Ancient origin of the integrin-mediated adhesion and signaling machinery

Arnau Sebé-Pedrós[a], Andrew J. Roger[b], Franz B. Lang[c], Nicole King[d], and Iñaki Ruiz-Trillo[a,e,1]

[a]Departament de Genètica and Institut de Recerca en Biodiversitat, Universitat de Barcelona, 08028 Barcelona, Spain; [b]Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada B3H 1X5; [c]Department of Biochemistry, Université de Montréal, Montréal, QC, Canada H3C 3J7; [d]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; and [e]Institució Catalana per a la Recerca i Estudis Avançats at Parc Científic de Barcelona, 08028 Barcelona, Spain

Edited* by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved April 28, 2010 (received for review February 24, 2010)

The evolution of animals (metazoans) from their unicellular ancestors required the emergence of novel mechanisms for cell adhesion and cell–cell communication. One of the most important cell adhesion mechanisms for metazoan development is integrin-mediated adhesion and signaling. The integrin adhesion complex mediates critical interactions between cells and the extracellular matrix, modulating several aspects of cell physiology. To date this machinery has been considered strictly metazoan specific. Here we report the results of a comparative genomic analysis of the integrin adhesion machinery, using genomic data from several unicellular relatives of Metazoa and Fungi. Unexpectedly, we found that core components of the integrin adhesion complex are encoded in the genome of the apusozoan protist *Amastigomonas* sp., and therefore their origins predate the divergence of Opisthokonta, the clade that includes metazoans and fungi. Furthermore, our analyses suggest that key components of this apparatus have been lost independently in fungi and choanoflagellates. Our data highlight the fact that many of the key genes that had formerly been cited as crucial for metazoan origins have a much earlier origin. This underscores the importance of gene cooption in the unicellular-to-multicellular transition that led to the emergence of the Metazoa.

cell adhesion | lateral gene transfer | metazoan origins | multicellularity

Little is known about how multicellular animals (metazoans) or fungi evolved from their single-celled or colonial ancestors. Cell adhesion and cell signaling are two important features of the multicellular metazoan lifestyle that were likely critical to the origin of Metazoa (1, 2). Recent data have shown that many of the major metazoan signaling pathways and cell adhesion systems are ubiquitous across the metazoan kingdom, including nonbilaterian lineages [sponges, placozoans, and cnidarians (3–6)]. These findings indicate that cell adhesion and cell signaling genes might have evolved before the origin of Metazoa. Consistent with this view, choanoflagellates, the unicellular putative sister group of Metazoa (7–11), have been shown to possess some genes involved in cell signaling and adhesion, such as tyrosine kinases and cadherins (1, 12–14). Expressed sequence tag surveys of other unicellular relatives of metazoans, such as *Capsaspora owczarzaki* and *Ministeria vibrans*, also yielded homologs of genes involved in metazoan cell adhesion and cell signaling (9, 15).

Here we report a comparative genomic survey of integrin-mediated adhesion machinery, a critical cell–matrix adhesion mechanism in metazoans that also plays a vital role in cell signaling (16–18). Integrin-mediated signaling occurs in two ways: as an "inside-out" signaling modulated through intracellular events, and as "outside-in" signaling that reacts via binding of a ligand to the receptor (17, 19, 20). Thus, integrins are involved in diverse cellular processes, including embryogenesis, cell spreading, cell migration, and proliferation (16–18). However, integrin adhesion and signaling seems to be absent from other multicellular organisms (e.g., plants and fungi) and is generally considered to be metazoan specific (2, 5, 21).

Integrins are heterodimeric transmembrane proteins composed of one α and one β subunit (17). The integrin-mediated process of linking the extracellular matrix to the intracellular actin cytoskeleton is made in concert with several cytoskeletal proteins that form adhesion-triggered signaling complexes (22): α-actinin and talin [both of which directly bind to the integrin β subunit (23–25)]; and paxillin and vinculin [both of which are scaffolding proteins that indirectly bind to integrin-β via talin and α-actinin (26, 27)]. An important element of the integrin adhesion machinery is the heterotrimer IPP complex, which is composed of ILK (integrin-linked kinase), PINCH (particularly interesting Cys-His–rich protein), and parvin (28, 29). This complex plays an important role in integrin-mediated signaling, regulating apoptosis, and cell dynamics (29). Finally, integrin-mediated signaling occurs mainly via two kinases known to be concentrated at the integrin adhesion machinery, namely c-Src tyrosine kinase and FAK (focal adhesion kinase) (22, 30, 31). Many other proteins are indirectly involved with the integrin adhesion complex (32), but here we focus on those most directly involved in the clustering of integrins into the adhesion complex (22).

The recent completion of genome sequences for five close relatives (some strictly unicellular, some colonial) of metazoans and fungi provides the opportunity to reconstruct the evolution of proteins required for integrin-mediated cell adhesion (ref. 33; see also http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html). By examining the genomes of the amoeba *C. owczarzaki*, two basal fungi, *Allomyces macrogynus* and *Spizellomyces punctatus*, the apusozoan *Amastigomonas* sp., and a choanoflagellate, *Proterospongia* sp., we find that the integrin adhesion and signaling machinery evolved in unicellular progenitors of apusozoan protists and opisthokonts (i.e., Fungi, choanoflagellates, and Metazoa). Integrin α and β and several other components of the integrin adhesion complex are absent from choanoflagellates and fungi and were presumably lost independently in these lineages. By comparing genome data from a broad sampling of unicellular taxa, we have been able to clarify the dynamic evolutionary history of the integrin adhesion complex.

