01217—ID derived from Prints version 42.0³⁴) (Figure 3E; Table S1).

In one type II ITGA, the amoebozoan sequence (*Flabellula citata* paralog 1) has an epidermal growth factor (EGF)-like domain (IPR000742) as its only novel domain located C-terminal to the ITGA head (Figure 2). Additionally, three amoebozoans uniquely have two EGF domains (Figure 3B): a Thrombospondin repeat-1 (TSP1) (IPR000884) (Figure 3C) and a carbohydrate-binding domain called a wall stress component (WSC) (IPR002889) (Figure 3D), all with cysteine-rich motifs located upstream (N-terminal) to the ITGA head (Figure 2). Using a standard BLAST-based approach to query the NCBI non-redundant protein (*nr*) database, we were unable to detect any other protein with significant similarity to the N-terminal domain composition of these sequences (EGF-TSP1-WSC) using an e-value cutoff $< 1e-10$. This suggests that this EGF-TSP1-WSC architecture may be unique to these amoebozoans (Table S1).

Seven of the amoebozoan ITGAs we designated as type II possess canonical (6–7 blades) β -propellers (Figure 2) and a different number of cation-binding motifs (Table S1). The other

five amoebozoans’ type II ITGAs have short (3–5 blades) β -propellers and varying numbers of cation-binding motifs (Figure 2; Table S2). Four amoebozoan ITGA homologs of this type contained both a signal peptide and a transmembrane domain (Figure 2). Eight others had either a signal peptide or a transmembrane region.

We also found ITGA homologs in several underexplored obazoan transcriptomes: the opisthokont *Ministeria vibrans* and the breviate *Lenisia limosa* (Figure 2). Obazoan ITGA proteins have at least three β -propeller blade (IPR013517) domains, with an FG-GAP domain (Figure 2; Figure S3Aa) and FG-GAP/cation-binding motif domains seemingly distributed randomly (Table S1). Animal-like KXGFFXKR cytoplasmic tail motifs were observed in most obazoan transcriptomes that were examined in this study (Figure 3B; Figure S2Ac). Among the non-animal obazoans, the seven novel obazoans’ ITGAs have an Ig-fold-like domain leg next to ITGA head (Figure 2; Figure S3Ad).

Ministeria vibrans has both type I and type II paralogs (Figure 2). In type II, paralogs 1 and 2 have a predicted long (10) ITGA β -propeller, but they do not contain transmembrane domains. The other paralog has a three β -propeller blades (one with FG-GAP/cation-binding motif), a cytosolic tail motif, and a transmembrane domain (Figure 2). In both paralogs 1 and 3, a signal peptide region was not predicted. *Ministeria vibrans* type II ITGA uniquely has an EGF-like domain (IPR000742) N-terminal to the ITGA head (Figure 2). All obazoans in our study have a signal peptide or a transmembrane region except for *Pigoraptor* spp., *Sphaeroforma arctica*, and *Syssomonas multiformis* (Figure 2), which lack both.

The unrooted tree of ITGA recovered all amoebozoan ITGAs in a clade (Figure 2), which is sister to most of Obazoa. However, a few opisthokont sequences are nested in this group. These include a paralog of *Capsaspora owczarzaki* (ITGA4: XP_004347912), both *M. vibrans* paralogs, the breviate (*Pygmaea* and *Lenisia*), and the lone *Syssomonas multiformis* ITGA homolog. The major lineages or sublineages of Amoebozoa were not recovered. Instead, amoebozoan ITGAs are divided into two clades. One group consists primarily of tubulinid amoebozoans (plus *M. vibrans*) and the other contains the rest of Amoebozoa.

It should be noted that in addition to these obazoan and amoebozoan taxa, the genome sequence of the evolutionary distant Stramenopile, a complex multicellular brown alga (*Ectocarpus siliculosus*), has been reported to have ITGA homologs.³⁵ The most canonical of these homologs is CBN77719 on NCBI. Therefore, we inferred an additional ITGA tree with this species along with another ITGA-like protein identified in a coccolithophorid haptophyte unicellular alga (*Emiliania huxleyi*: XP_005778208.1) (Figure S4). In this tree, we included some potentially homologous FG-GAP-containing proteins, GLPD and LINKIN (Figure S4), to observe their phylogenetic affinity to ITGA homologs. We observed a few more FG-GAP proteins in other taxa such as cyanobacteria (e.g., *Gloeobacter violaceus*: WP_011143221.1; *Nostoc punctiforme*: ACC84844.1; and *Crocospaera watsoni*: WP_007307935.1), Cryptophyta (*Guillardia theta*: XP_005842481.1), and a few other Stramenopiles (e.g., *Nannochloropsis* spp. CCMP1776: TFJ84058.1 and EWM21519 and *Cafeteria roenbergensis*: KAA0148838.1) on NCBI. These FG-GAP proteins either do not possess a transmembrane and a signal

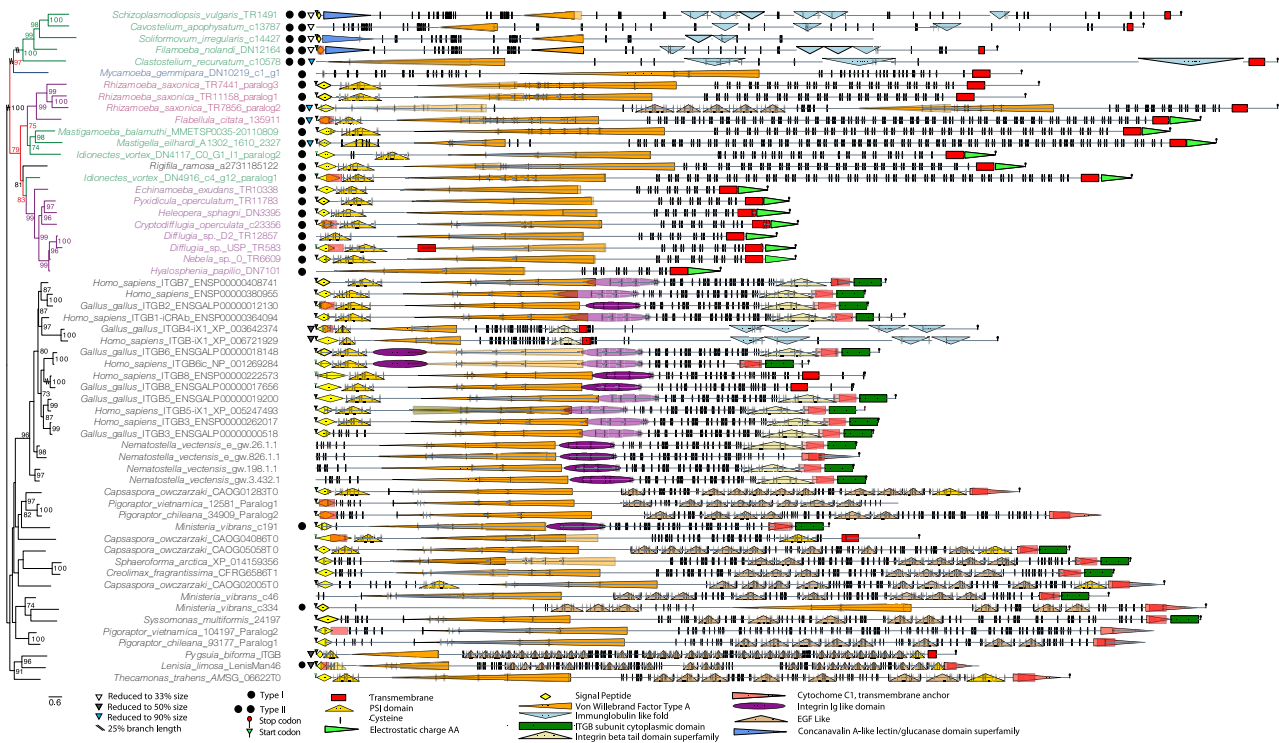


Figure 4. Protein domain architecture of ITGB from Amoebozoa

The domain architecture of ITGB for amoebozoan species is shown to the right side and the phylogenetic tree of Amoebozoa is shown at the left side of figure. Keys to each of domains color and shape are represented in the box at the bottom of the figure. Tree of amoebozoan ITGB 295 amino acids sites phylogeny of amoebozoan ITGB rooted with Obazoa. The tree was built using IQ-TREE v.1.5.5 under the LG+C60+F+G model of protein evolution. Numbers at nodes on the phylogenetic tree are MLBS support values derived from 1,000 ultrafast bootstrap replicates. Only values greater than 50% are shown. The color of branches amoebozoan sequences are illustrated according to Figure 3 in Kang et al.¹⁹ When domains are overlapped because of different software algorithms, these are represented in transparent colors. See also Figure S6 and Table S4.

peptide region, or they possess many transmembrane domains. These FG-GAP proteins fail to pass our minimum criteria of being a canonical ITGA. Using the identified ITGA homologs as queries, we recovered we identified several predicted β -propeller proteins in Archaea that have a similar architecture to the canonical ITGA using a standard BLAST-based approach against the *nr* database. An unrooted tree of β -propeller homologs recovered Archaea as paraphyletic across our phylogeny (Figure S5).

