

Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex

Henri-Marc Bourbon*

Centre de Biologie du Développement, UMR5547 CNRS/Toulouse III, IFR109, Université Paul Sabatier, 31062 Toulouse, France

Received March 21, 2008; Revised and accepted May 14, 2008

ABSTRACT

The multisubunit Mediator (MED) complex bridges DNA-bound transcriptional regulators to the RNA polymerase II (PolII) initiation machinery. In yeast, the 25 MED subunits are distributed within three core subcomplexes and a separable kinase module composed of Med12, Med13 and the Cdk8-CycC pair thought to control the reversible interaction between MED and PolII by phosphorylating repeated heptapeptides within the Rpb1 carboxyl-terminal domain (CTD). Here, MED conservation has been investigated across the eukaryotic kingdom. *Saccharomyces cerevisiae* Med2, Med3/Pgd1 and Med5/Nut1 subunits are apparent homologs of metazoan Med29/Intersex, Med27/Crsp34 and Med24/Trap100, respectively, and these and other 30 identified human MED subunits have detectable counterparts in the amoeba *Dictyostelium discoideum*, indicating that none is specific to metazoans. Indeed, animal/fungal subunits are also conserved in plants, green and red algae, entamoebids, oomycetes, diatoms, apicomplexans, ciliates and the ‘deep-branching’ protists *Trichomonas vaginalis* and *Giardia lamblia*. Surprisingly, although lacking CTD heptads, *T. vaginalis* displays 44 MED subunit homologs, including several CycC, Med12 and Med13 paralogs. Such observations have allowed the identification of a conserved 17-subunit framework around which peripheral subunits may be assembled, and support a very ancient eukaryotic origin for a large, four-module MED. The implications of this comprehensive work for MED structure–function relationships are discussed.

INTRODUCTION

In higher eukaryotes, a large variety of sequence-specific transcription factors regulate the expression of thousands of protein-coding genes, ensuring the proper development and functioning of the organism (1–3). The specificity of this transcriptional control occurs primarily through differential recruitment of the basal RNA polymerase II (PolII) initiation machinery to gene promoters (4,5). Transcription by PolII is an elaborate multi-step process that requires the fine-tuned assembly on core promoters of a massive pre-initiation complex (PIC) of more than 60 proteins, including the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIH (6–8). Biochemical studies in the budding yeast *Saccharomyces cerevisiae*, in mammals and the fruit fly *Drosophila melanogaster* have revealed a pivotal role played in this process by a large (1MDa) multi-protein entity termed Mediator (MED) (9). Comprising up to 30 distinct subunits in mammals, MED acts as a modular interface bridging diverse transcription factors arrayed on regulatory DNA regions to the PolII initiation machinery (10–13).

MED complexes are unable to directly contact transcriptional ‘enhancer’ or ‘silencer’ DNA elements. Instead, they physically interact with gene-specific transcription activators and repressors, PolII subunits and some GTFs, conveying DNA-directed signals to the basal initiation apparatus (10–14). The phosphorylation status of the highly repetitive heptads in the carboxyl(C)-terminal domain (CTD) of the largest PolII subunit Rpb1 is thought to play a critical role in orchestrating the interaction between the yeast polymerase and MED (15,16). Indeed, removal of all the CTD heptads is lethal *in vivo* (17) and precludes, *in vitro*, the formation of a stable PolII/MED ‘holoenzyme’ complex (15). While PolII holoenzyme possesses a hypo-phosphorylated CTD, heptads of the actively transcribing enzyme are heavily phosphorylated

*To whom correspondence should be addressed. Tel: +33 0561558288; Fax: +33 0561556507; Email: bourbon@cict.fr

(mainly on Ser2 and Ser5) and are no longer associated with MED (16,18). These data provided evidence that CTD Ser2/Ser5 phosphorylations facilitate dissociation of MED from the transcriptionally competent polymerase, to be recycled into a new PIC with unmodified PolII (18). In addition, biochemical studies from the budding yeast have indicated that MED and a subset of GTFs persist at the core promoter during the transition from initiation to elongation, suggesting a role in facilitating re-initiation (19,20).

