Comprehensive identification of SWI/SNF complex subunits underpins deep eukaryotic ancestry and reveals new plant components

Jorge Hernández-García1,7, Borja Diego-Martin1, Peggy Hsuanyu Kuo2, Yasaman Jami-Alahmadi3,4, Ajay A. Vashisht3,4, James Wohlschlegel3,4, Steven E. Jacobsen2,5,6, Miguel A. Blázquez1 & Javier Gallego-Bartolomé1✉

Over millions of years, eukaryotes evolved from unicellular to multicellular organisms with increasingly complex genomes and sophisticated gene expression networks. Consequently, chromatin regulators evolved to support this increased complexity. The ATP-dependent chromatin remodelers of the SWI/SNF family are multiprotein complexes that modulate nucleosome positioning and appear under different configurations, which perform distinct functions. While the composition, architecture, and activity of these subclasses are well understood in a limited number of fungal and animal model organisms, the lack of comprehensive information in other eukaryotic organisms precludes the identification of a reliable evolutionary model of SWI/SNF complexes. Here, we performed a systematic analysis using 36 species from animal, fungal, and plant lineages to assess the conservation of known SWI/SNF subunits across eukaryotes. We identified evolutionary relationships that allowed us to propose the composition of a hypothetical ancestral SWI/SNF complex in the last eukaryotic common ancestor. This last common ancestor appears to have undergone several rounds of lineage-specific subunit gains and losses, shaping the current conformation of the known subclasses in animals and fungi. In addition, our results unravel a plant SWI/SNF complex, reminiscent of the animal BAF subclass, which incorporates a set of plant-specific subunits of still unknown function.
Eukaryotic life has evolved for millions of years, giving rise to a wide diversity of living forms that range from unicellular organisms to complex multicellular species. The evolution of eukaryotes into complex multicellular organisms has been accompanied by the biological deployment of novel and sophisticated mechanisms to support larger genomes, epigenetic regulation, and intricate gene expression networks. Chromatin plays a prominent role in the regulation of transcriptional states, with the basic chromatin unit being the nucleosome, which consists of four histone pairs (H3, H4, H2A, and H2B) associated with ~147 bp of DNA. Nucleosomes affect many aspects of nuclear biology and act as physical barriers to proteins attempting to access genomic DNA. Moreover, nucleosomes also serve as recruitment platforms for diverse proteins and complexes involved in chromatin regulation. Thus, eukaryotes have evolved numerous ATP-dependent chromatin remodeling complexes to precisely control the nucleosome landscape. These multiprotein complexes contain a catalytic subunit that uses ATP hydrolysis to dissociate DNA from histones, as well as a variety of scaffold and regulatory subunits that trigger changes in the position or histone composition of nucleosomes. One of the best-understood chromatin remodelers is the SWI/SNF family which was first identified in yeast and is broadly conserved across eukaryotes.

Malfunction of this remodeler family has a profound impact on important nuclear processes associated with the misregulation of cell differentiation and cancer.

The function, recruitment, and composition of SWI/SNF complexes have been extensively studied in model organisms from yeast to mammals. Detailed characterization of these organisms has revealed that SWI/SNF complexes are predominantly organized in SWI/SNF-BAF and RSC-PBAF subclasses, as well as the non-canonical BAF (ncBAF) subclass that was recently identified in mammals. These subclasses show non-redundant functions and have signature subunits that define each architecture. Moreover, as a consequence of gene duplications in multicellular organisms, multiple SWI/SNF subunit paralogs can be selectively incorporated into these complexes in specific cell types or during certain developmental processes, such as in embryonic stem cells and during neuron differentiation. This diversity increases the number of possible SWI/SNF complex combinations and provides versatility for specific biological functions. It is thought that the composition and combinatorial assembly of SWI/SNF complexes have evolved in multicellular organisms to accommodate the demands of larger genomes, the presence of new epigenetic regulators like H1 and DNA methylation, and the complex transcriptional regulation required for multicellularity.

The composition and architecture of SWI/SNF complexes have been well characterized in S. cerevisiae and a handful of animal model organisms. However, equivalent information in other organisms, including plants, remains significantly limited. This gap in information limits our knowledge of how different configurations of SWI/SNF complexes originated and evolved in different taxa. Given the intimate connection between the architecture of the complexes and their functionality, filling this gap is a critical first step in the understanding of SWI/SNF biological functions across eukaryotes.

Here, we leverage the recent increase in genomic and transcriptomic data to describe the evolutionary history of SWI/SNF complexes in eukaryotes. Our comprehensive SWI/SNF subunit search approach across non-model metazoan and fungal species and plants, coupled with in vivo plant-based experiments, has allowed us to (i) establish the degree of conservation of different known subunits, (ii) identify a hypothetical simple complex present in a last eukaryotic common ancestor (LECA) to delineate a possible evolutionary trajectory of different SWI/SNF complexes, and (iii) identify uncharacterized plant-specific SWI/SNF subunits.