Results

Integrins. Outside Metazoa, we found four integrin β and four integrin α genes in *C. owczarzaki*, and one β and one α in *Amastigomonas* sp. Interestingly, we found one of the integrin β domains (the extracellular domain) in the cyanobacterium *Trichodesmium*

Author contributions: A.S.-P. performed research; A.S.-P., F.B.L., and I.R.-T. analyzed data; A.J.R., F.B.L., and N.K. contributed new reagents/analytic tools; I.R.-T. designed research; and A.S.-P., A.J.R., F.B.L., N.K., and I.R.-T. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GU320672–GU320675).

[1]To whom correspondence should be addressed. E-mail: inaki.ruiz@icrea.es.

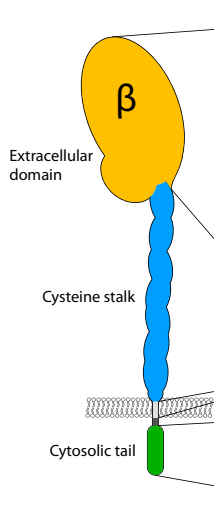
This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002257107/-/DCSupplemental.

erythraeum (34), which lacks all of the other integrin β domains. Integrins were not detected in any other examined eukaryote. Interestingly, although an integrin α ortholog was thought to be present in the choanoflagellate *M. brevicollis* (12), we failed to detect a bona fide integrin α in either *Monosiga brevicollis* or *Proterospongia* sp. The putative integrin α from *M. brevicollis* (XP_001749484) did not pass any of our criteria (for example, reverse blast did not give integrin α hits; *Methods*). The *M. brevicollis* gene XP_001749484 shares with integrin α homologs the presence of some FG-GAP repeats domains, which are not specific to integrin α and are found in other nonintegrin proteins. A phylogeny made from FG-GAP repeats shows the *M. brevicollis* putative integrin α homolog clustering with nonintegrin bacterial proteins but not with integrin α (Fig. S1). On the other hand, our phylogenetic analysis of integrin β (Fig. S2) shows the four *C. owczarzaki* integrins clustering together with a bootstrap value (BV) of 60% (BV = 95% if the most divergent *C. owczarzaki* integrin β homolog is deleted from the analysis). The integrin β homologs of *Amastigomonas* sp. and *T. erythraeum* group together, with a BV of 85%. We were unable to recover any other integrin adhesion complex components in *Trichodesmium erythraeum*, and no other sequenced bacterial genome encodes the integrin β domain or any other component of the integrin adhesion complex.

We next analyzed whether β integrins from *C. owczarzaki* and *Amastigomonas* sp. and the integrin β extracellular domain from *T. erythraeum* have conserved the functional domains and motifs present in metazoan integrins (Fig. 1 and Fig. S3). The cation-binding motifs MIDAS, ADMIDAS, and LIMB, which are located in the extracellular domain (35, 36), are well conserved in the different nonmetazoan integrin β, except for *C. owczarzaki* integrin β4 (Fig. 1*A*). Moreover, *C. owczarzaki* integrin β1, β2, and β3 and *Amastigomonas* sp. integrin β have a clear expansion of the cysteine-rich stalk (Fig. 1*A* and Fig. S3), which accounts for their longer size relative to metazoan integrin β proteins. Other key motifs in metazoan integrin β proteins are the cytoplasmic integrin α-interacting motif and the NPXY motif, which plays a key role in protein interactions (17, 20, 37, 38). Both motifs are well conserved in *C. owczarzaki* integrin β1–β3 and *Amastigomonas* sp. integrin β (Fig. 1*A*). Finally, both *C. owczarzaki* and *Amastigomonas* sp. integrin β have predicted signal peptides and transmembrane domains.

We also examined the evolutionary conservation of nonmetazoan integrin α (Fig. 1*B*). Metazoan integrin α homologs typically have large extracellular regions with seven FG-GAP repeats that form a β propeller structure (36), with three DXD/NXD/NXXXD cation-binding motifs in the last three FG-GAP repeats (39). One specific and diagnostic feature of integrin α is

A Integrin β



Motif/Function	Metazoa	*Capsaspora β1*	*Capsaspora β2*	*Capsaspora β3*	*Capsaspora β4*	*Amastigomonas*	*Trichodesmium*
MIDAS Mg^{2+} and Mn^{2+} cations binding.	D	D^{119}	D^{123}	D^{238}	D^{128}	D^{115}	D^{59}
	S	S^{121}	S^{125}	S^{240}	S^{130}	S^{117}	S^{61}
	S	S^{123}	S^{127}	T^{242}	G^{132}	S^{119}	S^{63}
	E	E^{232}	E^{220}	N^{342}	N^{220}	E^{214}	E^{142}
	D	D^{268}	D^{248}	D^{370}	D^{252}	D^{245}	D^{173}
ADMIDAS Ca^{2+} cations binding.	S	S^{123}	S^{127}	T^{242}	G^{132}	S^{119}	S^{63}
	D	D^{126}	D^{130}	D^{245}	S^{135}	D^{122}	D^{66}
	D	D^{127}	D^{131}	D^{246}	D^{136}	D^{123}	D^{67}
	A	A^{357}	A^{338}	A^{454}	A^{338}	F^{330}	F^{246}
	D	D^{268}	D^{248}	D^{370}	D^{252}	D^{245}	D^{173}
LIMB Ca^{2+} cations binding.	D/E	E^{169}	D^{170}	E^{288}	E^{167}	D^{155}	D^{98}
	N	N^{227}	N^{216}	N^{338}	D^{215}	N^{209}	N^{138}
	D	D^{229}	D^{217}	D^{339}	D^{217}	D^{211}	D^{139}
	P	P^{231}	P^{219}	P^{341}	P^{219}	P^{213}	P^{141}
	E	E^{232}	E^{220}	N^{342}	N^{220}	E^{214}	E^{142}
CXCXXCXC	3	9	6	7	2	7	
KLLXXX**D** Interaction with integrin α.	1	1	1	1	0	1	
NPXY Interaction with intracellular scaffolding and signaling proteins.	1-2	1	1	1	0	2	