The 3D structure of purified *S. cerevisiae* PolII holoenzyme complex, as reconstructed from electron microscopy (EM) images, reveals an extended MED that consists of three separate subdomains of approximately equal mass, termed 'Head', 'Middle' and 'Tail', wrapping around the globular polymerase (21,22). Multiple contacts between MED and PolII extend from Head to the intersection between Middle and Tail domains (14,23). The structural and functional organization of the MED complex has been mainly explored in the budding yeast (Figure 1A). Here, the 21 core subunits are distributed among three modules that roughly correspond to the Tail, Middle and Head domains seen in the 3D holoenzyme structure (14,24,25). The Tail module includes the Med2, Med3, Med5, Med14, Med15 and Med16 subunits, several of which directly interact with DNA-bound transcriptional activators and repressors (26–29). The Middle module, containing the Med1, Med4, Med7, Med9, Med10, Med21 and Med31 subunits, directly interacts with CTD in biochemical assays and is thought to function primarily in transferring regulatory inputs from activators and repressors to the Head module, PolII and GTFs, at a post-binding stage (24). In addition to transcription factors and CTD, both the Middle and Tail modules also physically interact with several GTFs, notably TFIID and TFIIE (24,30). The Head module, comprising the Med6, Med8, Med11, Med17, Med18, Med19, Med20 and Med22 subunits, has been proposed to play a general role in transcription. And indeed, a recombinant Head module interacts with a reconstituted PolII–TFIIF complex and stimulates basal transcription *in vitro* (31). The TATA box-binding protein (TBP), an essential TFIID component, interacts with the amino(N)-terminal region of Med8 within a Med8–Med18–Med20 Head module triad (32). Lastly, in addition to the three core subcomplexes, a separable four protein regulatory module, composed of Med12 and Med13 plus the cyclin-dependent kinase Cdk8 and its cyclin partner CycC, appears mainly involved in transcriptional repression in exponentially growing yeast cells (33,34). The regulatory activity of the so-called Cdk8 module involves phosphorylations of specific proteins, including transcriptional activators, core MED subunits, some GTFs and Rpb1 CTD heptads prior to transcriptional initiation (19,34–40).

Extensive protein sequence analyses have indicated that 22 of the 25 budding yeast MED subunits have detectable homologs among the 33 mammalian subunits identified to date (12,41,42). It is noteworthy that the three remaining *S. cerevisiae* subunits (i.e. Med2, Med3 and Med5) belong to the Tail module (Figure 1A) and together with Med15 are capable of assembling a stable subcomplex

when co-expressed in insect cells (43). Available data provide compelling evidence that the functional organization of yeast MED into four modules (i.e. the core complex plus the Cdk8 module) has been conserved in metazoans (33,41,44), and significant similarities in the overall shapes of isolated yeast and mammalian MED complexes are in fact revealed by EM analyses (14,22). However, it should also be emphasized that eight mammalian MED subunits (i.e. Med23–30) have so far been identified only in metazoan complexes (42).

The ability of MED to act as a signal transducer from transcriptional regulators to the general PolII initiation machinery is likely to have played a major role in the evolutionary diversification of eukaryotes. It is assumed that animals and fungi diverged relatively recently and belong to the so-called opisthokont 'supergroup' (45,46). In contrast, a systematic search for MED subunit homologs in other eukaryotic kingdoms has not been performed to date. To investigate the structural conservation of MED among a broader sample of eukaryotes, I have taken advantage of the rapidly expanding collection of sequenced (>90%) genomes from species representing the following supergroups or phyla: Microsporidia, Plantae, Rodophyta, Amoebozoa, Heterokonta, Ciliata, Kinetoplastida, Trichomonadida (i.e. *Trichomonas vaginalis*) and Diplomonadida (i.e. *Giardia lamblia*). Of note, the Amitochondriate parasitic protists *T. vaginalis* and *G. lamblia*, often considered to represent deeply branching eukaryotes (47,48), lack PolII CTD heptads (49). The comparative genomic approach applied here to MED subunits from a spectrum of 70 eukaryotes, ranging from the most primitive unicellular organisms to mammals, leads to several conclusions: (i) first, it shows that all the known budding yeast MED subunits, including Med2, Med3 and Med5, have structural counterparts in insects and mammals; (ii) second, it identifies a set of core subunits detectable in most eukaryotic taxa, including Trichomonadida and Diplomonadida; (iii) third, these data indicate that repetitive CTD heptads may not be critical for assembling a PolII holoenzyme *in vivo* and (iv) fourth and last, it provides evidence that no MED subunit is specific to animals. Taken together, these data provide compelling support for an ancient four-module MED that appeared early on during eukaryotic evolution, apparently before acquisition of PolII CTD repetitive heptads. Finally, I speculate that this same set of about 30 detectably conserved MED subunits may also have contributed centrally to the diversification of transcriptional programs.