Results
Identification of SWI/SNF subunits across different eukaryotic lineages. To investigate the degree of SWI/SNF conservation among eukaryotes, we developed a primary sequence-dependent approach based on a step-by-step phylogeny-driven search of different SWI/SNF subunits. To unify the names of the different subunits across species we used the established HUGO nomenclature (SMARC) and requested the assignment of new HUGO names for those subunits that previously lacked this nomenclature (Table 1).

Previously characterized SWI/SNF subunit protein sequences from humans, baker’s, and fission yeast were used as starting queries (Table 1 and Supplementary Table 1). We then performed iterative BLASTP/Phmmer searches in the proteomes of phylogenetically-related species. We selected numerous species within Metazoa, Fungi, and Archaeplastida, encompassing a diverse range of species within each lineage phylogeny, with a special focus on the plant kingdom (Supplementary Table 2). We also included two protozoan species with well-annotated genomes: the endo-parasitic kinetoplastid Trypanosoma brucei and the free-living amebozoan Dictyostelium discoideum in which no extensive study of SWI/SNF subunits exists to our knowledge.

Our analysis identified a set of subunits (SMARCA, SMARCB, SMARCC, SMARCD, and SMARCN) that were unambiguously present in all studied lineages (Fig. 1). We, therefore, conclude that these subunits were part of ancient SWI/SNF complexes of the LECA. Interestingly, we find the protozoan T. brucei does not contain orthologs of any of these subunits, suggesting SWI/SNF complexes may be completely absent in this species.

Plant LFR proteins are true SMARC orthologs. Previous studies have proposed a relationship between the metazoan SMARCF1/2 subunits (ARID1 and ARID2) and the fungal Swi1/Rsc9 proteins. In animals, SMARCF1 proteins contain a signature amino-terminal AT-rich interaction domain (ARID) domain, followed by an Armadillo fold (ARM). Animal SMARCF2s are characterized by a similar N-terminal structure, including an ARID and ARM domain followed by an RFX DNA-binding domain (DBD), and a C-terminal zinc finger (Fig. 2a). We find that these features are conserved in SMARCF1 and SMARCF2 orthologs in most opisthokonts and that the lack of an automatically detectable ARM domain in S. cerevisiae Swi1 and Rsc9 is an exception among fungi (Fig. 2a and Supplementary Fig. 1a). Moreover, ARM and ARID domain phylogenetic analyses support a common evolutionary origin for SMARCF1 and SMARCF2 in animals and fungi (Supplementary Fig. 1b, c), with species-specific motif loss.

Our analyses also suggest A. thaliana LFR protein and its plant orthologs share a common ancestry with opisthokonts SMARC proteins since all of them contain a phylogenetically-related ARM domain. Additionally, we find some chlorophyta algae display a conserved ARID domain in their LFR orthologs (Supplementary Fig. 1a, c). While SMARC orthologs appear absent in some algal species, this result may be due to lineage-specific losses or poor genome annotation. Additionally, red algae also contain the ARID-ARM architecture in their SWI/SNF homologs. Therefore, we propose that a SMARCf subunit composed of ARID and ARM domains was also part of an ancestral SWI/SNF complex, and has been maintained in the major lineages of eukaryotes, with limited losses in specific lineages. Lastly, a SMARCf duplication in an opisthokont common ancestor and the introduction of an RFX DBD domain was associated with the appearance of the
Table 1 SWI/SNF complex components in reference species.

| SMARC name   | Metazoa (human) | Fungi (baker’s yeast/other fungi) | Plants (Arabidopsis) |
|--------------|----------------|-----------------------------------|----------------------|
| SMARCA2/4    | BRM, BRG1      | Snf2, Sth1                         | BRM, SYD, CHR12, CHR23 |
| SMARCB       | BAF47          | Snf5, Sth1                         | BSH                  |
| SMARCC       | BAF155, BAF170 | Swi3, Rsc8                         | SWI3A, SWI3B, SWI3C, SWI3D |
| SMARCD       | BAF60A, BAF60B, BAF60C | Snf12, Rsc6 | SWP73A, SWP73B |
| SMARCEa      | BAF57          | Ssr4 (S. pombe, Q9P7Y0; Gonapodya prolifera, A0A139AVV0) | - |
| SMARCEb      | ARID1A, ARID1B, ARID2 | Swi1, Rsc9 | LFR |
| SMARCH       | DPF1, DPF2, DPF3, PHF10 | Swp82, Rsc7(Npl6) | TPF1, TPF2 |
| SMARCI       | BRD7, BRD9     | Rsc1, Rsc2, Rsc4                    | - |
| SMARCJ       | BCL7A, BCL7B, BCL7C | AGAB1IDRAFT_112712 (A. bisporus, K5W213), UMA10422 (U. maydis, A0A0D1DVN6), RhirC2_747354 (R. irregularis, PKK70026) | BDH1, BDH2 |
| SMARCM       | GLTSCR1/1L     | -                                 | BRIPl, BRIPl2 |
| SMARCL       | SS18, CREST, SS18L2 | RISS18/Rig_105530 (R. irregularis, A0A015JMG5) | GFI1, GFI2 |
| SMARCM       | BCL11A, BCL11B | -                                 | - |
| SMARCN       | ACTL6A, ACTL6B, ACTB | Arp7, Arp9 | ARP4, ARP7 |
|              |                 | Snf6                              | - |
|              |                 | Rtt102                            | - |
|              |                 | Rsc3, Rsc30                       | - |
|              |                 | Rsc58                             | - |
|              |                 | Rsc14(Ldb7)                       | - |
|              |                 | Htl1                              | - |