B Integrin α

Motif/Position/Function	Metazoa	*Capsaspora α1*	*Capsaspora α2*	*Capsaspora α3*	*Capsaspora α4*	*Amastigomonas*
DXD/NXD/NXXXD Extracellular. Cations binding.	3	3	3	3	1	3
KXG**FFXR** Cytosolic tail. Interaction with integrin β.	1	1	1	1	0	1

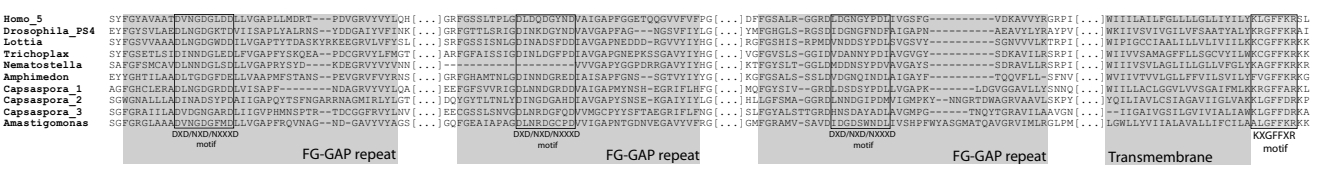


Fig. 1. Comparison of the functional domains and amino acid motifs between canonical metazoan and nonmetazoan integrins. (*A*) Integrin β amino acid motifs and (*B*) amino acid motifs and schematical alignment of integrin α (20, 35–39). Integrin β is divided into the integrin β extracellular domain (orange), integrin stalk (blue), interaction motif (gray), and cytoplasmic tail (green). The *T. erytrhaeum* homolog possesses only the integrin β domain. For the MIDAS, ADMIDAS, and LIMB motifs within the integrin β domain, we indicate the canonical metazoan amino acids and their positions within the whole protein. In red are shown nonconserved position and motifs. For the remaining motifs, the number of motifs present is indicated. Integrin α is divided into the three FG-GAP cation binding motifs and the integrin α–β interacting motif after the transmembrane domain. Key amino acids experimentally determined in the α–β interacting motifs are depicted in bold (17).

the short cytoplasmic tail that contains a KXGFFXR motif that interacts with integrin β. A signal peptide and a transmembrane domain are also typically found in integrin α homologs. These motifs are conserved, with some minor modifications, in the integrin α homologs of *C. owczarzaki* and *Amastigomonas* sp., but not in *C. owczarzaki* integrin α4, which has only two of the three cation-binding motifs and does not have a predicted signal peptide (Fig. 1*B*).

Scaffolding Proteins. Our investigations show that all scaffolding proteins involved in the integrin adhesion apparatus (that is, α-actinin, vinculin, paxillin, and talin) are common among unikonts (i.e., Opisthokonts+Amoebozoa; Figs. 2 and 3). Phylogenetic analyses of these proteins show, in general, topologies in agreement with organismal phylogeny (Figs. S4*A* and S5 *A* and *B*), except for paxillin, which did not have enough phylogenetic signal to recover a statistically significant topology.

IPP Complex. A complete IPP complex with all three components is only present in Metazoa, *C. owczarzaki*, the chytrid fungus *Batrachochytrium dendrobatidis*, and the apusozoan *Amastigomonas* sp. (Figs. 2 and 3). In the *Amastigomonas* sp. genome we found a partial gene encoding the N-terminal part of the ILK protein, which is composed of the three consecutive ankyrin repeats but failed to find a characteristic C terminus, which is a Ser/Thr kinase domain. Phylogenetic inference based on an alignment of the three ankyrin repeats shows that the putative ILK of *Amastigomonas* sp. branches within the canonical ILK homologs (Fig. S5*C*). Because the genome coverage of *Amastigomonas* sp. is at present still low, it is possible that the C-terminal part of *Amastigomonas* sp. ILK homolog is indeed present but not represented in the current assembly. In any case, the three components of the IPP complex are missing in both of the choanoflagellates, *M. brevicollis* and *Proterospongia* sp., and fungi other than *B. dendrobatidis*. Interestingly, *A. macrogynus* and *S. punctatus* possess just one component (PINCH and ILK, respectively) of the IPP complex (Figs. 2 and 3). A phylogenetic tree of ILK and several related kinases estimated from an alignment of the kinase domain alone shows that *C. owczarzaki*, *S. punctatus*, and *B. dendrobatidis* ILKs are related to metazoan ILKs (Fig. S5*D*). Similarly, phylogenetic trees of parvin and PINCH show a topology in agreement with organismal phylogeny (Fig. S6).

c-Src Tyrosine Kinase and FAK. Our searches show that c-Src is present in Metazoa, choanoflagellates, and *C. owczarzaki* (Fig. 2). The phylogenetic analysis of this protein family, which includes Abl kinases as an outgroup, shows that both choanoflagellates and *C. owczarzaki* c-Src tyrosine kinases group with metazoan ones (Fig. S4*B*). On the other hand, bona fide FAK are only present in Metazoa and *C. owczarzaki* (Fig. 2). *C. owczarzaki* FAK have all of the functional domains involved in its protein–protein interactions (31) (Fig. S7*A*). Interestingly, *M. brevicollis* has a gene encoding a tyrosine kinase domain that, by phylogenetic analysis, seems to be related to FAK (Fig. S7*B*), even though the predicted protein does not have the canonical domain structure of FAK.