MATERIALS AND METHODS

Database searches

In this work, completed genome sequences (>90%) from 70 eukaryotes, including animals, fungi, land plants, green and red algae, amoebae, oomycetes, diatoms and parasitic protists, were examined. The entire list of species, references and web sites for genome projects are compiled in dataset S1 (Supplementary Material online). PSI-Blast, BlastP and TblastN searches (50,51) were undertaken at

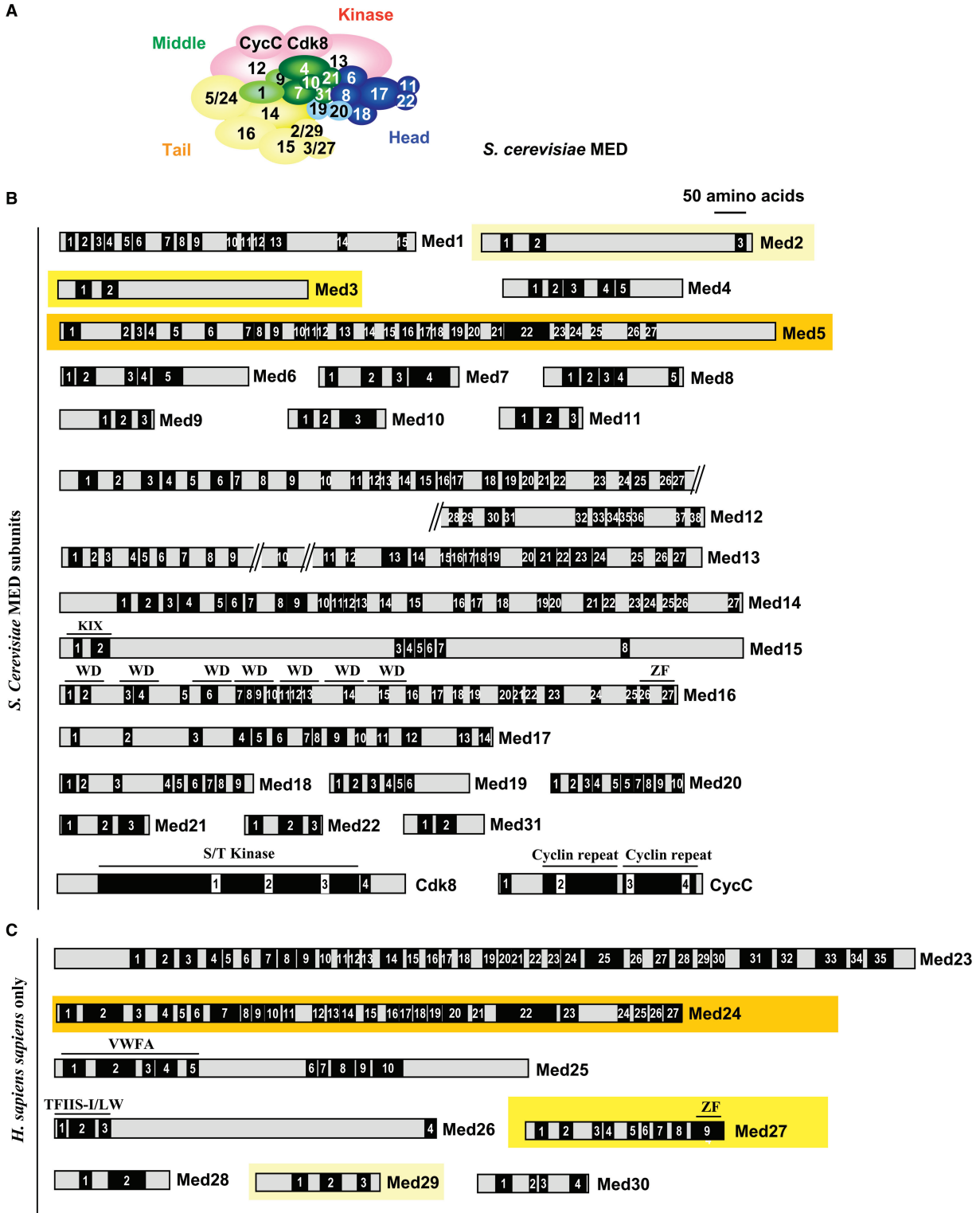


Figure 1. Conserved motifs and domains among fungal and human MED subunits. (A) Architectural organization of *S. cerevisiae* MED. Subunits have been assigned to Tail (yellow), Middle (green), Head (blue) and Cdk8 kinase (red) modules according to an integrated interaction map from ref. (25), those essential for cell viability being depicted by darker colours. (B–C) Distribution of conserved SSMs within the primary sequences of the 25 identified *S. cerevisiae* MED subunits (A) and of the eight human MED subunits found thus far only in animals (B). SSMs are represented by black boxes, except those (for Cdk8 and CycC) located within broadly conserved kinase and cyclin repeat domains (in black), which are depicted by white boxes. As indicated, for each subunit, they have been numbered from the N- toward the C-terminus. For Med15, Med16, Med25, Med26 and Med27 the extents of the distinguishable domains found in many functionally unrelated proteins are shown above the corresponding SSMs. Of note, for most subunits the SSMs are distributed throughout the primary sequences. Drawings of fungal Med2, Med3 and Med5 subunits and those of their predicted human counterparts Med29, Med27 and Med24, respectively, are included within boxes coloured according to the novel orthology assignments shown in Figure 2.