Integrin beta

We identified ITGB homologs in 19 of 113 amoebozoan genomes and transcriptomes examined. Of these, only a single homolog was identified in each taxon, except for three ITGB paralogs in *Rhizamoeba saxonica* and two paralogs in *Idionectes vortex* (Figure 4). We find amoebozoan ITGBs are present in all three major clades of Amoebozoa. However, like ITGA, these proteins are not found in the four amoebozoan taxa that have genome sequences (dictyostelids, entamoebids, acanthamoebids, and *Protostelium fungivorum*). We were able to find an ITGB homolog in the genome of the amoebozoan *Mastigamoeba balamuthi* ATCC 30984 (Figure 5). For ITGB, there were five introns in the genomic scaffold CBKX010020319.1 and eight introns in CBKX010020318.1. These introns are canonical *M. balamuthi* introns,³¹ and the other genes on the scaffold are

mastigamoebid/amoebozoan in nature. For example, on these ITGB genomic scaffolds (CBKX010020318.1 and CBKX010020319.1), we uncovered putative open reading frames (ORFs) for tenascin and PA14-like proteins. To illustrate that these scaffolds are amoebozoan in provenance, we inferred phylogenetic trees of these proteins, tenascin (extracellular glycoprotein) and PA14 (membrane adhesion protein), by searching for homologs in our extensive datasets. Both proteins clearly show amoebozoan signal (shown as a subset in Figure 5). Additionally, the phylogenetic signal of other ORFs on these genomic scaffolds shows amoebozoan signal (data not shown) and is not from contamination in the genome project. Furthermore, the presence of introns within the genomic scaffold for ITGB demonstrates that this protein/genomic scaffold is eukaryotic and likely part of the genome, which was obtained from an axenic culture. Interestingly, these genes neighboring ITGB are also membrane associated, which could be evidence that they are linked to the cytoplasmic membrane with co-regulated transcription.

The size range of ITGB paralogs in Amoebozoa varies widely from ~600 to 3,300 amino acids (Table S3). In addition to the novel ITGB proteins in Amoebozoa, we also found previously unreported ITGB homologs in two obazoa transcriptomes as well in a recently described clade closely related to Amorphea (CRuMs).³⁶ Based on domain and motif architecture from InterProScan outputs, we classify non-metazoan ITGBs into two

Introns in the *Mastigamoeba balamuthi* ITGB on Genomic Scaffolds

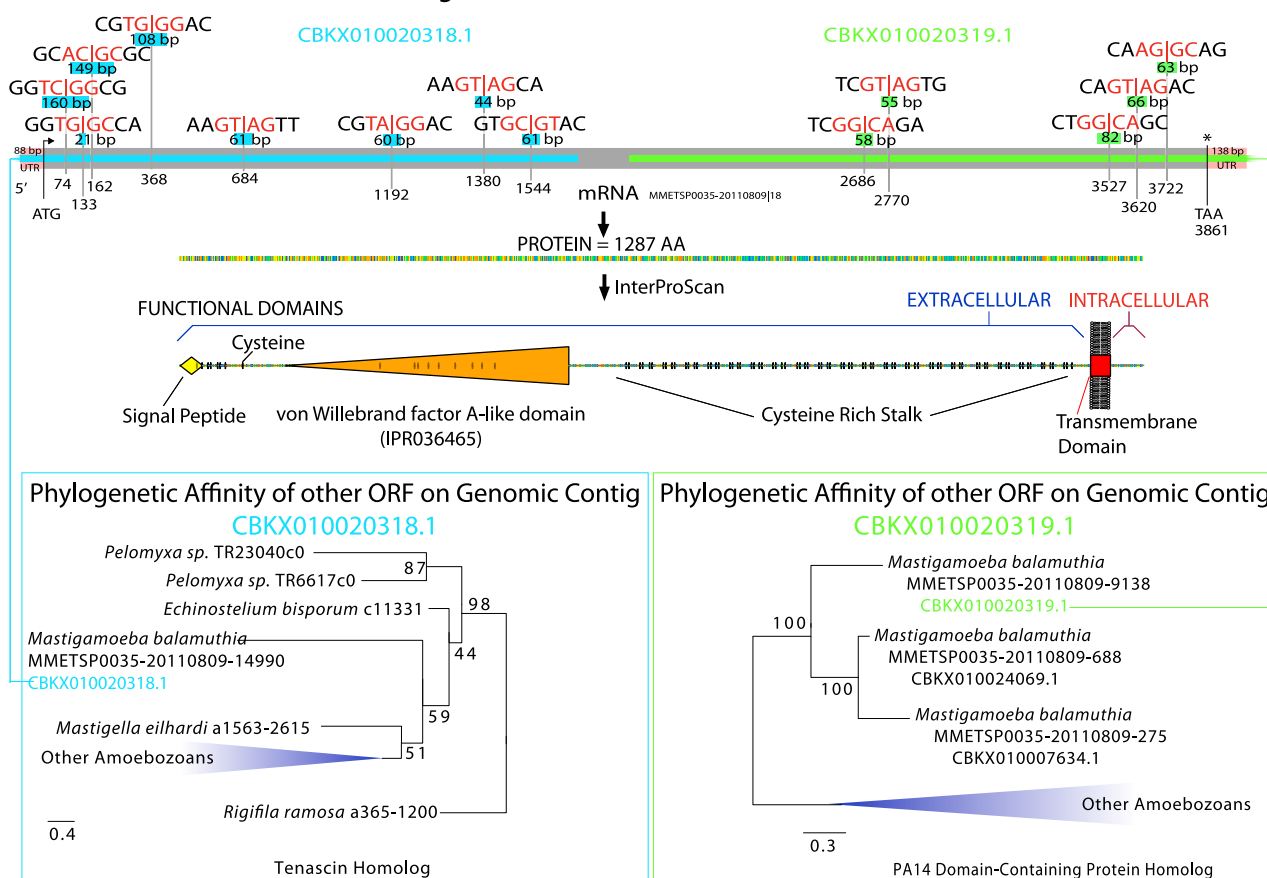


Figure 5. The presence of ITGB gene in a *Mastigamoeba balamuthi* genome sequence

Intron/exon boundary sequences are shown along with corresponding intron size. The genome contig that possesses the ITGB gene is colored blue and green. The protein domain architecture of the ITGB gene is shown along with protein size. The adjacent tenascin gene on the ITGB genome contig is shown left and right along with the inferred phylogenetic tree.

representative types: ITGB that are similar to a canonical ITGB (type I) and unique domain architecture proteins that are non-canonical ITGB-like proteins (type II).