Discussion

Our analyses show that the integrin-mediated cell adhesion machinery is not specific to metazoans, as previously thought (2, 5, 21). We found that the apusozoan *Amastigomonas* sp. has the integrin adhesion machinery, including all of the components of the canonical metazoan complex, except for the signaling molecules FAK and c-Src. Recent multigene analyses suggest that apusozoans are related to opisthokonts, most likely falling outside of this clade as their nearest sister group (8, 40, 41). However, they have also been proposed to be sister group to amoebozoans or represent a deeper-branching eukaryotic lineage, although these proposals derive from single-gene and statistically weakly supported phylogenies (see ref. 41 for a discussion). In any case, apusozoans clearly fall outside opisthokonts in all multigene phylogenetic analyses, and they do not share the characteristic translation elongation factor 1-α (EF1-α) insertion, a synapomorphy unique to the opisthokont lineages (see ref. 8 for *Ancyromonas* and *Apusomonas* and Fig. S8 for *Amastigomonas*). Regardless of whether apusozoans are (*i*) sister group to opisthokonts (40–42), (*ii*) sister group to amoebozans (41), or (*iii*) a deep eukaryotic lineage (41), our conclusion that many of the components of the integrin adhesome evolved well before the origin of Metazoa and Fungi is still valid. If apusozoans are sister group to amoebozoans or a deeper eukaryotic lineage, then these core integrin compo-

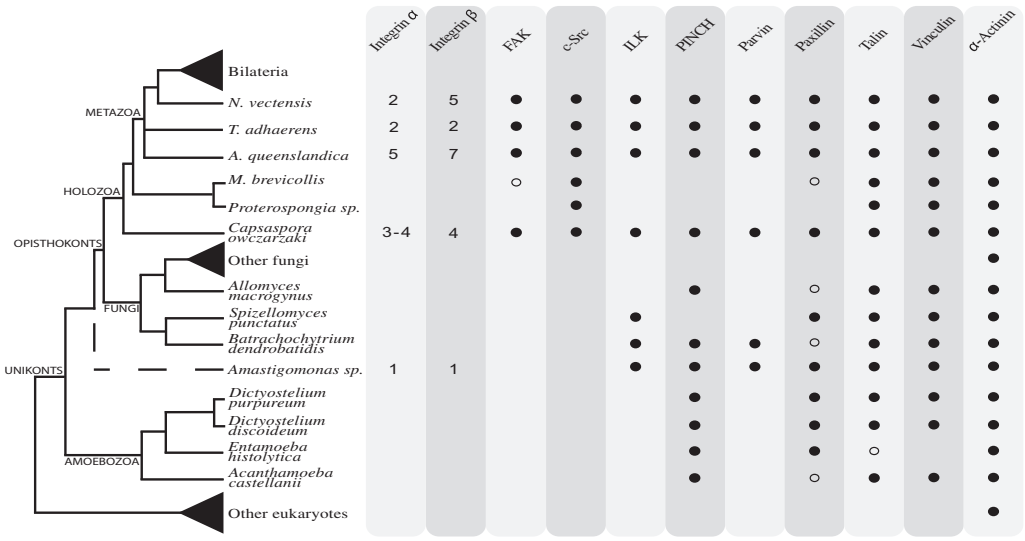


Fig. 2. Schematic representation of the eukaryotic tree of life showing the distribution of the different components of the integrin adhesion complex. The number of integrin homologs is shown. A black dot indicates the presence of clear homologs, whereas a hollow dot indicates the presence of putative or degenerate homologs. Absence of a dot indicates that a homolog is lacking in that taxon. The phylogenetic relationships are based in several recent phylogenetic studies (8, 9, 15, 40, 64, 65).

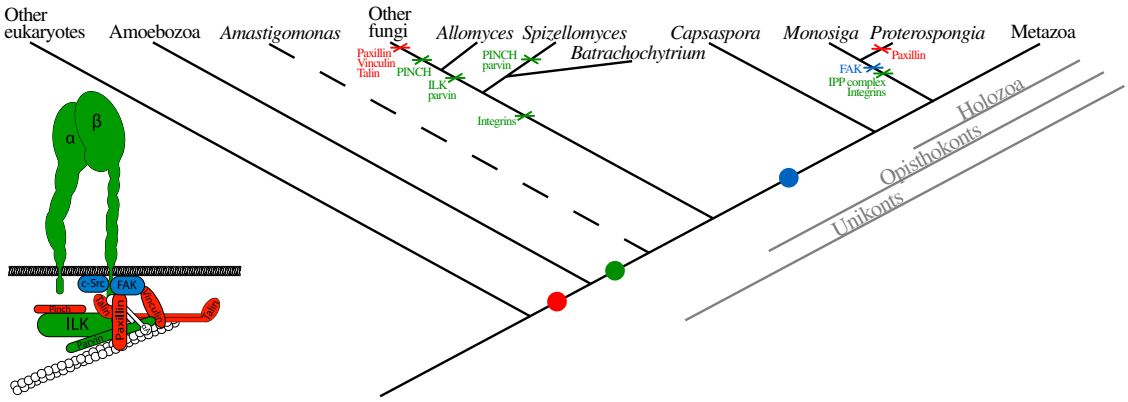


Fig. 3. Schematic representation of integrin-mediated cell-adhesion and cell-signaling evolution. *Left:* The canonical metazoan integrin adhesion complex. The colors correspond to the three main steps in the evolution of the integrin adhesion mechanism, as shown in the cladogram. Dots indicate origin, and crosses indicate losses. The branch leading to *Amastigomonas* sp. is shown dashed, because its phylogenetic position remains unresolved (see main text for discussion).

nents must also have been secondarily lost in the amoebozoan taxa whose genomes have been characterized to date.