the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=euk), at the Max Planck Institute for developmental biology (MPI) (http://toolkit.tuebingen.mpg.de/psi_blast) or at dedicated websites (see dataset S1). Queries included the entire primary sequences or portions encompassing one or several of the most highly conserved regions (i.e. 'signature sequence motifs' or SSMs; see below) of known or predicted MED subunits. In the initial part of this study, PSI-Blast analyses were mostly performed using the BLOSUM62 'substitution matrix' and with an inclusion threshold of 0.001. A 'phylogenomics' approach was applied to increase the probability of identifying true 'orthologs'. For example, the entire primary sequences of predicted *Caenorhabditis elegans* MED subunits were used as queries to readily identify *C. briggsae* counterparts. When short sequences corresponding to SSMs were used as inputs in TBLASTN analyses, the 'expect' (*E*) threshold was generally 10 (default) and the 'low complexity filter' was mostly omitted. The Exon–intron organizations were deduced by comparing genomic sequences with expressed sequence tags (when available) and/or inspection for exons flanked by consensus intron splice sites maintaining proper open reading frames. Also, note that the availability of genome sequences from closely related species (e.g. *Ciona intestinalis* and *Ciona savignyi*) helped to discriminate among alternative gene prediction models. Accession numbers for the protein or genomic sequences reported throughout this study are given in dataset S6 and the conceptual primary sequences of genuine or predicted MED subunits are compiled in dataset S7.