Our first representative type of ITGB (type I) has a similar domain architecture to metazoan ITGB (Figure 4). In metazoan ITGB, which we refer herein to as “canonical ITGB proteins,” there are seven domains: a plexin-semaphorin-integrin (PSI) (IPR002165), a von Willebrand factor (IPR002369), a Ig-like domain (IPR013783), three to four EGF modules (IPR013111), an integrin β -tail subunit (IPR012896), transmembrane, and signal peptide domains (Figure 4; Figure S2B).⁵ From the 19 amoebozoan taxa in which we identified an ITGB, 14 taxa have type I (Figure 4). Most amoebozoans in this class appear to have the von Willebrand type A (vWA) domain, β subunit stalk region, and the PSI domains except two amoebozoans that lack PSI domains and one amoebozoan that possesses a duplicated vWA domain (Figure 4).

In metazoans, the metal-ion-dependent adhesive site (MIDAS) and a synergistic metal-ion-binding site (SyMBS) (the positive regulatory site for integrin activation³) motif is essential for integrin-ligand binding,^{3,5} but the site adjacent to MIDAS (ADMIDAS) motif (the negative regulatory³⁷ site for integrin inhibition) is not

absolutely necessary for integrin activation (Figure S2A).³⁷ MIDAS is well conserved, especially in amoebozoan β -I domains (Figures 6A and 6B), and SyMBS is well conserved across the amoebozoans (Figures 6C and 6D). However, ADMIDAS motifs are quite variable in Amoebzoa compared with the metazoan consensus pattern (Figure 6B; Table S3). The consensus pattern of the cysteine-rich stalk motif (CRSM) in ITGB homologs of Opisthokonta is CXCXXCXC^{12,38} (Figure S2Bf); in Metazoa, there are typically 3–4 CRSMs in a β subunit stalk region.⁵ Amoebozoan ITGB homologs have the CRSM pattern in the β subunit stalk region of DXCGXCGGX⁽³⁻⁶⁾CXGC (Figure 6F), ranging from three to seven times (Table S3). In β -tail motifs (IPR012896), the cytosolic tail motif (NPXY/F [IPR021157])³⁹ is present across Amoebzoa (Figure 6G) except one amoebozoan (Table S3), whereas an ITGA-interacting motif (LLXXHDKRKE [IPR014836]) is absent in Amoebzoa. Twelve amoebozoans in this class possess amino acids with a predicted electrostatic charge where the ITGA-interacting motif (LLXXHDKRKE) would be present in metazoan ITGB (Figures 4 and 6H).

A signal peptide and a transmembrane domain were predicted in most amoebozoan type I ITGB in this class, but three

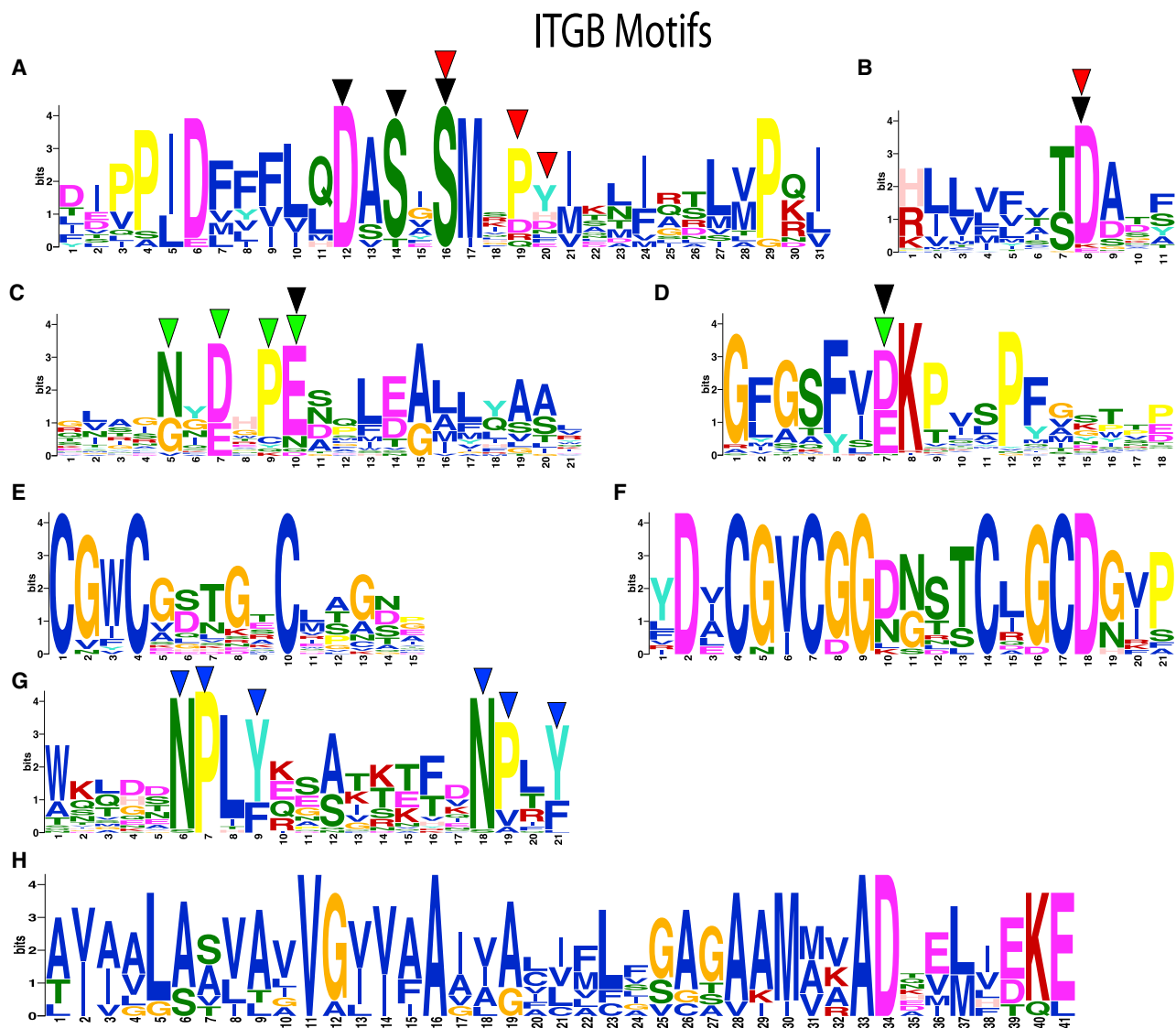


Figure 6. Most significantly enriched motifs in 18 amoebozoan ITGB homologs

(A–F) Three divalent cation-binding motifs, MIDAS, ADMIDAS, and SyMBS, are shown. These consensus motifs are obtained from type I amoebozoan ITGB homolog sequences only. (A) MIDAS (black arrow) and ADMIDAS (red arrow) motifs. (B) A shared amino acid (indicated by red/black arrows) in MIDAS and ADMIDAS motifs. (C) SyMBS motif represented by green arrow, and a shared amino acid of MIDAS motif indicated by black arrow. (D) SyMBS motif indicated by a green arrow is shown. (E and F) Cysteine-rich motifs (E) and PSI domain (F) and the cysteine-rich motif stalk. (G and H) Motifs found at the C-terminal are shown. (G) NPXY motif shown by blue arrow found in most amoebozoans' ITGB except three amoebozoan ITGB homologs. (H) An electrostatic charge motif is shown. This motif is discovered in type I ITGB homologs only.