We have also shown that *C. owczarzaki*, a specific sister group to choanoflagellates and Metazoa (9, 15, 43), has a canonical metazoan-type integrin adhesion and signaling machinery with a full repertoire of integrin adhesion complex components. Therefore, the canonical metazoan-type integrin adhesion machinery is probably specific to holozoans; that is, it originated before the divergence of *C. owczarzaki* from choanoflagellates+Metazoa, but likely after the Fungi+nucleariid+fonticulid clade had split from Holozoa (Fig. 3). Another possible scenario is that a canonical metazoan-type integrin adhesion machinery was present in the ancestor of apusozoans and opisthokonts, and both FAK and c-Src were subsequently lost within the apusozoan lineage.

A major implication of our taxon-rich comparative genomic survey is the emergence of a more complex scenario for the evolution of the integrin adhesion complex (Figs. 2 and 3). Under this scenario, which should be further tested with genome data from additional eukaryotic lineages, it is evident that several independent losses and diversifications of main components of the integrin adhesome have occurred over the course of evolution. For example, integrin α and β homologs seem to be absent from choanoflagellates and fungi. In fact, each choanoflagellate and fungal taxon we examined harbors a distinctive repertoire of integrin adhesome components resulting from different lineage-specific losses. This is most obvious in Fungi, where the loss of the IPP complex seems to be gradual, with the chytrid fungi taxa retaining all or some of the IPP components despite their lack of integrins. Specifically, *B. dendrobatidis* has the full IPP complex (ILK, PINCH, and parvin), whereas *S. punctatus* has just one of the components (ILK), and *A. macrogynus* has just PINCH. It is unclear what cellular functions the IPP complex components present in *B. dendrobatidis*, *S. punctatus*, and *A. macrogynus* might have in the absence of integrin subunits.

Of major interest for the origin of metazoans is the fact that choanoflagellates, which are the closest sister group of Metazoa (7, 8, 10, 11), have also lost many of the integrin components. Specifically, the two choanoflagellates analyzed here, *M. brevicollis* and *Proterospongia* sp., lack both integrin β and α, the full IPP complex, and one of the signaling molecules involved in the integrin adhesome, FAK (although *M. brevicollis* has a protein with a FAK-related tyrosine kinase domain; Fig. S7). Choanoflagellates do have c-Src, but they act in a different context than that of integrin adhesion, as recently demonstrated experimentally in *M. brevicollis* (44). Moreover, lineage-specific diversifications (independent of those occurring in metazoans) of both integrin α and β have occurred within the *C. owczarzaki* lineage. What roles these various homologs play in *C. owczarzaki* biology remains to be determined.

It is possible that functional differences between metazoan and nonmetazoan integrins exist that would explain their conservation in unicellular vs. multicellular contexts. Our analysis of the functional domains shows that both metazoan and nonmetazoan integrins are quite similar, the only difference being the longer size of the protein in nonmetazoan ones (Fig. 1 and Fig. S3). More importantly, both integrin α and β in *C. owczarzaki* and in *Amastigomonas* sp. possess all of the critical interacting amino acid motifs in their cytoplasmic tails (Fig. 1). Thus, we can assume that they too work as heterodimers and that they interact and function similarly to metazoan homologs. Functional analysis will be needed to test this hypothesis.

Many of the scaffolding proteins (talin, vinculin, paxillin) most likely evolved in the common ancestor of amoebozoans and opisthokonts, where they had ancestrally different functions (as in present day amoebozoans). They were coopted to further work as part of a metazoan-specific integrin adhesome (i.e., their presence in opisthokonts and in amoeboans should not be interpreted as a signature of an ancient integrin-mediated adhesion apparatus) (Fig. 3). For example, it has been shown that the talin homolog of the amoebozoan *D. discoideum* interacts with an NPXY motif (the same motif found in integrin β) of the cytoplasmic tail of an adhesion molecule called SibA (45). Thus, it is possible that an ancestral integrin β independently acquired an NPXY motif allowing it to recruit talin.

Our comparative genomic study not only deciphers the evolutionary history of the integrin adhesome, but it also highlights the importance of a broad taxonomic sampling in these kinds of studies. In particular, a broader taxonomic sampling within nonbilaterian metazoans was key for the realization that many key genes in bilaterian development are indeed present in triploblastic metazoans (3–6, 46–49). Similarly, genome data from unicellular metazoan-related lineages is pushing back the times of origin of many gene families formerly believed to be metazoan specific to well into the Proterozoic. Such is the case, for example, of tyrosine kinases (14, 50, 51), some transcription factors (12, 52), membrane-associated guanylate kinases (53), or cadherines (13). The integrin-mediated signaling and adhesion machinery here presented add another striking example to this pattern and suggest that some of these protein families may have emerged even earlier in eukaryote evolution before the divergence of opisthokonts. Investigation of a variety of additional genomes from unicellular opisthokonts and other more distantly related protistan

lineages will be required to more precisely pinpoint the origins of these systems in early eukaryote evolution.