Identification of evolutionarily conserved signature motifs for opisthokont MED subunits

SSMs typical of each MED subunit identified from opisthokonts were inferred from MAFFT alignments (52,53) done at the Kyoto University (<http://align.bmr.kyushu-u.ac.jp/mafft/online/server/>), as evolutionarily conserved motifs comprising at least seven amino-acid residues and present in at least (i) 34 out of 40 aligned sequences for subunits common to fungi/animals, (ii) 16 out of 20 for subunits previously thought to be present only in fungi (i.e. Med2, Med3 and Med5) or (iii) 17 out of 20 for those previously thought to be present only in animals (i.e. Med23–30). Note that equivalent numbers of representative animal and fungal species were examined to minimize kingdom-specific biases.

Structural validation of MED subunit homologs

Candidate proteins were assigned as predicted MED subunits by PSI-Blast analyses undertaken at MPI (http://toolkit.tuebingen.mpg.de/psi_blast). Whole-sequence alignments generated by MAFFT analyses (above) were used as inputs to derive position-specific scoring matrixes (PSSMs). Only sequences previously assigned as potential MED subunits by PSI-Blast analyses (*E*-values >0.001) and secondary structure predictions (below) were included in the alignments. Indeed, only 'jump-starting' PSI-Blast analyses using as inputs PSSMs generated from MAFFT

alignments that included validated proteins, have allowed detection of remote homologs. Also, note that inclusion of sequences from related species (when available) led to improved alignments (particularly for none-opisthokonts). Non-redundant eukaryotic sequences or in-house databases were searched using the Smith–Waterman algorithm (54). The *E*-values given in the text are from round 2 or 3. In many cases, MED subunits were also predicted from conserved domain searches at NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), at the European Bioinformatics Institute (EBI) (<http://www.ebi.ac.uk/InterProScan/>), and finally at MPI using the HHpred interactive server (<http://toolkit.tuebingen.mpg.de/hhpred>) for protein homology detection and structure prediction. HHpred results [i.e. 'matching probabilities' (Prob) and *E*-values] for genuine or predicted MED subunits, from a representative panel of eukaryotes (Figure 3B), are compiled in dataset S8. Secondary structure predictions were done at the 'Pôle Bio-Informatique Lyonnais' (PBIL; http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html) and at MPI (<http://toolkit.tuebingen.mpg.de/hhpred>).

RESULTS AND DISCUSSION

Identification of SSMs for each fungal/metazoan MED subunit

When this work was initiated, MED complexes had been isolated only from a small number of animals including a few mammals, the insect *D. melanogaster* and the nematode *C. elegans* as well as from the distantly related yeasts *S. cerevisiae* and *Schizosaccharomyces pombe* (55–60). All these eukaryotes belong to the Opisthokonta supergroup. In order to identify putative MED components in other eukaryotic supergroups, kingdoms and phyla, characteristic 'signature motifs' should be defined for (i) the 22 subunits common to mammals and yeasts, (ii) the eight subunits (Med23–Med30) identified thus far only in animals and (iii) the three *S. cerevisiae* Tail module subunits (Med2, Med3 and Med5) lacking known counterparts in any species outside the *Saccharomyces* group. To this end, a series of PSI-Blast, BlastP and TBLASTN analyses (50,51) were performed as previously reported (41), using the sequences of known MED subunits as queries against >90% completed genome sequences and/or predicted proteomes of 20 metazoans and 20 fungi. These include representative flat and round worms, insects, ascidians, fish, frogs, chicken and mammals, as well as filamentous ascomycetes, basidiomycetes and the zygomycete *Rhizopus oryzae* (listed in Supplementary dataset S1). Conceptual protein sequences obtained by this phylogenomics approach were eventually assigned as putative MED subunits through PSI-Blast analyses using as queries whole-sequence alignments generated by MAFFT (52,53) (see Materials and methods section). Apparent homologs were readily detected for most of the examined opisthokonts (Figure S5 and dataset S2). Of note, the conceptual primary sequences of many annotated proteins appeared incomplete and/or contained improperly assigned portions (e.g. *S. pombe* Cdk8/Prk1 C-terminus and Med13,

Med15 and Med20 N-termini; see dataset S8; details available upon request).