See also [Figures S2](#) and [S3](#) and [Table S4](#).

amoebozoan ITGB type I lack signal peptides ([Figure 4](#)). Again, we do not discard the possibility of truncated ITGBs since these are present in metazoans^{22,23} and *Creolimax fragrantissima*.¹⁵

The ITGB proteins of non-metazoan obazoans have domain/motif architectures very similar to a canonical metazoan ITGB ([Figure 4](#); [Figures S3Ba–S3Bh](#); see [Figures S2Ba–S2Bh](#) for metazoan motifs). All of the obazoan ITGB MIDAS, ADMIDAS, and SyMBS motifs were well conserved ([Figures S2Ba–S2Bd](#)). At the b-tail motifs (IPR012896) of Obazoa, the cytosolic tail motif (NPXY/F) is present in all obazoan ITGB homologs ([Figure S3Bg](#)), but we did not observe the ITGA-interacting motif (LLXXXHDRKE) in ITGB of

Breviatea. We found a metazoan GXXXG motif in three amoebozoan ITGB ([Table S3](#)), suggesting that this motif is not animal specific. There are multiple tandem repeats of the opisthokont CRSM pattern in Obazoa ITGB, and both breviate taxa have a clear expansion of β subunit stalk region as previously reported ([Figure 4](#)).^{11–16}

We also investigated a possible “choanoflagellate ITGB gene” that was recently reported to be present in the transcriptome of *Didymoeca costata*.⁴⁰ The protein domain architecture has a serine protease flanking vWA, PSI domain. However, it is missing the CRSM, NPXY/F, LLXXXHDRKE, and GXXXG motifs. To our

surprise, *Rigifila ramosa*, which was recently placed in a novel supergroup sister to Amorphea called CRuMs,³⁶ has a protein that contains all the canonical domains of ITGB such as vWA, amoebozoan CRSM pattern, and C-terminal motifs (Figure 4) and all typical cation-binding motifs (Table S3). We were unable to find an ITGA homolog in either *Rigifila ramosa* and *Didymoeca costata*.

Our second representative class of ITGB, which we refer to as type II, are large proteins (~1,400–3,600 amino acids) that have novel domain architectures unobserved before in ITGB. Of the 19 amoebozoan taxa that have ITGB, five have an ITGB assigned to our type II class (Figure 4). Type II amoebozoan ITGB domain architectures include Laminin G3 (LamG3) (IPR013320), vWD (IPR001846), Ig-like fold (IPR013783), EGF-like (IPR013032), and a cyclic-nucleotide-binding (IPR000595) domains (Figure 4). A canonical human ITGB has four EGF domains^{3,5} that coincide with the location of Ig regions in amoebozoan ITGB. Three amoebozoan ITGB in this class have an extracellular EGF-like domain at the N terminus, at the location of the PSI domain (Figure 4). LamG3 (CSM) and EGF-like domains are recovered in two amoebozoan ITGB (Figure 4). The cation-binding motifs MIDAS and ADMIDAS are hypervariable in type II representatives compared with the metazoan consensus pattern, but SyMBS is well conserved except in *Filamoeba nolandi* (Table S3).

We selected three ITGB homologs from animals (*Homo sapiens*, *Gallus*, and *Nematostella vectensis*) as a selective reference ITGB sequences, because integrin proteins are highly conserved among Metazoa. We observed a clade of type II amoebozoan ITGB, breviatea with apusomonads ITGB, type I amoebozoan with the CRuMs *Rigifila limosa* ITGB homolog, and a clade of opisthokonts (Figure 4). With the placement of *Rigifila*, Amoebozoa is not recovered as a monophyletic group.

Proteins that are similar to ITGB have been discovered in Amoebozoa and a choanoflagellate as well as in bacteria. In order to verify our putative ITGBs were not actually members of these proteins classes, we predicted their protein domain architectures, added these to our protein alignments of ITGB, and built phylogenetic trees. The “similar to integrin beta” (Sib) protein is a cell-adhesive molecule that is structurally and functionally similar to ITGB, and it was discovered in *Dictyostelium discoideum*.⁴¹ However, the dictyostelids do not appear to have a canonical ITGB, only Sib proteins. Additionally, an ITGB homolog was shown in cyanobacteria, but it contained only one of the ITGB domains.¹² We inferred an additional ITGB tree with these species (the choanoflagellate *Didymoeca costata*, *Dictyostelium discoideum*, and cyanobacteria) (Figure S6). In this tree, a clade of type II ITGB is placed between the cyanobacteria *Trichodesmium* and the choanoflagellate *Didymoeca*.

Scaffold cytoplasmic proteins

In our comparative transcriptome and genome analysis, we observed the following intracellular scaffold cytoskeletal adhesion proteins: talin, α -actinin, vinculin, PINCH, paxillin, and ILK throughout Amoebozoa. These six adhesion proteins have previously been reported to be present in Amoebozoa.¹² Since then, the number of the core consensus adhesomes has been expanded.⁴² Filamin, kindlin, and tensin are three consensus adhesion proteins and cross-actin linkers that interact with ITGB.⁴² Across the diversity of Amoebozoa, we have observed

two putative adhesion proteins (i.e., filamin and tensin) that were not known in Amoebozoa (see STAR Methods). Talin, PINCH, filamin, and paxillin, key players in the integrin adhesion, were not present in seven amoebozoan transcriptomes (Figure 1A). However, their absence is suspected to be an artifact of low gene coverage in these data based on global scores from BUSCO (a tool that assesses genome completeness based on near-universal single-copy orthologs) (Table S4). Across the amoebozoan clade, we also noticed a large number of what appear to be truncated talin proteins compared with the rest of the adhesion proteins that we investigated (Figure 1B).

ILK was scattered across the Amoebozoa, unlike the case in Apusomonada and Opisthokonta,^{12–16} where ILK appears to be universally present in IMAC-containing taxa. α -Actinin is the only signaling protein that is present in transcriptomic datasets all of the amoebozoan taxa. Consistent with previous reports, we did not detect focal adhesion kinase (FAK), cellular proto-oncogene tyrosine-protein kinase Src (cSRC), or parvin proteins in any of the Amoebozoa- or Breviatea¹¹-predicted proteome data examined. Since the gene coverage for the transcriptome of *Pygmaea bifurcata* (Breviatea) is high and genomes are available for *Lenisia limosa* (Breviatea) and *Entamoeba* spp. (Amoebozoa), we conclude that breviatea and *Entamoeba* spp. likely do not possess vinculin proteins (Figure 1). The genome of the brown alga *Ectocarpus siliculosus* has talin and α -actinin homologs,³⁵ but it does not contain any other integrin-signaling adhesion protein homologs.

DISCUSSION

Our findings show that the IMAC is not exclusive to obazoaans, as previously proposed.^{11,12,14–16} We report a nearly full set of IMAC proteins is present in amorphous taxa including five amoebozoan species (*Flabellula citata*, *Rhizamoeba saxonica*, *Cryptodifflugia operculatum*, *Mastigamoeba balamuthi*, and *Nebela* sp.) and two previously unsurveyed obazoan taxa, *Lenisia limosa* (Breviatea) and *Ministeria vibrans* (Opisthokonta). Our increased taxon sampling that spans the breadth of the known diversity within Amoebozoa permitted discovery of these proteins. While the integrin proteins undoubtedly played a role in the origins of animal multicellularity,^{12,43,44} we do not find any evidence of integrins in the few multicellular lineages of Amoebozoa, those that socially aggregate to form emergent fruiting-body structures (*Copromyxa protea* and the model dictyostelids). Interestingly, secondary loss of these proteins seems to be rampant, as the majority of amoebozoan taxa appear to lack integrins or some of the IMAC components, but the caveat is that some of these data are based on incomplete transcriptomic data. Therefore, their absence may be due to the available data. Nonetheless, the most parsimonious explanation is that the IMAC was present in the ancestor of Amoebozoa. An analogous scenario of secondary loss of the IMAC plays out in obazoan evolution, in which we see that the closest relative of animals, the Choanoflagellata,^{40,45,46} as well as the Nucleomyxozoa (also referred to as the Holomycota, the fungal lineage), are devoid of key IMAC components (namely the integrins).^{11,12}

The functional domains of ITGA in amoebozoans are similar to the canonical ITGA found in animals and their close relatives such as the opisthokont filastereans *Capsaspora owczarzakii*,¹²

Ministeria vibrans, and *Pigoraptor* spp.¹³ Given the homologous nature of the functional domain architecture in amoebozoan ITGA, it is likely that their function at the cellular level is similar to those in Metazoa. Therefore, we suspect the last common ancestor of Amorphea (LCAA) probably already had ITGA within its genome. Of note, there are potential ITGA homologs in Archaea, it is possible that ITGA homologs predate the eukaryotic last common ancestor, but were retained mostly in Amorphea. We also suspect that the ITGA Ig-like domains possessed by some amoebozoans and obazoa are equivalent to metazoan ITGA legs originally thought to be specific to metazoans.^{11,12} An ITGA leg domain was probably in the ITGA of the LCAA and secondarily lost in some taxa.