Integrin β in the Cyanobacterium *T. erythraeum* Is Derived from a Lateral Gene Transfer Event. Our search revealed the presence of a gene encoding an incomplete integrin β in the *T. erythraeum*. We believe the most plausible scenario to explain this observation is an interdomain lateral gene transfer (LGT) event in the eukaryote-to-prokaryote direction, because integrin β is present in many eukaryote taxa but only in a single known prokaryotic genome. The lack of introns in the *Amastigomonas* integrin β (in contrast to the other integrins described herein) may have facilitated its integration into a cyanobacterial genome as would the property of natural competence (i.e., the ability to take up DNA) known in Cyanobacteria (54). Even though eukaryote-to-prokaryote LGT events are not as common as LGTs in the opposite direction, other cases have been described in *T. erythraeum* (55).

Conclusions

We have demonstrated that a near-complete integrin adhesion complex had evolved in a unicellular common ancestor of metazoans and fungi and still exists in the Apusozoa, the putative sister group to opisthokonts. Furthermore, we have shown that the origins of most of the scaffolding elements of current integrin adhesion complex predate the origins of integrin proteins themselves, suggesting that an ancient scaffolding machinery was coopted to the integrin adhesion system. Moreover, the origin of the IPP complex probably represented one of the first signaling modules, coupling the integrin adhesion machinery with cell signaling to control cell behavior. Novel signaling systems based on tyrosine kinases appeared at a later stage, most likely within holozoans. Another implication of our analyses is that lineage-specific diversifications and lineage-specific losses have played a major role in the evolution of the integrin adhesome in opisthokonts. For example, both fungi and choanoflagellates have lost several important components of the integrin adhesion complex from their ancestors. Finally, from our study and that done by Abedin and King (13), we can conclude that the major cell–cell and cell–matrix adhesion mechanisms in metazoans, those mediated by cadherins and integrins, respectively, have a deeper evolutionary origin than previously thought. This adds to the growing evidence that major cell signaling and cell adhesion pathways crucial to metazoan development were present in premetazoan lineages (12, 50, 51, 53). Thus, the answers to what triggered the unicellular-to-multicellular transition that gave rise to metazoans may lie not only in the acquisition of new genes but also in the cooption of ancestral proteins into new functions and the evolution of more complex interactions.

Methods

Gene Searches. We performed searches for the two integrin subunits (α and β) plus all of the other proteins that are directly involved in the integrin-mediated adhesion and signaling complex (see the Introduction). Those proteins include α-actinin, vinculin, talin, paxillin, ILK, PINCH, parvin, FAK, and c-Src. A primary search to collect putative initial candidates was performed using the basic local alignment sequence tool (BLAST: blastp and tblastn) using *Homo sapiens* integrin adhesion proteins as queries and an e-value threshold of 10^{-05}. We blasted against completed or ongoing genome project databases at the National Center for Biotechnology Information (NCBI), the Joint Genome Institute, and the Broad Institute (see Fig. 2 for a list of the taxa considered), as well as against the *Amphimedon queenslandica* protein and genome database (Dr. Bernard M. Degnan). *C. owczarzaki* and *S. punctatus* genome assemblies and annotations are available at the Broad Institute Web site (http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html). In the case of *Proterospongia* sp., *Amastigomonas* sp., *A. macrogynus*, and *Acanthamoeba castellanii*, we assembled the trace data using the WGS assembler. We then annotated the genes of interest using both Genomescan (56) and Augustus (57) and performed local BLAST searches against both annotations. Assemblies and annotations for these taxa are available upon request (*Appendix S1*).

When the BLAST searches of genome data described above returned significant "hits", the sequences obtained were then reciprocally searched against the NCBI protein database by BLAST to confirm the validity of the sequences retrieved with the initial search (58). To identify distant homologs that might have escaped these simple searches, two additional methods were used. The same BLAST search was repeated using homologs from nonmetazoan taxa, such as *Dictyostelium*, *Capsaspora*, or *Amastigomonas*, as queries instead of *Homo* sequences. Additionally, for integrin β, integrin α, vinculin, talin, and FAK, we performed protein domain searches using HMMER3.0b2 (59) against the same genome databases, plus six-frame translations of all studied genomes. Finally, we checked the protein domain structure of all putative positives by searching the Pfam (http://pfam.sanger.ac.uk/search) and SMART (http://smart.embl-heidelberg.de/) databases. Signal peptides were identified using the SignalP 3.0 Server (60).

Confirmation of *C. owczarzaki* Integrin β by PCR. We confirmed the presence of integrin β in *C. owczarzaki* by RT-PCR and 3′ RACE PCR. mRNA was extracted using a Dynabeads mRNA purification kit (Invitrogen), and subsequent RT-PCR was performed using SuperScript III First Strand Synthesis kit (Invitrogen). The full sequences of the 5′ and 3′ ends of the four distinct *C. owczarzaki* integrin β cDNAs were obtained by RACE, using nested PCR with primers designed from initial analyses of the genome data. Both coding and noncoding strands were sequenced using an ABI Prism BigDye Termination Cycle Sequencing Kit (Applied Biosystems). New sequences were deposited in GenBank under accession nos. GU320672–GU320675.

Phylogenetic Analyses. Alignments were constructed for all proteins using the Muscle (61) plug-in of Geneious software (Biomatters), which were then manually inspected and edited. Only those species and those positions that were unambiguously aligned were included in the final phylogenetic analyses. Maximum likelihood (ML) phylogenetic trees were estimated by RaxML (62) using the PROTGAMMAWAGI model, which uses the Whelan and Goldman amino acid exchangeabilities and accounts for among-site rate variation with a four-category discrete gamma approximation and a proportion of invariable sites (WAG+ Γ+I). Statistical support for bipartitions was estimated by performing 100-bootstrap replicates using RaxML and the same model.