The initial part of this study extends previous cross-species comparisons (33,41,44,61–65), and provides further support for the widely accepted proposal that at least 22 MED subunits from *S. cerevisiae* are conserved in animals (42). MAFFT alignments revealed widely varying numbers of conserved sequence islands within the subunits, from two (for Med3, Med28 and Med31) up to 38 (for Med12). These conserved elements allowed me to assign SSMs (Figure 1B and C and datasets S2–3). As expected, many SSMs are included in regions previously defined as conserved protein domains from systematic database searches, e.g. InterPro domains (66) (listed in dataset S8). It should be stressed that for most SSMs only a few residues have remained unchanged during evolution (datasets S2–3). This divergence is in keeping with the low conservation of most MED proteins. In an extreme example, the ‘universal’ MED component Med31 (see hereafter) comprises only 11 evolutionarily conserved amino acid residues (out of a total of ~150) distributed over two SSMs (dataset S2, p. 109). Indeed, for most SSMs many positions are highly variable though amino-acid changes mostly conserve some biochemical character (hydrophilic, hydrophobic or small size; datasets S2–3). This is the reason why structural counterparts from distantly related species (e.g. *S. cerevisiae* versus human) are not readily apparent for many MED subunits in classical BLAST analyses.

The SSMs defined here are distributed throughout the primary sequences of MED subunits (with the notable exception of Med15; Figure 1B and C and see also below). However, many of them are included within regions delineated as inter-subunit interaction domains (25), while some others are presumably involved in functional connections with dedicated PolIII subunits or GTFs. Consistent with this view, structural modelling predicts that many SSMs adopt amphiphilic α -helices, i.e. with hydrophilic and hydrophobic faces on opposite sides (not shown). With few exceptions, the relative positions of SSMs within the whole sequences have been rather well conserved during evolution, providing additional hallmarks typical of each MED subunit. Lastly, a few SSMs correspond to distinguishable domains identified in many functionally distinct proteins (Figure 1B and C). For example, Med16 includes seven WD repeats in its N-terminal half plus a C-terminal C2-C2 zinc-finger (ZF) motif, while Med27 contains a distinct C-terminal C2-HC ZF motif, Med15 comprises a KIX domain (65,67), Med25 displays a VWFA domain (68), and Med26 exhibits a TFIIS-I/LW domain (69,70). The additional SSMs specifically recovered in Med15–16 and Med25–27 (Figure 1B and C) were thus key elements for distinguishing putative MED subunits among a number of candidate proteins. Similarly, Cdk8 and CycC counterparts were inferred among various Cdk and cyclin family members by the presence of specific evolutionarily conserved motifs (Figure 1B and datasets S2–3), notably an N-terminal motif (SSM#1) of CycC prone to adopt a mobile α -helical conformation (71), and for Cdk8 of a typical kinase

activation segment as well as a C-terminal kinase domain extension motif (SSMs #2 and #4, respectively).

MED subunit conservation, loss and duplication among animals and fungi

Regarding the structural conservation of MED among metazoans several findings should be emphasized. First, except for Med11, Med16 and Med24–26 in a few species (see below), all the remaining mammalian MED subunits have at least one detectable counterpart in the 20 examined animals. Second, *C. elegans* and the related nematode *C. briggsae* display two Med1 and Med27 paralogs as well as four putative MED subunits that have not been reported to date, i.e. Med9, Med24 (Lin-25 in *C. elegans*), Med26 and Med30 (Figure S5 and dataset S2; see also Figure 3B for *C. elegans*). Human Trap100 and worm Lin-25 exhibit 10% identity and 26% similarity over 1183 amino acids (PSI-Blast *E*-value 1.0e-114). The identification here of Lin-25 as the *C. elegans* Med24 homolog is fully consistent with the independent experimental observations that *lin-25/med24* and *sur-2/mdt-23/med23* mutants show indiscernible phenotypes (72), while mammalian Med24 (Trap100) forms a submodule with Med23 (Sur2) (73,74) (below). Third, extensive searches failed to detect Med16 and Med25 counterparts in any of the examined worms. It is intriguing that the malarial vector *Anopheles gambiae*, but not the yellow fever mosquito *Aedes aegypti*, apparently also lacks Med16 and Med25 subunits. Though the sequences of some genomes remain to be fully completed, these data suggest that a few metazoans have lost some MED subunits. Conversely, some species have accumulated several MED subunit paralogs. For example, frogs, fish and worms are predicted to possess two Med1 subunits [large and small forms; note that the *C. elegans* small form has been reported as Med1L in (42)]. Similarly, the mouse, human and chicken genomes encode two Med12, Med13 and Cdk8 paralogs (i.e. Med12/Med12L, Med13/Med13L and Cdk8/Cdk8L, respectively) but a single C-type cyclin, whereas frog and fish genomes display two or three Med13 paralogs but only single Med12, Cdk8 and CycC counterparts (Figure S5). It thus appears possible that in higher chordates core MED may associate with alternative Cdk8 modules. Finally, the three fish species examined are the only metazoans possessing two Med30 paralogs (dataset S2, pp. 107–108).