Some novel functional domains in type II representatives are integrin-binding domains that are present in ECM proteins such as LamG,⁴⁷ TSP1,⁴⁸ and EGF-like domains.⁴⁹ Could the last common ancestor of type II representatives or LCAA contain these integrin-binding domains? Other functional domains in amoebozoan ITGA such as protein kinase, TSP1 together with WSC, and intramolecular chaperone are not found in Obazoa. Domain shuffling is an important mechanism for creating novel genes in evolution.⁵⁰ It has been suggested that many domains in the ECM are found in unrelated proteins.⁵¹ Whether these proteins were parts of the domain shuffling in ITGA of the last common ancestor of Amoebozoa, which resulted in ECM evolution remains to be seen and their function remains elusive. However, based on our results, it is possible that we have predicted novel integrin proteins with catalytic as well as adhesive properties.

The domains of amoebozoan ITGB are very similar to the canonical metazoan ITGB, and we suspect that the last common ancestor of Amoebozoa retained these canonical ITGB structures, including the CRSM motif and three metal-binding sites. The cytosolic tail motif “GFFKR” is probably specific to Obazoa, and there may be no interaction in the C-terminal intracellular region of integrins in Amoebozoa. The presence of transmembrane and the signal peptide regions are sporadic in Amorphea, including Metazoa; a previous study showed shedding (evolution from membrane docked to a secretory function) of integrins occurs.²³ The amoebozoan ITGB cysteine pattern of CGXCGGXXXXCXGC possesses more amino acids, including glycine, compared with Obazoa (Figure 6F). Glycine is a small non-polar amino acid⁵² and due to its small size, glycine positions are often highly conserved within proteins.^{53,54} However, the reason for the differences in glycine content remains a mystery.

The type II ITGB form an independent clade, and it is a sister to the other amoebozoans and *R. ramosa* in Figure 4. LamG3 and vWD in ITGB have never been observed in Obazoa, and the functions of these type II ITGB proteins are unknown. Some unicellular protists such as *breviata*, *Capsaspora owczarzaki*, and apusomonads have a long expansion of the CRSM for ITGB.^{11,12} We suspect that *Breviatea* with a long integrin legs form a long integrin heterodimerization complex at the extracellular area of the cell, as both ITGA and ITGB have long legs of nearly equal proportions in *Pygmaia bifurcata* (Figures 2 and 3).¹¹

Our prediction is that most of the Discosea, which includes *Acanthamoeba* (again with several genomes available) and many other sampled transcriptomes in our study, have lost these proteins. However, one representative discosean (*Mycamoeba*

gemmipara) has canonical ITGB homolog (Figure 4). Additionally, four other amoebozoan taxa with genome sequences available also appear to have lost ITGB in their evolution. There are amoebozoans in which we have observed only ITGA or ITGB receptors, without their counterpart integrin protein. A study by Li et al.⁵⁵ and Schneider et al.⁵⁶ showed that homodimerization of ITGA or ITGB might occur by clustering of integrin in a lipid raft. Our results show that many of the amoebozoan species singly possess either ITGA or ITGB, and yet they possess all the signal components of IMAC. These amoebozoan with singular integrins may be capable of initiating the integrin signaling pathway by a similar process.

Also, the question remains that the complexity of extra domains of integrin are not species related and their compositions are a single multidomain protein that is present only in Amoebozoa. Phylogenetically, these orthologs (i.e., amoebozoan ITGA and ITGB) form novel clades independent to known organisms and most likely are ancestral to the Amoebozoa as a whole. In order to further examine the overall distributions and protein architectures, genomic sequencing and further increased sampling of taxa are necessary. Additionally, localization and gene knockout strategies are necessary to understand the function of these “multicellularity” proteins in unicellular amoebozoans. Recent work in unicellular holozoans shows that they do function in adhesion to substrates.⁵⁷

Our most comprehensive comparative genomic and transcriptomic survey of Amoebozoa has revealed the presence of a full set of IMAC across the Amoebozoa. To date, this complex was believed to be Obazoa specific; our data show that these IMAC proteins predate Obazoa. Therefore, the last common ancestor of Amorphea contains the necessary machinery of IMAC; yet, what functional role they play in these unicellular protists remains unknown. In future work, localization and functionality studies of these IMAC proteins will be conducted and will be critical for unlocking their mysteries.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Culturing of *Mycamoeba gemmipara*
- METHOD DETAILS
 - Transcriptome library construction
 - Collection and identification of novel IMAC proteins
 - Confirmation of amoebozoan integrins
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Phylogenetic trees

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2021.04.076>.

ACKNOWLEDGMENTS

This project was supported in part by US National Science Foundation (NSF), Division of Environmental Biology (DEB) grants 1456054 and 2100888 (<https://www.nsf.gov/>), awarded to M.W.B. We thank Prof. Andrew J. Roger (Dalhousie University) for advanced access to *Mastigamoeba balamuthi* transcriptome, which was supported by grant MOP-142349 from the Canadian Institutes of Health Research, awarded to A.J. Roger. We thank the anonymous reviewers for the constructive criticism and helpful comments, leading to a better manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, supervision, project administration, and funding acquisition, M.W.B.; resources, D.J.G.L., C.W.S., and M.W.B.; data generation, S.K., D.J.G.L., M.W.B., A.K.T., R.E.J., and C.W.S.; formal analysis and investigation, S.K., A.K.T., and M.W.B.; writing – original draft, S.K. and M.W.B.; writing – review editing, S.K., A.K.T., C.W.S., R.E.J., D.J.G.L., and M.W.B.; manuscript revision, A.K.T., M.W.B., S.K., and C.W.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 30, 2020