Non-SMARC names correspond to lineage-specific subunits previously identified. New SMARC names assigned by the HUGO Gene Nomenclature Committee are indicated in italics.
SMARCD is an animal-specific subunit with already established SMARC nomenclature.
SMARCF naming for ARID-type subunits has been employed in a limited number of databases and articles.
All names correspond to reference species: human for Metazoa, Saccharomyces cerevisiae for Fungi, and Arabidopsis for Plants. Subunits not present in the reference species but present in the lineage are indicated.

The absence of BAF57-like proteins in S. cerevisiae had originally led to the assumption that SMARC subunits were exclusive to Metazoa. However, other fungal SWI/SNF complexes contain the Ssr4 subunit, for which we have found evidence of a shared evolutionary origin with animal BAF57. Although Ssr4 lacks the characteristic HMG domain, it contains a region homologous to the animal BAF57 NHRL1 domain, which partially overlaps with a previously defined kinesin-like coiled-coil region (KLC2). Importantly, the filasterean Capsaspora owczarzaki contains what could be considered an intermediate SMARC protein that presents characteristics of both metazoan BAF57 and fungal Ssr4, including the N-terminal Ssr4 domain (DUF1750), an HMG domain, and the C-terminal NHRL1 domain (Supplementary Fig. 2). Similar to SMARCH, no homologs for SMARCE were found outside opisthokonts, suggesting SMARCE may have originated in an opisthokontan common ancestor.

We readily identified orthologs of the bromodomain subunit BRD7/9 in most species, with the exceptions of bacteria and fission yeasts. In these yeasts, we identified Rsc1/2/4 proteins and most Archaeplastida species have bromodomain proteins highly similar to BRD7/9. In contrast, no BRD7/9 orthologs in D. discoideum and most plants but not within fungi, suggesting this subunit was lost in an early fungal ancestor (Fig. 1 and Supplementary Fig. 3). The two predicted SMARC orthologs in A. thaliana that we named BCL-domain homolog 1 (BDH1) and BDH2 (AT4G22320 and AT5G55210, respectively) have been identified with other SWI/SNF subunits in affinity purification experiments. However, no functional data connecting these proteins to plant SWI/SNF complexes is available. These results suggest the origin of the SMARCJ subunit in the LECA.

Outside of Metazoa, we find BCL7/SMARCC orthologs in D. discoideum and most plants but not within fungi, suggesting this subunit was lost in an early fungal ancestor (Fig. 1 and Supplementary Fig. 3). The two predicted SMARC orthologs in A. thaliana that we named BCL-domain homolog 1 (BDH1) and BDH2 (AT4G22320 and AT5G55210, respectively) have been identified with other SWI/SNF subunits in affinity purification experiments. However, no functional data connecting these proteins to plant SWI/SNF complexes is available. These results suggest the origin of the SMARCJ subunit in the LECA.

GLTSCR1/SMARCK and SS18/SMARCL were also confidently identified in animals, fungi, and plants. SMARCK, a signature subunit of the mammalian non-canonical BAF (ncBAF) complexes, is widespread among nearly all plant species analysed, as...
well as in fungi, being absent only in ascomycetes (Fig. 1). The two SMARCK subunits we detected in A. thaliana have been already characterized as functional components of a SWI/SNF complex that incorporates the A. thaliana BRM ATPase (AtSMARCA)\textsuperscript{27,30,31}. For SMARCL, we find an SSXT domain-driven phylogenetic connection between SS18 and the previously reported fission yeast-specific Snf30 subunit\textsuperscript{25}. The A. thaliana SMARCL orthologs we identify belong to the GIF1/AN3 family of transcriptional activators\textsuperscript{32} and have previously been associated with plant SWI/SNF complexes\textsuperscript{27,30,31}. These results suggest SMARCK and SMARCL were present in the LECA.

**Lineage-specific yeast SWI/SNF subunits.** As a confirmation of the existence of lineage-specific SWI/SNF subunits, we find that ten of the S. cerevisiae subunits without reported orthologs in humans or flies indeed had orthologs only in the fungal lineage (Fig. 1). Moreover, a thorough search among all fungal proteomes available in NCBI and MycoCosm suggests many of these subunits can only be found within the order Saccharomycetales—such as Snf6 and Rtt102—, or the family Saccharomycetaceae—such as Lbd7/Rsc14 and Htl1. Rsc3/30-related proteins are slightly more widespread, as similar sequences are found in ascomycetes and basidiomycetes. However, no experimental evidence supports this potential functional conservation. Taf14 is found in S. cerevisiae and S. pombe and the distantly related R. irregularis, suggesting either a single horizontal transfer event between fungi or multiple loss events during fungal evolution. Rsc58 is confidently found in two species, S. cerevisiae and S. pombe, and has been functionally associated with the RSC complex in both species\textsuperscript{25}.