Regarding the structural conservation of MED among fungi, several findings also deserve to be discussed. First, Med4, Med6–8, Med10–11, Med14, Med17 and Med21–22 are all essential for cell viability in *S. cerevisiae* (75). Further, each has a single detectable structural homolog in the 20 investigated fungi (Figure S5 and dataset S2). These data support the notion that these 10 core subunits play critical architectural roles within MED (see below). Second, the Tail module subunits Med2, Med3 (Pgd1) and Med5 (Nut1), identified to date only in *S. cerevisiae* and other Saccharomycotinae, possess detectable counterparts in 18 or 19 of the 20 examined fungi (Figure S5; see also below). Of note, *S. cerevisiae* Med2 and Med3/Pgd1 form a functional pair (27,28) and interact with Med15 (Gal11) to assemble a separable triad *in vivo* (76).

As apparent Med15/Gal11 subunits could be detected in all the 20 investigated fungi, these data suggest a Med2–Med3–Med15 triad conserved throughout fungal evolution. The interacting domains have not yet been identified for any member of the *S. cerevisiae* Med2–Med3–Med15 triad. Of note, the N-terminal KIX domain of Med15 is absent in the four basidiomycetes, that instead display an internal ARID/BRIGHT domain (not shown).

Third, the human pathogenic yeast *Candida albicans*, but not the closely related candidal species *C. dubliniensis* (not shown), is predicted to possess 14 Med2 paralogs (dataset S2, pp. 6–8). Twelve of them have been previously referred to as the CTA2 family (77). In agreement with a likely MED subunit, CTA2 was initially identified in a one-hybrid screen in *S. cerevisiae* for *C. albicans* proteins with transcriptional activating properties (78). Though their expression and incorporation within MED have yet to be confirmed, it has been proposed that CTA2 family members may contribute to the increased prevalence and virulence of *C. albicans* versus *C. dubliniensis* (77).

Fourth, despite extensive searches, Med5 and Med16 counterparts were not detected in *S. pombe* (Figure 3B) and in the related fission yeasts *Schizosaccharomyces octosporus* and *Schizosaccharomyces japonicus* (not shown). In contrast, likely Med5 (above) and Med16 subunits could be detected in all the other examined fungi. Of note, *S. cerevisiae* Med5/Nut1 and Med16/Sin4 interact in a two-hybrid assay (25) and Med5/Nut1 is lost from MED purified from a *SIN4* (*MED16*) deletion strain (43). Taken together, these data indicate that the fission yeasts may lack a part of the Tail module, as suggested above for worms.

Lastly, the *P. chrysosporium* and *C. cinereus* basidiomycetes as well as the zygomycete *R. oryzae* possess single counterparts for the entire set of identified budding yeast MED subunits (Figure S5 and dataset S2; for *C. cinereus* and *R. oryzae*, see also Figure 3B and dataset S8). In addition to this 25 subunit set, single likely equivalents of the metazoan Med23 (Sur2) and Med25 (Arc92) subunits could be also identified in *R. oryzae* (*E*-values 0.0 and 3.0e-47, respectively) (Figure 3B and datasets S5 and S8) as well as in the related mucorale *Phycomyces blakesleeanus* (not shown). Candidate Med25 subunits were also detected in the four basidiomycetes (*E*-values ranged from 3.0e-34 to 7.0e-43; datasets S5 and S8), but are apparently lacking in the investigated ascomycetes. Although detectable in *R. oryzae*, Med23 equivalents could not be identified in any of the 19 examined basidiomycetes/ascomycetes. These data suggest that Med23 and Med25 may have been incorporated within MED before the divergence between fungi and animals, then lost in some of the ancestors of present day fungal subphyla (see also below).