Bayesian analyses were performed with MrBayes 3.1 (63), using the WAG+Γ+I model of evolution, with four chains, a subsampling frequency of 100, and two parallel runs. Runs were stopped when the average SD of split frequencies of the two parallel runs was <0.01, usually around 1,000,000 generations. The two LnL graphs were checked and an appropriate burn-in length established; stationarity of the chain typically occurred after ≈15% of the generations. Bayesian posterior probabilities were used for assessing the confidence values of each bipartition.

ACKNOWLEDGMENTS. We thank the Joint Genome Institute (JGI), the Broad Institute, and the Baylor College of Medicine (BCM) for making data publicly available; Bernard Degnan for access to the *A. queenslandica* genome data; Jason Stajich for sharing unpublished *B. dendrobatidis* genome data; Kim C. Worley and the team of the *A. castellanii* genome project for accession to the genome data; Manuel Palacín, Romain Derelle, Alex de Mendoza, and Hiroshi Suga for helpful insights; and other members of UNICORN, Gertraud Burger, Michael W. Gray, and Peter W. H. Holland. Preliminary sequence data were obtained from the JGI, the Broad Institute, BCM, and National Center for Biotechnology Information Web sites. The genome sequences of *C.owczarzaki*, *A. macrogynus*, *S. punctatus, Amastigomonas* sp., and *Proterospongia* sp. are being determined by the Broad Institute of Massachusetts Institute of Technology/Harvard University under the auspices of the National Human Genome Research Institute and within the UNICORN initiative. This work was supported by an Institució Catalana per a la Recerca i Estudis Avançats contract, European Research Council Starting Grant 206883, and Grant BFU2008-02839/BMC from Ministerio de Ciencia e Innovación (MICINN) (to I.R.-T.). A.S.-P.'s salary was supported by a pregraduate Formación de Personal Universitario grant from MICINN. A.J.R.'s contribution was supported by Grant MOP 62809 from the Canadian Institutes of Health Research, and B.F.L.'s contribution by the Canadian Research Chair Program.

1. King N (2004) The unicellular ancestry of animal development. *Dev Cell* 7:313–325.
2. Rokas A (2008) The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet* 42:235–251.
3. Putnam NH, et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86–94.
4. Srivastava M, et al. (2008) The Trichoplax genome and the nature of placozoans. *Nature* 454:955–960.