**A common eukaryotic ancestor for the SMARCG subunits.** Based on limited homology searches, the 2x tandem PHD-containing BAF45/SMARCG subunits were considered metazoan SWI/SNF-specific subunits, while the chromatin remodeling complex (CRC)-domain Swp82/Rsc7 subunits were considered fungi-specific\textsuperscript{15,33}. Our comprehensive analysis here establishes a previously disregarded common evolutionary origin for both sets of subunits, which is also shared by plants (Fig. 3a and Supplementary Fig. 4). First, the lack of a characteristic PHD domain in the S. cerevisiae Swp82 and Rsc7 proteins appears to be an exception among fungi. In fact, most fungal Swp82/Rsc7
**Fig. 2** SMARCF subunits are found in all eukaryotes. **a** Domain architecture of known SMARCF subunits in H. sapiens, S. cerevisiae, C. subellipsoidea, and A. thaliana. Scale bar, 100 amino acids. **b** Graphical summary of the evolution of SMARCF domain architectures as predicted from Supplementary Fig. 1. The scale bar indicates the primary sequence length. Phylogram represents the suggested relationship between SMARCF subunits. Arrow indicates the duplication of a single SMARC into SMARCF1 and SMARCF2 in a fungal and animal common ancestor. Domains are predicted based on Pfam and InterProScan hits, and depicted as colored boxes as indicated in the figure. ARM fold represents a series of ARM-fold hits (IPR016024 and IPR00025) and BAF250C (PF12031/ IPR033388); ARID, AT-rich interaction domain (PF01388/IPR001606); RFX DBD, RFX DNA-binding domain (PF02257/IPR003150); C2H2 Zf, zinc finger, C2H2 type (IPF00096). H.s, Homo sapiens; S.c, Saccharomyces cerevisiae; R. irregularis, Roslogus irregularis; C.s, Coccomyxa subellipsoidea; A. castellanii, Acanthamoeba castellanii; A. thaliana, Arabidopsis thaliana; C. braunii, Chara braunii. “A. castellanii (Armooboeza), is a close relative of D. discoideum with a bona fide SMARCF subunit.”

**Fig. 3** Evolutionary conservation of SMARCF domains. **a** Phylogenetic tree of SMARCF domains show conservation across metazoans, fungi, and plants. Scale bar indicates evolutionary time. **b** Multiple structure alignment of SMARCF1/2 domains used for structural predictions. N- and C-terminal regions are highly conserved across species.

**Fig. 4** Evolutionary conservation of SWI/SNF chromatin remodeler subunits. **a** Phylogenetic tree of SWI/SNF subunits show conservation across metazoans, fungi, and plants. Scale bar indicates evolutionary time. **b** Multiple structure alignment of SWI/SNF subunits used for structural predictions. N- and C-terminal regions are highly conserved across species.

**Fig. 5** Evolutionary conservation of SWI/SNF chromatin remodeler complexes. **a** Phylogenetic tree of SWI/SNF complexes show conservation across metazoans, fungi, and plants. Scale bar indicates evolutionary time. **b** Multiple structure alignment of SWI/SNF complexes used for structural predictions. N- and C-terminal regions are highly conserved across species.

**SMARCG1/TPF1 is a SWI/SNF subunit in plants.** The plant model organism A. thaliana contains two SMARCG subunit paralogs, TPF1 and TPF2. TPF1 displays nuclear sublocalization (Supplementary Fig. 5c) and is able to co-immunoprecipitate multiple known plant SWI/SNF subunits in immunoprecipitation mass spectrometry (IP-MS) experiments (Table 2 and Supplementary Data 1). Of the four plant SWI/SNF ATPases, only CHR12 was identified in the TPF1 bait-based IP-MS experiments suggesting TPF1 complexes preferentially incorporate the CHR12 paralog. Interestingly, three highly-enriched uncharacterized TPF1 interactors were previously identified in separate plant IP-MS experiments using different bait SWI/SNF subunits. These uncharacterized proteins are SAWANDEE HOMEODOMAIN HOMOLOG 2 (SHH2), and ATIG23730 and ATIG106500, which we refer to as PLANT-SPECIFIC SWI/SNF-ASSOCIATED PROTEIN 1 (PSA1) and PSA2, respectively (Table 2). While no functional data exists for PSA1 and PSA2, SHH2 has been reported to bind H3K9me2 in vitro. Furthermore, when SHH2 is used as an IP-MS bait it pulls down a protein complex similar to the protein complex identified with TPF1 (Table 2 and Supplementary Data 1). Here we identify CHR12 and CHR23 paralogs, as well as the TPF1 paralog TPF2. Moreover, three proteins were highly-enriched: BROMODO-MAIN 5 (BRD5), and two paralogs of a protein we name ONE PHD FINGER 1 (OPF1) and OPF2. Both BRD5 and OPF1 were detected in TPF1 IP-MS experiments. Importantly, these uncharacterized SWI/SNF interactors, SHH2, PSA1, PSA2, BRD5, and OPF1/2, are streptophyte-specific, suggesting lineage-specific subunits have also evolved in plants (Supplementary Fig. 6a). In summary, we have identified a plant SWI/SNF subclass reminiscent of the animal BAF complex that selectively incorporates the ATPases CHR12 or CHR23 as well as a set of plant-specific subunits of still unknown function.