Revised: March 17, 2021

Accepted: April 28, 2021

Published: June 1, 2021

REFERENCES

- Hynes, R.O. (2002). Integrins: bidirectional, allosteric signaling machines. *Cell* 110, 673–687.
- LaFlamme, S.E., Mathew-Steiner, S., Singh, N., Colello-Borges, D., and Nieves, B. (2018). Integrin and microtubule crosstalk in the regulation of cellular processes. *Cell. Mol. Life Sci.* 75, 4177–4185.
- Zhang, K., and Chen, J. (2012). The regulation of integrin function by divalent cations. *Cell Adhes. Migr.* 6, 20–29.
- Xia, W., and Springer, T.A. (2014). Metal ion and ligand binding of integrin $\alpha 5 \beta 1$. *Proc. Natl. Acad. Sci. USA* 111, 17863–17868.
- Campbell, I.D., and Humphries, M.J. (2011). Integrin structure, activation, and interactions. *Cold Spring Harb. Perspect. Biol.* 3, a004994.
- Lee, J.-O., Rieu, P., Arnaout, M.A., and Liddington, R. (1995). Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell* 80, 631–638.
- Xiong, J.-P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D.L., Joachimiak, A., Goodman, S.L., and Arnaout, M.A. (2001). Crystal structure of the extracellular segment of integrin $\alpha V \beta 3$. *Science* 294, 339–345.
- LaFlamme, S.E., and Auer, K.L. (1996). Integrin signaling. *Semin. Cancer Biol.* 7, 111–118.
- Harburger, D.S., and Calderwood, D.A. (2009). Integrin signalling at a glance. *J. Cell Sci.* 122, 159–163.
- Horton, E.R., Byron, A., Askari, J.A., Ng, D.H.J., Millon-Frémillon, A., Robertson, J., Koper, E.J., Paul, N.R., Warwood, S., Knight, D., et al. (2015). Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. *Nat. Cell Biol.* 17, 1577–1587.
- Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A.A., Lang, B.F., Simpson, A.G., and Roger, A.J. (2013). Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc. Biol. Sci.* 280, 20131755.
- Sebé-Pedrós, A., Roger, A.J., Lang, F.B., King, N., and Ruiz-Trillo, I. (2010). Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc. Natl. Acad. Sci. USA* 107, 10142–10147.
- Hehenberger, E., Tikhonenkov, D.V., Kolisko, M., Del Campo, J., Esaulov, A.S., Mylnikov, A.P., and Keeling, P.J. (2017). Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. *Curr. Biol.* 27, 2043–2050.e6.
- Grau-Bové, X., Torruella, G., Donachie, S., Suga, H., Leonard, G., Richards, T.A., and Ruiz-Trillo, I. (2017). Dynamics of genomic innovation in the unicellular ancestry of animals. *eLife* 6, e26036.
- de Mendoza, A., Suga, H., Permanyer, J., Irimia, M., and Ruiz-Trillo, I. (2015). Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *eLife* 4, e08904.
- Dudin, O., Ondracka, A., Grau-Bové, X.A.B., Haraldsen, A., Toyoda, A., Suga, H., Bråte, J., and Ruiz-Trillo, I. (2019). A unicellular relative of animals generates an epithelium-like cell layer by actomyosin-dependent cellularization. *eLife* 8, e49801.
- Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., et al. (2019). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J. Eukaryot. Microbiol.* 66, 4–119.
- Cavalier-Smith, T. (2017). Origin of animal multicellularity: precursors, causes, consequences—the choanoflagellate/sponge transition, neurogenesis and the Cambrian explosion. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, 20150476.
- Kang, S., Tice, A.K., Spiegel, F.W., Silberman, J.D., Pánek, T., Čepička, I., Kostka, M., Kosakyan, A., Alcántara, D.M.C., Roger, A.J., et al. (2017). Between a pod and a hard test: the deep evolution of amoebae. *Mol. Biol. Evol.* 34, 2258–2270.
- Hamann, E., Gruber-Vodicka, H., Kleiner, M., Tegetmeyer, H.E., Riedel, D., Littmann, S., Chen, J., Milucka, J., Viehweger, B., Becker, K.W., et al. (2016). Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. *Nature* 534, 254–258.
- Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., et al. (2017). InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* 45 (D1), D190–D199.
- Jin, R., Trikha, M., Cai, Y., Grignon, D., and Honn, K.V. (2007). A naturally occurring truncated $\beta 3$ integrin in tumor cells: native anti-integrin involved in tumor cell motility. *Cancer Biol. Ther.* 6, 1559–1568.
- Gomez, I.G., Tang, J., Wilson, C.L., Yan, W., Heinecke, J.W., Harlan, J.M., and Raines, E.W. (2012). Metalloproteinase-mediated Shedding of Integrin $\beta 2$ promotes macrophage efflux from inflammatory sites. *J. Biol. Chem.* 287, 4581–4589.
- De Melker, A.A., Kramer, D., Kuikman, I., and Sonnenberg, A. (1997). The two phenylalanines in the GFFKR motif of the integrin $\alpha 6 A$ subunit are essential for heterodimerization. *Biochem. J.* 328, 529–537.
- Singh, R., Schilde, C., and Schaap, P. (2016). A core phylogeny of Dictyostelia inferred from genomes representative of the eight major and minor taxonomic divisions of the group. *BMC Evol. Biol.* 16, 251.
- Schaap, P., Barrantes, I., Minx, P., Sasaki, N., Anderson, R.W., Bénard, M., Biggar, K.K., Buchler, N.E., Bundschuh, R., Chen, X., et al. (2015). The *Physarum polycephalum* Genome Reveals Extensive Use of Prokaryotic Two-Component and Metazoan-Type Tyrosine Kinase Signaling. *Genome Biol. Evol.* 8, 109–125.
- Loftus, B., Anderson, I., Davies, R., Alsmark, U.C.M., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R.P., Mann, B.J., et al. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature* 433, 865–868.
- Clarke, M., Lohan, A.J., Liu, B., Lagkouravdos, I., Roy, S., Zafar, N., Bertelli, C., Schilde, C., Kianianmomeni, A., Bürglin, T.R., et al. (2013). Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol.* 14, R11.
- Hillmann, F., Forbes, G., Novohradská, S., Ferling, I., Riege, K., Groth, M., Westermann, M., Marz, M., Spaller, T., Winckler, T., et al. (2018). Multiple

- Roots of Fruiting Body Formation in Amoebozoa. *Genome Biol. Evol.* **10**, 591–606.
30. Nývltová, E., Šuták, R., Harant, K., Šedínová, M., Hrdý, I., Pačes, J., Vlček, Č., and Tachezy, J. (2013). NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. *Proc. Natl. Acad. Sci. USA* **110**, 7371–7376.
31. Žárský, V., Klimeš, V., Pačes, J., Vlček, Č., Nývltová, E., Hrdý, I., Pyrih, J., Mach, J., Barlow, L., Eliáš, M., et al. (2021). *Mastigamoeba balamuthi* genome and the nature of the free-living ancestor of *Entamoeba*. *Mol. Biol. Evol.* **2021**, msab020.
32. Metcalf, J.A., Funkhouser-Jones, L.J., Brileya, K., Reysenbach, A.-L., and Bordenstein, S.R. (2014). Antibacterial gene transfer across the tree of life. *eLife* **3**, e04266.
33. Clarke, J., Cota, E., Fowler, S.B., and Hamill, S.J. (1999). Folding studies of immunoglobulin-like β -sandwich proteins suggest that they share a common folding pathway. *Structure* **7**, 1145–1153.
34. Attwood, T.K., Bradley, P., Flower, D.R., Gaulton, A., Maudling, N., Mitchell, A.L., Moulton, G., Nordlie, A., Paine, K., Taylor, P., et al. (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.* **31**, 400–402.
35. Cock, J.M., Sterck, L., Rouzé, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.-M., Badger, J.H., et al. (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621.
36. Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A.K., Shiratori, T., Ishida, K.-I., Hashimoto, T., Simpson, A.G.B., and Roger, A.J. (2018). Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biol. Evol.* **10**, 427–433.
37. Mould, A.P., Barton, S.J., Askari, J.A., Craig, S.E., and Humphries, M.J. (2003). Role of ADMIDAS cation-binding site in ligand recognition by integrin $\alpha 5 \beta 1$. *J. Biol. Chem.* **278**, 51622–51629.
38. Tan, S.-M., Walters, S.E., Mathew, E.C., Robinson, M.K., Drbal, K., Shaw, J.M., and Law, S.K.A. (2001). Defining the repeating elements in the cysteine-rich region (CRR) of the CD18 integrin $\beta 2$ subunit. *FEBS Lett.* **505**, 27–30.
39. Calderwood, D.A., Fujioka, Y., de Pereda, J.M., García-Alvarez, B., Nakamoto, T., Margolis, B., McGlade, C.J., Liddington, R.C., and Ginsberg, M.H. (2003). Integrin β cytoplasmic domain interactions with phosphotyrosine-binding domains: a structural prototype for diversity in integrin signaling. *Proc. Natl. Acad. Sci. USA* **100**, 2272–2277.
40. Richter, D., Fozouni, P., Eisen, M., and King, N. (2018). The ancestral animal genetic toolkit revealed by diverse choanoflagellate transcriptomes. *eLife* **7**, e34226.
41. Cornillon, S., Gebbie, L., Benghezal, M., Nair, P., Keller, S., Wehrle-Haller, B., Charette, S.J., Brückert, F., Letourneur, F., and Cosson, P. (2006). An adhesion molecule in free-living *Dictyostelium amoebae* with integrin β features. *EMBO Rep.* **7**, 617–621.
42. Humphries, J.D., Chastney, M.R., Askari, J.A., and Humphries, M.J. (2019). Signal transduction via integrin adhesion complexes. *Curr. Opin. Cell Biol.* **56**, 14–21.
43. Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedrós, A., Brown, M.W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., et al. (2013). The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 2325.
44. Brunet, T., and King, N. (2017). The Origin of Animal Multicellularity and Cell Differentiation. *Dev. Cell* **43**, 124–140.
45. King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788.
46. Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., et al. (2013). Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* **14**, R15.
47. Pulido, D., Hussain, S.-A., and Hohenester, E. (2017). Crystal Structure of the Heterotrimeric Integrin-Binding Region of Laminin-111. *Structure* **25**, 530–535.
48. Calzada, M.J., Annis, D.S., Zeng, B., Marcinkiewicz, C., Banas, B., Lawler, J., Mosher, D.F., and Roberts, D.D. (2004). Identification of novel $\beta 1$ integrin binding sites in the type 1 and type 2 repeats of thrombospondin-1. *J. Biol. Chem.* **279**, 41734–41743.
49. Hynes, R.O. (2012). The evolution of metazoan extracellular matrix. *J. Cell Biol.* **196**, 671–679.
50. Babushok, D.V., Ohshima, K., Ostertag, E.M., Chen, X., Wang, Y., Mandal, P.K., Okada, N., Abrams, C.S., and Kazazian, H.H., Jr. (2007). A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. *Genome Res.* **17**, 1129–1138.
51. Adams, J.C. (2013). Extracellular Matrix Evolution: An Overview. In *Evolution of Extracellular Matrix*, F.W. Keeley, and R.P. Mecham, eds. (Springer), pp. 1–25.
52. Lodish, H., Darnell, J.E., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., and Matsudaira, P. (2008). *Molecular cell biology* (Macmillan).
53. Creighton, T.E. (1993). *Proteins: structures and molecular properties* (Macmillan).
54. Branden, C.I., and Tooze, J. (2012). *Introduction to protein structure* (Garland Science).
55. Li, R., Mitra, N., Gratkowski, H., Vilaire, G., Litvinov, R., Nagasami, C., Weisel, J.W., Lear, J.D., DeGrado, W.F., and Bennett, J.S. (2003). Activation of integrin $\alpha 5 \beta 3$ by modulation of transmembrane helix associations. *Science* **300**, 795–798.
56. Schneider, D., and Engelman, D.M. (2004). Involvement of transmembrane domain interactions in signal transduction by α / β integrins. *J. Biol. Chem.* **279**, 9840–9846.
57. Parra-Acero, H., Harcet, M., Sánchez-Pons, N., Casacuberta, E., Brown, N.H., Dudin, O., and Ruiz-Trillo, I. (2020). Integrin-Mediated Adhesion in the Unicellular Holozoan *Capsaspora owczarzaki*. *Curr. Biol.* **30**, 4270–4275.e4.
58. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659.
59. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652.
60. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512.
61. Möller, S., Croning, M.D., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653.
62. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.
63. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60.
64. Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., and Stoeckert, C.J. (2002). Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. *Current Protocols in Bioinformatics* (John Wiley & Sons).
65. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.
66. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
67. Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic

- p>informative regions from multiple sequence alignments.
- BMC Evol. Biol.*
- 10, 210.
68. Nguyen, L.T., Schmidt, H.A., Haeseler, A., and Minh, B.Q. (2014). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274.
 69. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
 70. Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423.
 71. MacManes, M.D. (2018). The Oyster River Protocol: A Multi Assembler and Kmer Approach For de novo Transcriptome Assembly. *PeerJ* 6, e5428.
 72. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.
 73. Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 33, 3387–3395.
 74. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* 33, 1635–1638.
 75. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47 (D1), D427–D432.
 76. Van Dongen, S.M. (2000). Graph clustering by flow simulation. PhD thesis (Center for Math and Computer Science).
 77. Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO. *Methods Mol. Biol.* 1962, 227–245.
 78. Blandenier, Q., Seppey, C.V.W., Singer, D., Vlimant, M., Simon, A., Duckert, C., and Lara, E. (2017). *Mycamoeba gemmipara* nov. gen., nov. sp., the First Cultured Member of the Environmental Dermamoebidae Clade LKM74 and its Unusual Life Cycle. *J. Eukaryot. Microbiol.* 64, 257–265.
 79. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
 80. Onsbring, H., Tice, A.K., Barton, B.T., Brown, M.W., and Ettema, T.J.G. (2019). An efficient single-cell transcriptomics workflow to assess protist diversity and lifestyle. *bioRxiv*. <https://doi.org/10.1101/782235>.
 81. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
2-propanol(certified ACS)	Sigma-Aldrich	Catalog Number: I9516-25ML
Ethanol (pure) 200 proof	Fisher Sciences	Catalog Number: BP2818100
Critical commercial assays		
Nextera XT DNA Library Prep Kit	Illumina	Catalog Number: FC-131-1096
Nextera XT index kit v2 set A	Illumina	Catalog Number: FC-131-2001
Nextera XT index kit v2 set B	Illumina	Catalog Number: FC-131-2002
Nextera XT index kit v2 set C	Illumina	Catalog Number: FC-131-2003
Nextera XT index kit v2 set C	Illumina	Catalog Number: FC-131-2003
Deposited data		
<i>Mastigamoeba balamuthi</i> (transcriptome)	NCBI Bioproject PRJNA724719	Assembly - https://doi.org/10.5061/dryad.gxd2547jk
<i>Mycamoeba gemmipara</i> (transcriptome)	NCBI Bioproject PRJNA724710	Assembly - https://doi.org/10.5061/dryad.gxd2547jk
Phylogenetic trees and alignments	Dyrad	https://datadryad.org/stash/share/UzDxrdpQ-o0T2lcXkP_nDrG42JwYfNPYRVqZ8roFGBI
Taxa examined and data sources, see Table S5	N/A	N/A
Software and algorithms		
CD-HIT	Li and Godzik ⁵⁸	https://github.com/weizhongli/cdhit
Trinity v2.4.0	Grabherr et al. ⁵⁹	https://github.com/trinityrnaseq/trinityrnaseq/
TransDecoder v 5.5.0	Haas et al. ⁶⁰	https://transdecoder.github.io/
TmHm v 2.0	Möller et al. ⁶¹	http://www.cbs.dtu.dk/services/TMHMM/
Bowtie2 v 2.3.4.3	Langmead and Salzberg ⁶²	https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.3.3.1
Diamond v 0.9.25	Buchfink et al. ⁶³	https://github.com/bbuchfink/diamond
OrthoMCL v 5.0	Fischer et al. ⁶⁴	https://orthomcl.org/orthomcl/
Trimomatic v 0.35	Bolger et al. ⁶⁵	https://github.com/timflutre/trimomatic
Mafft-Linsi v 7	Katoh and Standley ⁶⁶	https://mafft.cbrc.jp/alignment/server/
Bmge v 1.12	Criscuolo and Gribaldo ⁶⁷	ftp://ftp.pasteur.fr/pub/GenSoft/projects/BMGE
SEQUENCER v 5.4.6.		http://www.genecodes.com/
BLAST 2.2.30+		http://blast.ncbi.nlm.nih.gov/blast.ncbi.nlm.nih.gov/Blast.cgi
IQTREE v 1.5.5	Nguyen et al. ⁶⁸	http://www.iqtree.org
Rsem	Li and Dewey ⁶⁹	https://github.com/deweylab/RSEM
SignalIP 5.0	Almagro Armenteros et al. ⁷⁰	http://www.cbs.dtu.dk/services/SignalP/
Oyster River Protocol v 2.1.1	MacManes ⁷¹	https://oyster-river-protocol.readthedocs.io/en/latest/
Meme-Suite v 5.0.4	Bailey et al. ⁷²	https://meme-suite.org/index.html
DeepLoc –1.0	Almagro Armenteros et al. ⁷³	http://www.cbs.dtu.dk/services/DeepLoc/
InterProScan 5.27-66.0	Finn et al. ²¹	https://www.ebi.ac.uk/interpro/search/sequence-search
ETE3	Huerta-Cepas et al. ⁷⁴	http://etetoolkit.org/documentation/ete-view/
Pfam	El-Gebali et al. ⁷⁵	https://pfam.xfam.org/
Mcl	Van Dongen ⁷⁶	https://micans.org/mcl/
BUSCO v5	Seppely et al. ⁷⁷	https://busco.ezlab.org

RESOURCE AVAILABILITY

Lead contact

Further information and request for resources should be directed to, and will be fulfilled by, the Lead Contact Matthew W. Brown (matthew.brown@msstate.edu).