5. Nichols SA, Dirks W, Pearse JS, King N (2006) Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci USA* 103:12451–12456.
6. Larroux C, et al. (2008) Genesis and expansion of metazoan transcription factor gene classes. *Mol Biol Evol* 25:980–996.
7. Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G (2002) The closest unicellular relatives of animals. *Curr Biol* 12:1773–1778.
8. Steenkamp ET, Wright J, Baldauf SL (2006) The protistan origins of animals and fungi. *Mol Biol Evol* 23:93–106.
9. Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF (2008) A phylogenomic investigation into the origin of metazoa. *Mol Biol Evol* 25:664–672.
10. Ruiz-Trillo I, Lane CE, Archibald JM, Roger AJ (2006) Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts *Capsaspora owczarzaki* and *Sphaeroforma arctica*. *J Eukaryot Microbiol* 53:1–6.
11. Carr M, Leadbeater BS, Hassan R, Nelson M, Baldauf SL (2008) Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc Natl Acad Sci USA* 105:16641–16646.
12. King N, et al. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788.
13. Abedin M, King N (2008) The premetazoan ancestry of cadherins. *Science* 319: 946–948.
14. Suga H, et al. (2008) Ancient divergence of animal protein tyrosine kinase genes demonstrated by a gene family tree including choanoflagellate genes. *FEBS Lett* 582: 815–818.
15. Shalchian-Tabrizi K, et al. (2008) Multigene phylogeny of choanozoa and the origin of animals. *PLoS ONE* 3:e2098.
16. Hynes RO (1992) Integrins: Versatility, modulation, and signaling in cell adhesion. *Cell* 69:11–25.
17. Hynes RO (2002) Integrins: Bidirectional, allosteric signaling machines. *Cell* 110: 673–687.
18. Harburger DS, Calderwood DA (2009) Integrin signalling at a glance. *J Cell Sci* 122: 159–163.
19. O'Toole TE, et al. (1994) Integrin cytoplasmic domains mediate inside-out signal transduction. *J Cell Biol* 124:1047–1059.
20. Dedhar S, Hannigan GE (1996) Integrin cytoplasmic interactions and bidirectional transmembrane signalling. *Curr Opin Cell Biol* 8:657–669.
21. Whittaker CA, Hynes RO (2002) Distribution and evolution of von Willebrand/integrin A domains: Widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell* 13:3369–3387.
22. LaFlamme SE, Auer KL (1996) Integrin signaling. *Semin Cancer Biol* 7:111–118.
23. Sjöblom B, Salmazo A, Djinović-Carugo K (2008) Alpha-actinin structure and regulation. *Cell Mol Life Sci* 65:2688–2701.
24. Wegener KL, et al. (2007) Structural basis of integrin activation by talin. *Cell* 128: 171–182.
25. Critchley DR (2009) Biochemical and structural properties of the integrin-associated cytoskeletal protein talin. *Annu Rev Biophys* 38:235–254.
26. Ziegler WH, Liddington RC, Critchley DR (2006) The structure and regulation of vinculin. *Trends Cell Biol* 16:453–460.
27. Deakin NO, Turner CE (2008) Paxillin comes of age. *J Cell Sci* 121:2435–2444.
28. Legate KR, Montañez E, Kudlacek O, Fässler R (2006) ILK, PINCH and parvin: The tIPP of integrin signalling. *Nat Rev Mol Cell Biol* 7:20–31.
29. Nikolopoulos SN, Turner CE (2001) Integrin-linked kinase (ILK) binding to paxillin LD1 motif regulates ILK localization to focal adhesions. *J Biol Chem* 276:23499–23505.
30. Arias-Salgado EG, et al. (2003) Src kinase activation by direct interaction with the integrin beta cytoplasmic domain. *Proc Natl Acad Sci USA* 100:13298–13302.
31. Parsons JT, Martin KH, Slack JK, Taylor JM, Weed SA (2000) Focal adhesion kinase: A regulator of focal adhesion dynamics and cell movement. *Oncogene* 19:5606–5613.
32. Zaidel-Bar R (2009) Evolution of complexity in the integrin adhesome. *J Cell Biol* 186: 317–321.
33. Ruiz-Trillo I, et al. (2007) The origins of multicellularity: A multi-taxon genome initiative. *Trends Genet* 23:113–118.
34. Johnson MS, Lu N, Denessiouk K, Heino J, Gullberg D (2009) Integrins during evolution: Evolutionary trees and model organisms. *Biochim Biophys Acta* 1788: 779–789.
35. Valdramidou D, Humphries MJ, Mould AP (2008) Distinct roles of beta1 metal ion-dependent adhesion site (MIDAS), adjacent to MIDAS (ADMIDAS), and ligand-associated metal-binding site (LIMBS) cation-binding sites in ligand recognition by integrin alpha2beta1. *J Biol Chem* 283:32704–32714.
36. Xiong JP, et al. (2001) Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* 294:339–345.
37. Brower DL, Brower SM, Hayward DC, Ball EE (1997) Molecular evolution of integrins: Genes encoding integrin beta subunits from a coral and a sponge. *Proc Natl Acad Sci USA* 94:9182–9187.
38. Tahiliani PD, Singh L, Auer KL, LaFlamme SE (1997) The role of conserved amino acid motifs within the integrin beta3 cytoplasmic domain in triggering focal adhesion kinase phosphorylation. *J Biol Chem* 272:7892–7898.
39. Knack BA, et al. (2008) Unexpected diversity of cnidarian integrins: Expression during coral gastrulation. *BMC Evol Biol* 8:136.
40. Brown MW, Spiegel FW, Silberman JD (2009) Phylogeny of the "forgotten" cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. *Mol Biol Evol* 26:2699–2709.
41. Kim E, Simpson AG, Graham LE (2006) Evolutionary relationships of apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. *Mol Biol Evol* 23: 2455–2466.
42. Cavalier-Smith T, Chao EE (2003) Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *J Mol Evol* 56:540–563.
43. Ruiz-Trillo I, Inagaki Y, Davis LA, Sperstad S, Landfald B, Roger AJ (2004) *Capsaspora owczarzaki* is an independent opisthokont lineage. *Curr Biol* 14(22):R946–947.
44. Li W, Young SL, King N, Miller WT (2008) Signaling properties of a non-metazoan Src kinase and the evolutionary history of Src negative regulation. *J Biol Chem* 283: 15491–15501.
45. Cornillon S, et al. (2006) An adhesion molecule in free-living *Dictyostelium amoebae* with integrin beta features. *EMBO Rep* 7:617–621.
46. Simionato E, et al. (2007) Origin and diversification of the basic helix-loop-helix gene family in metazoans: Insights from comparative genomics. *BMC Evol Biol* 7:33.
47. Gauthier M, Degnan BM (2008) The transcription factor NF-kappaB in the demosponge *Amphimedon queenslandica*: Insights on the evolutionary origin of the Rel homology domain. *Dev Genes Evol* 218:23–32.
48. Technau U, et al. (2005) Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet* 21:633–639.
49. Yamada A, Pang K, Martindale MQ, Tochinai S (2007) Surprisingly complex T-box gene complement in diploblastic metazoans. *Evol Dev* 9:220–230.
50. Pincus D, Letunic I, Bork P, Lim WA (2008) Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. *Proc Natl Acad Sci USA* 105:9680–9684.
51. Manning G, Young SL, Miller WT, Zhai Y (2008) The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc Natl Acad Sci USA* 105:9674–9679.
52. Degnan BM, Vervoort M, Larroux C, Richards GS (2009) Early evolution of metazoan transcription factors. *Curr Opin Genet Dev* 19:591–599.
53. de Mendoza A, Suga H, Ruiz-Trillo I (2010) Evolution of the MAGUK protein gene family in premetazoan lineages. *BMC Evol Biol* 10:93.
54. Johnsborg O, Eldholm V, Håvarstein LS (2007) Natural genetic transformation: Prevalence, mechanisms and function. *Res Microbiol* 158:767–778.
55. Layton BE, et al. (2008) Collagen's triglycine repeat number and phylogeny suggest an interdomain transfer event from a Devonian or Silurian organism into *Trichodesmium erythraeum*. *J Mol Evol* 66:539–554.
56. Yeh RF, Lim LP, Burge CB (2001) Computational inference of homologous gene structures in the human genome. *Genome Res* 11:803–816.
57. Stanke M, et al. (2006) AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34(Web Server issue):W435–W439.
58. Moreno-Hagelsieb G, Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24:319–324.
59. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763.
60. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
61. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
62. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
63. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
64. Minge MA, et al. (2008) Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proc Biol Sci* 276:597–604.
65. Liu Y, et al. (2009) Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. *BMC Evol Biol* 9:272.