**Discussion**

SWI/SNF chromatin remodelers have been studied for decades in different model organisms and a wealth of information is available pertaining to their function, targets, and architecture. Our systematic evolutionary approach here confirms the strong conservation of multiple subunits. Crucially, our approach here also identifies unnotified phylogenetic connections between some SWI/SNF subunits and identifies SMARCG as a conserved SWI/SNF subunit in plants. A possible explanation for previously
missed links is that previous phylogenetic analyses were conducted using a limited number of metazoan, fungal, and plant protein sequences. With this limited dataset, large phylogenetic distances between species could complicate the identification of orthologous genes. Thus, our approach of concatenated searches in the proteomes of phylogenetically related species is a powerful method to unravel previously hidden phylogenetic relationships across evolutionarily distant species.

Of all previously reported metazoan SWI/SNF subunits, we find only the BAF-specific subunit BCL11 lacks homologs in other lineages, while orthologs for all other SWI/SNF subunits can be found in more distant lineages like fungi and plants. Indeed, our work reveals three phylogenetic relationships that change the evolutionary model of SWI/SNF complexes in eukaryotes: (i) the opisthokontan origin of SMARCE, (ii) the origin of SMARCF in a LECA instead of a last opisthokontan ancestor, and (iii) the origin of SMARCG subunits in a LECA instead of a metazoan ancestor.

Although the BAF57 subunit has been historically defined as an HMG-domain protein specific to Metazoa, the presence of an NHRLI domain both in BAF57 and fungal Ssr4 subunits suggests a phylogenetic link and that both proteins have a
common ancestor that gave rise to the SMARCE family in fungi and animals. In fact, two pieces of evidence support the hypothesis that the NHRLI domain is a reliable indicator of SMARCE conservation. First, two recent cryoEM studies of the mammalian BAF complex were able to resolve only the regions of BAF57 corresponding to the NHRLI domain (amino acids 220–298) or the entire KLCC domain that parallels KLCC domain architecture in Swp82 and Rsc7 proteins. RiRsc7 and RgRsc7, proteins appear to perform a similar structural role within the complex. Alternatively, Snf6 and BAF57/Ssr4 could have had different ancestral origins and undergone convergent evolution towards a similar function in the complex.

The common origin of the plant LFR protein and the signature SMARCF subunits Swi5/ARID1 and Rsc9/ARID2 in yeast and animals is supported by our phylogenetic analysis of the ARM domain. Our findings are consistent with the previously described and prominent role the ARM domain plays in the architecture of fungal and animal SWI/SNF complexes.

Table 2 Plant BAF subunits found by TPF1 and SHH2 IP-MS.

| Name      | Protein ID | TPF1 experiment 1 | TPF1 experiment 2 | SHH2 experiment 1 | SHH2 experiment 2 |
|-----------|------------|-------------------|-------------------|-------------------|-------------------|
|           |            | Col-0 #11 #4      | Col-0 #11 #4      | Col-0 #10 #21     | Col-0 #10 #21     |
| TPF1      | AT3G52100  | - 36 36           | - 24 24           | - 31 30           | - 31 31           |
| ARP7      | AT3G60830  | 24 32 32          | - 23 21           | 28 31 31          | 26 31 31          |
| ARP4      | AT1G18450  | 25 30 31          | - 24 21           | 27 30 29          | 25 29 29          |
| PSA2      | AT1G06500  | - 30 30           | 20 18             | - 28 28           | - 29 28           |
| SHH2      | AT3G18380  | - 29 29           | - 21 19           | - 29 28           | - 29 29           |
| PSA1      | AT1G32730  | - 29 28           | - 21 19           | - 29 28           | - 29 29           |
| BDH1      | AT4G22320  | - 29 30           | - 20 18           | - 27 27           | - 28 27           |
| OFP1      | AT1G05620  | - 29 28           | - 20 18           | - 33 33           | - 33 33           |
| SWI3B     | AT2G33610  | - 29 26           | - 22 19           | - 30 29           | - 30 29           |
| SWI3A     | AT2G47620  | - 28 26           | - 23 21           | - 30 30           | - 30 30           |
| CHR12     | AT3G06600  | - 28 26           | - 22 20           | - 30 29           | - 30 29           |
| SWP73B    | AT5G14170  | - 28 26           | - 22 20           | 25 30 29          | 25 29 29          |
| BRD5      | AT1G58025  | - 27 25           | - 20 18           | - 29 27           | - 29 28           |
| LFR       | AT3G22990  | - 25 23           | - 21 18           | 23 30 28          | - 30 30           |
| BSH       | AT3G17590  | - 25 -             | - 21 19           | - 30 29           | - 30 30           |
| OPF2      | AT3G20280  | - - -             | - - -             | - 30 30           | - 30 30           |
| TPF2      | AT3G08020  | - - -             | - - -             | - 28 27           | - 29 29           |
| BDH2      | AT5G55210  | - - -             | - - -             | - 27 -             | - 28 27           |
| CHR23     | AT5G19310  | - - -             | - - -             | - 27 24           | - 29 29           |