Materials availability

This study did not generate any new reagents.

Data and code availability

The *Mycamoeba gemmipara* and *Mastigamoeba balamuthi* transcriptome raw reads are archived on NCBI under BioProject PRJNA724710 and PRJNA724719, which also shown in [Key resources table](#). Assemblies of these data are provided on the DRYAD data repository (<https://doi.org/10.5061/dryad.gxd2547jk>). All protein and nucleotide sequences of each gene, all protein alignments (trimmed and untrimmed) used for the phylogenetic analyses, and phylogenetic trees are also available on the DRYAD repository under the above citation.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The genomic level data examined in this project are listed in [Table S5](#).

Culturing of *Mycamoeba gemmipara*

Mycamoeba gemmipara was obtained from Q. Blandenier and cultured as in Blandenier et al.⁷⁸ Briefly, amoebae were grown on weak malt yeast extract agar media (1 l distilled H₂O, 0.75 g K₂PO₄, 0.002 g yeast extract, 0.002 g malt extract, 15 g Bacto agar). Amoebae were transferred from feeding fronts to wMY plates streaked with *E. coli* (strain MG1655 (ATCC 700926)) for the generation of mono-eukaryotic and clonal cultures.

METHOD DETAILS

Transcriptome library construction

Ultra-low input/single-cell cDNA library construction based on Smart-Seq2

About 50 *M. gemmipara* cells were scraped off of an agar plate using a 30-gauge platinum wire placed directly into a 200 μ L thin-walled PCR tube. The cells were subjected to a modified version of Smart-Seq2⁷⁹ that includes an additional freeze thaw step for cell lysis described in Onsbring et al.⁸⁰ for mRNA extraction and cDNA library preparation. The resulting cDNA libraries were prepared for sequencing on the Illumina platform using a Nextera XT DNA Library Prep Kit (Illumina, CA) following the manufacturer's protocol with dual index primers. The *M. gemmipara* library was pooled with libraries for an unrelated study and sequenced using an Illumina HiSeq 4000 at Genome Quebec.

Transcriptomic sequencing assembly

Low quality bases, adaptor sequences, and Smart-Seq2 primer sites were removed from the raw sequencing reads of *M. gemmipara* with TRIMMOMATIC v 0.35.⁶⁵ Surviving reads were assembled using the *de novo* assembly program TRINITY.⁶⁰ Nucleotide sequences were translated and open reading frames were predicted with TRANSDCODER v 5.5.0 (<https://github.com/TransDecoder/TransDecoder>). After initial examination, the predicted proteomes of some taxa contained clearly truncated predicted proteins of interest. We employed two additional strategies to improve the assembly of these transcriptomes or individual sequences. To improve the overall assembly for some taxa, the raw reads were assembled more rigorously following the steps outlined in the Oyster River Protocol.⁷¹ For individual sequences of interest, contigs generated by our automated assembly were blasted (BLASTN) back to the raw nucleotide data and to the assembly for each transcriptome. Hits were collected and assembled using SEQUENCHER v 5.4.6 (GeneCodes, Madison, WI, USA). Each transcriptome was analyzed for completeness using the default methodology of BUSCO v 5.⁷⁷

Collection and identification of novel IMAC proteins

Canonical IMAC proteins and their protein architecture

We selected human IMAC proteins from NCBI: ITGA5:P08648, ITGB1:NP_002202, Talin:AAF27330, Parvin:AAH16713, PINCH:NP_060450.2, vinculin:AAH39174, FAK:AAA35819 Paxillin:AAC50104 ILK:NP_001014794, α -actinin:AAC17470, Filamin:AAF72339, and Tensin:AAG33700, as well as previously reported integrin sequences from protistan species. INTERPROSCAN 5.27-66.0²¹ was used to determine these IMAC proteins domain architecture along with SIGNALIP v 5.0⁷⁰ and TMHMM v 2.0⁸¹. MEME-SUITE 5.0.4⁷² was used to examine integrin motifs. DEEPLoc v 1.0⁷³ was used to determine the subcellular localization of integrin proteins. We set the minimum criteria of canonical IMAC proteins based on their architecture and motifs. We used OrthoMCL to assign IMAC proteins their own ortholog numbers.

For integrins alpha and beta we collected putative novel sequences from our newly examined data for further analyses using two strategies. In our first strategy we created a novel ortholog database by adding the transcriptomes and proteomes listed in [Table S5](#)

to OrthoMCL v.5.0. To do this, an all-against-all BLAST using DIAMOND-BLASTP was conducted using each protein from the aforementioned data as queries. Up to 1,000 hits that met our significance threshold of $e\text{-value} < 1e-5$ were collected. These Blast results were then clustered using the OrthoMCL pipeline methodology. We then collected all sequences within orthogroups that contained metazoan or previously identified protistan integrins from our newly constructed database. These sequences were again used as queries in an all-versus-all BLASTP against one another. All hits above our threshold ($e\text{-value} 1e-10$) were collected and MCL clustered. In an effort to reduce redundancy in our database of putative integrins all sequences were then subjected to clustering using CD-HIT at 0.95 global sequence identity. We also removed any sequences less than 500 amino acids in length. In our second strategy we collected proteins that had predicted Pfam domains ($e\text{-value} < 1e-10$) that are hallmarks of canonical integrins. For ITGA we collected any sequence containing FG-GAP (PF01839). For ITGB we collected any sequences that contained one or more of the following PSI_integrin (PF17205), integrin_beta (PF00362), integrin_b_cyt (PF08725), or integrin B tail (PF07965). For the remaining IMAC proteins, we used BLASTP to search our novel ortholog database using human and previously identified protistan integrins as queries. Sequences with hits ($e\text{-value} < 1e-10$) were collected for further examination.

Finally, the identity of the putative IMAC proteins were confirmed or rejected by comparing results of analyses using INTERPROSCAN, SIGNALIP, TMHMM, DEEPLoc, and MEME-SUITE against our minimum criteria established based on the examining the output of these programs when canonical or previously identified IMAC proteins were used as input. Subsequently, our collected IMAC proteins were used to search newly available genomic and transcriptomic data using BLASTP as it became available.

Confirmation of amoebozoan integrins

Presence of integrin in the *Mastigamoeba* genome

To confirm the presence of these genes in the genome, we used the integrin transcripts of *M. balamuthi* as a query against the whole shotgun genome data (CBKX000000000; <https://www.ncbi.nlm.nih.gov/nuccore/CBKX000000000.1>) using BLAST.³⁰ The common splicing sites were searched in integrins intron and exon boundaries by assembling the integrin genomic contig and integrin transcript in SEQUENCHER v 5.4.6. The genome contigs of which contained integrin gene were searched for additional ORFs to examine the phylogenetic affinity of other ORFs on the genomic contigs. We inferred a maximum likelihood phylogenetic tree of tenascin (adjacent to the ITGB gene on CBKX010020318.1) and PA14 (adjacent to the ITGB gene on CBKX010020319.1), we used same conditions as the integrin phylogenetic trees (see below).

QUANTIFICATION AND STATISTICAL ANALYSIS

Phylogenetic trees

To examine the evolutionary relationships of ITGA and ITGB within Amorphea, protein sequences were aligned by MAFFT-LINSI with the parameters “-maxiterate 1000” and “-local pair. Ambiguous sites were trimmed from the alignments using BMGE⁶⁷ with a gap penalty of 0.8. Maximum likelihood (ML) trees were inferred from these trimmed alignments in IQTREE v 1.5.5⁶⁸ under the LG model with the C60 series model of site heterogeneity. Each tree is ML bootstrapped (MLBS) by 1,000 pseudoreplicates. A custom PYTHON script implementing ETE-TOOLKIT (www.etetoolkit.org) was used to map protein domain architectures (INTERPROSCAN domain IDs) onto the resulting trees.