Data represents log2 LFQ intensity values of two independent IP-MS experiments in two independent TPF1-3xFLAG and SHH2-3xFLAG lines compared to untransformed Col-0 controls. Shown interactors followed the criteria LFQ control < LFQ transgenic/10 in at least one TPF1-3xFLAG transgenic line in both independent experiments. Corresponding log2 LFQ values for these proteins in the SHH2-3xFLAG experiments are shown on the right. Values for the paralogs OPF2, TPF2, BDH2, and CHR23, which were only detected in the SHH2-3xFLAG experiments, are also included. See Supplementary Data 1 for complete datasets and analyses.
subunits. Interestingly, structural analyses have shown that DPF2 and Rsc7 occupy similar positions in their respective complexes. On the other hand, the plant SMARCG subunit (TPF) has incorporated a Tudor-like domain, which could be involved in the recognition of methylated histones, similar to the described function of PHD domains. Consistent with its relationship with animal SMARCG subunits, TPF1 is also found in vivo in a plant SWI/SNF complex that specifically includes the functionally redundant CHR12 or CHR23 ATPases, in addition to numerous functionally uncharacterized plant-specific proteins. Among these subunits, functional information only exists for SHH2, which appears to bind H3K9me2 in vitro through its SAWADEE domain. This activity is consistent with its paralog SHH1, which contributes to the plant de novo DNA methylation pathway called RNA-directed DNA methylation (RdDM). While SHH1 appears to be angiosperm-specific, SHH2 orthologs are found in most streptophytes. The failure to identify any RdDM components by SHH2 IP-MS suggests SHH2 has ancestral functions in the SWI/SNF complex and implies a neo-functionalization of SHH1 in RdDM. Furthermore, our IP-MS findings that TPF1 bait fails to identify TPF2, while SHH2 bait identified both TPF1 and TPF2, suggests that this SWI/SNF complex only utilizes one TPF protein. We extend this finding to OPF paralogs. Future functional studies will help clarify these questions. Importantly, none of the uncharacterized putative SWI/SNF subunits we identify here (SHH2, PSA1, PSA2, BRD5, and OPF1/2), nor TPF1/2 have been previously identified in published experiments using the plant BRAHMA (BRM) ATPase as bait, or even other proteins that interact with BRM, such as SWI3C, BRIP1/2, BRD1/2/13, and SS18/GIF1 (Table 1). However, recent BRM bait experiments detected plant SMARCK (BRIP1/2), SMARCI (BRD1/2/13), and SMARCL (GIF2) that were not found in TPF1 or SHH2 bait-based IP-MS experiments. These results suggest plants have at least two different SWI/SNF subclasses (Fig. 4). One subclass would specifically incorporate the PHD-containing proteins TPF1/2 and OPF1/2, the bromodomain protein BRD5, SHH2, PSA1 and PSA2, the CHR12 or CHR23 ATPases, and SMARCF/LFR. A second subclass would incorporate the BRM ATPase, together with the BRIP1/2, BRD1/2/13, and SS18/GIF1 subunits (Fig. 4). It’s worth noting that the composition of these subclasses is reminiscent of metazoan BAF-PBAF and ncBAF, respectively. Consistent with these findings, SMARCF/LFR, the BAF/PBAF signature subunit, is detected in TPF1 and SHH2 bait experiments but is not identified in two recent studies using BRM as the bait.

![Diagram of SWI/SNF complexes](image-url)
In summary, we have identified a plant canonical BAF-like (cBAF) complex that likely evolved from the ancestral form that branched into the SWI/SNF-BAF and RSC-PBAF complexes during opisthokont evolution, and subsequently incorporated a set of plant-specific subunits of yet unknown function. The broad conservation of multiple subunits across diverse lineages suggests that SWI/SNF complexes have been conserved through millions of years as important factors for controlling chromatin accessibility. During the evolution of specific lineages, these remodeler complexes have diverged into different subclasses and have incorporated species-specific subunits, resulting in a diverse array of architectures and compositions in the extant lineages. A notable exception appears to be the loss of subunits in the parasite T. brucei. However, this might be explained by the preference in this organism for posttranscriptional regulation of polycistronic genes as a mechanism of regulating gene expression81. How SWI/SNF complexes evolved into their current architectures and what was the composition of the ancestral SWI/SNF in LECA remains to be described in detail. The evolutionary analyses of multiple lineages as well as details to speculate about the composition of the SWI/SNF complex in the LECA. For these data, we propose a model for the evolution of the SWI/SNF complexes (Fig. 4) where an ancient complex in LECA first divided into two main subclasses (a-cBAF and a-cNAcBAF), present in all extant eukaryotes, followed by divergence of a-cBAF into two separate subclasses (a-ABF and a-ABF) characteristic of the opisthokont lineage. Subsequently, we suggest further loss and incorporation of lineage-specific subunits shaped the current architectures of extant SWI/SNF complexes. We hope our evolutionarily informed model will facilitate future SWI/SNF functional analyses across a broad range of species.

**Methods**

**Identification of SWI/SNF complex subunit sequences in eukaryotes.** Ortholog searches for each subunit in the SWI/SNF complex were performed following a phylogeny-based step-by-step look-up. Previously characterized protein sequences from the subunits of the Homo sapiens BAF-PBAF-NcBAF complexes, and its orthologous complexes in baker’s yeast (S.witofn and RSC) were used as starting queries (Table 1). All the resulting hits were repeatedly used as queries in new searches until no new hits appeared. A combination of BLASTP and Phmmer searches using protein sequences as queries were performed on proteome databases searches using protein sequences as queries were performed on proteome databases until no new hits appeared. A combination of BLASTP and Phmmer searches using protein sequences as queries were performed on proteome databases until no new hits appeared. A combination of BLASTP and Phmmer searches using protein sequences as queries were performed on proteome databases. Robust hits of cold acetone and after the last wash, the pellet was air-dried.

**Phylogenetic analysis.** For SMARCC, PHD domain sequences were extracted based on Pfam/InterProScan automatic annotation (ARID, PF01388/PIR001466), ARID sequences shorter than 50 amino acids were curated with the BLASTP/Phmmer query list when these were successfully found, establishing a priori BlastP score thresholds for all proteins, we relied on the following procedure: First, we performed reciprocal BLAST using the subjects as queries and InterProScan (https://www.ebi.ac.uk/interpro) analyses to confirm the presence of known domains and protein architectures. HMM searches using Pfam were also performed for several subunits. These searches served both as complementary confirmation, and to identify distantly related orthologs. Robust hits that did not match the expected domain composition were used to propose previously overlooked connections between sequences, attributable to domain loss/gains. When no hits were found, tax sampling was extended to entire lineage proteomes in NCBI, UniProt, PhylomeX, Mycocosm, and oneKP databases to confirm the absence or presence of specific proteins or domains. Manual curation was performed for inconclusive or unclear cases. Finally, positive hits were included in the BLASTP/Phmmer query list when these were successfully found, restarting the search process.

The eluted peptides were then ionized via nanoelectrospray ionization, and mass spectrometry data were acquired using an Orbitrap Fusion™ Lumos™ “Tribrid” Mass Spectrometer (Thermo Fisher Scientific) with an MS1 resolution of 120,000 followed by sequential MS2 scans at a resolution of 15,000. For the SHH2-3xFLAG samples, tryptic-digested protein samples were analyzed using a Q Exactive Plus mass spectrometer using a Thermos Easy-nLC system coupled to a Thermo Q-Exactive MS. For experiment 2 of TFPI-3xFLAG, samples were prepared and analysed at the Wohlschlegel Lab at UCLA (Los Angeles, US). Briefly, proteins were reduced and alkylated using 5 mM Tris (2-carboxyethyl) phosphine and 10 mM iodoacetamide, respectively. Protein digestion was achieved by sequential addition of endopeptidase Lys-C (BioLabs) and proteinase K (Pierce®) at a 1:1 ratio (peptide/protein ratio) overnight. The digested samples were quenched by the addition of formic acid to a final concentration of 10% for 10 min at 10,000 × g, and filtered using a 40 μm cell strainer. Samples were rotated for 3 h at 4 °C with 200 μl of Anti-FLAG M2 magnetic beads (Sigma, cat#M8823) that were previously blocked with 5% BSA. The magnetic beads were captured and washed three times with IP buffer and supernatant. The samples were eluted for 30 min with 150 μl of 3xFLAG peptide (Sigma, cat#F4799) at a concentration of 250 μg/ml in IP buffer without NP40. This was repeated two more times and the eluates were combined. Proteins were precipitated by the addition of TCA to a final concentration of 20% and incubated for 30 min on ice followed by 30 min at 4 °C centrifugation at 12,000 × g. HMM Searches were analysed by tandem mass spectrometry data were acquired using an Orbitrap Fusion™ “Lumos” Tribrid™ Mass Spectrometer (Thermo Fisher Scientific) with an MS1 resolution of 120,000 followed by sequential MS2 scans at a resolution of 15,000. For the SHH2-3xFLAG samples, tryptic-digested protein samples were analysed using a Q Exactive Plus mass spectrometer using a Thermos Easy-nLC system coupled to a Thermo Q-Exactive MS. For experiment 2 of TFPI-3xFLAG, samples were prepared and analysed at the
Bacterial suspensions were then mixed adjusting to a...
Clayton, C. Regulation of gene expression in trypanosomatids: living with
Law, J. A. et al. Polymerase IV occupancy at RNA-directed DNA methylation
Wallace, I. M., O

Sang, Y. et al. Mutations in two non-canonical Arabidopsis SWI2/SNF2

31. Yu, Y. et al. BRAHMA-interacting proteins BRIP1 and BRIP2 are core
subunits of Arabidopsis SWI/SNF complexes. Nat. Plants 6, 996–1007 (2020).

32. Kim, J. H. & Tsukaya, H. Regulation of plant growth and development by
the GROWTH-REGULATING FACTOR and GRF-INTERACTING FACTOR
duo. J. Exp. Bot. 66, 6093–6107 (2015).

33. Wilson, B., Erdjument-Bromage, H., Tempst, P. & Cairns, B. R. The RSC
chromatin remodeling complex bears an essential fungal-specific protein subunit
with broad functional roles. Genetics 172, 795–809 (2006).

34. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The
Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc.
10, 845–858 (2015).

35. Lu, R. & Wang, G. G. Tudor: a versatile family of histone methylation
readers. Nature 503, 829–843 (2013).

36. Marcum, R. D., Reyes, A. A. & He, Y. Structural insights into the
evolutionarily conserved BAF chromatin remodeling complex. Biology https://
doi.org/10.3390/biology9070146 (2020).

37. Zhang, H. et al. DTF1 is a core component of RNA-directed DNA
methylation and may assist in the recruitment of Pol IV. Proc. Natl Acad. Sci.
USA 110, 8290–8295 (2013).

38. Pandey, R. et al. Analysis of histone acetyltransferase and histone deacetylase
families of Arabidopsis thaliana suggests functional diversification of
chromatin modification among multicellular eukaryotes. Nucleic Acids Res.
30, 5036–5055 (2002).

39. He, S. et al. Structure of nucleosome-bound human BAF complex. Nat. Comm.
39, 110–114 (2018).

40. Law, J. A. et al. Polymerase IV occupancy at RNA-directed DNA methylation
sites requires SHH1. Nature 498, 385–389 (2013).

41. Clayton, C. Regulation of gene expression in trypanosomatids: living with
polycistronic transcription. Open Biol. 9, 190072 (2019).

42. Wallace, I. M., O’Sullivan, O., Higgins, D. G. & Notredame, C. M-Coffee:
combining multiple sequence alignment methods with T-Coffee. Nucleic Acids
Res. 34, 1692–1699 (2006).

43. Clayton, C. Regulation of gene expression in trypanosomatids: living with
polycistronic transcription. Open Biol. 9, 190072 (2019).

44. Wallace, I. M., O’Sullivan, O., Higgins, D. G. & Notredame, C. M-Coffee:
combining multiple sequence alignment methods with T-Coffee. Nucleic Acids
Res. 34, 1692–1699 (2006).

45. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-
throughput functional analysis of genes in planta. Plant Physiol. 133, 462–469
(2003).

46. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold.
Nature 596, 583–589 (2021).

47. Pettersen, E. F. et al. UCSF chimeraX: structure visualization for researchers,
educators, and developers. Protein Sci. 30, 70–82 (2021).

Acknowledgements
We thank Ceegay Lee and Nathan Cai for their technical support. We thank Drs Jake
Harris and Bruno Catarino for their critical reading of the manuscript. This work
was supported by grants NIH Grant R35 GM130272 [to S.E.J.], RY2C2018-024108-1 [to
J.G.-B.] funded by MCIN/AEI/10.13039/501100011033 and by ‘‘ESF Investing in your
future’’, and PID2019-108577GA-I00 [to J.G.-B.] and PID2019-110717GB [to M.A.B.]
funded by MCIN/AEI/10.13039/501100011033. S.E.J. is a Howard Hughes Medical
Institute investigator.

Author contributions
J.H.-G. conducted all phylogenetic analyses. B.D.-M. and J.G.-B. conducted the confocal
microscopy experiment. P.H.-K. and J.G.-B. generated transgenic lines. B.D.-M., P.H.-K.,
and J.G.-B. conducted the immunoprecipitation experiments. Y.J.-A., A.A.V., and J.W.
conducted the mass spectrometry analyses. S.E.J supervised the research and edited the
manuscript. J.H.-G., M.A.B., and J.G.-B. conceived the original experimental design and
wrote the paper.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material
available at https://doi.org/10.1038/s42003-022-03490-x.

Correspondence and requests for materials should be addressed to Javier Gallego-
Bartolomé.

Peer review information Communications Biology thanks Jinghua Yang and the other
anonymous, reviewer(s) for their contribution to the peer review of this work. Primary
Handling Editors: Luke R. Ginham. Peer reviewer reports are available.

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

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