Metazoan Med29, Med27 and Med24 are apparent structural homologs of fungal Med2, Med3 and Med5 Tail subunits

The detection of likely Med2, Med3/Pgd1 and Med5/Nut1 subunits in most of the 20 examined fungal proteomes (above) led me to examine whether these three Tail

module components might also be conserved in higher eukaryotes. First, among the eight MED subunits thus far identified only in animals (Figure 1C), a close inspection of those displaying similar sizes and/or number of SSMs suggested that Med2 and Med5 might be the missing fungal counterparts of metazoan Med29 (known as Intersex in *D. melanogaster*) and Med24 (Trap100), respectively (Figure 1, compare panels B and C). Second, overall size consideration and the presence of a domain homologous to the C-terminal ZF motif of the metazoan Med27 (Crsp34) subunit at the C-terminal end of likely fungal Med3 subunits (i.e. *S. pombe*, *Aspergillus nidulans*, *A. fumigatus*, *Coccidioides posadasii*, *C. cinereus*, *P. chrysosporium* and *R. oryzae*; dataset S2, p. 9) indicated that human Med27/Crsp34 could be equivalent to *S. cerevisiae* Med3/Pgd1. Note that the apparent *S. pombe* Med3 homolog (i.e. Pcm3) was found in purified MED and had been already assigned as a fission yeast Med27 homolog (42). Using overall alignments as query sequences (dataset S4) all these predictions were very well supported by ‘jump-starting’ PSI-Blast analyses (see Materials and methods section). As shown in Figure 2A, structural similarities between *S. cerevisiae* Med2 and human Med29/Intersex were then readily detected with an *E*-value of 3.0e-32, both primary sequences displaying 10% identity and 23% similarity over 181 amino acids. For Med27/Crsp34 versus Med3/Pgd1 the *E*-value was 4.0e-12, with 13% identity and 25% similarity over 198 amino acids (Figure 2B). Given that *S. cerevisiae* and some other ascomycetes lack the Med27 ZF domain, its functional significance in MED activity remains to be deciphered. However, it is intriguing that *S. pombe* Med27/Pcm3 is located proximally to the Med18–Med20 pair on the periphery of the Head domain (79) and mammalian Med29/Intersex interacts physically with Med20/Trfp (80). Altogether these data indicate that the apparent counterparts of the *S. cerevisiae* Tail subunits Med2 and Med3 belong to the Head module both in fission yeasts and in higher eukaryotes, and suggest that structural rearrangements might have occurred during evolution. This hypothesis is consistent with the fact that *S. pombe* MED apparently lacks many Tail subunits (above). Interestingly, the fission yeast Med15 homolog has been also linked to the Head domain (79). Thus, it is possible that Med2, Med3 and Med15 interact with Head module subunits, not only in *S. pombe* and in higher eukaryotes but also in *S. cerevisiae*. Such a possibility in the budding yeast is indeed supported by expression profiling analyses (34) and by the recent discovery of unanticipated physical links between the Tail and Head modules (81).

Regarding fungal Med5, structural similarities with metazoan Med24 are apparent over their entire length (dataset S4, pp. 8–12). Indeed, human Med24/Trap100 and *S. cerevisiae* Med5/Nut1 exhibit 9% identity and 22% similarity over 1176 amino acids (*E*-value 1.0e-112) (Figure 2C). Moreover, secondary structure models for fungal Med5 and metazoan Med24 fully matched (not shown). Thus, despite the low level of overall identity, these observations are good evidence for the proposed orthology assignment. Consistent with this, apparent

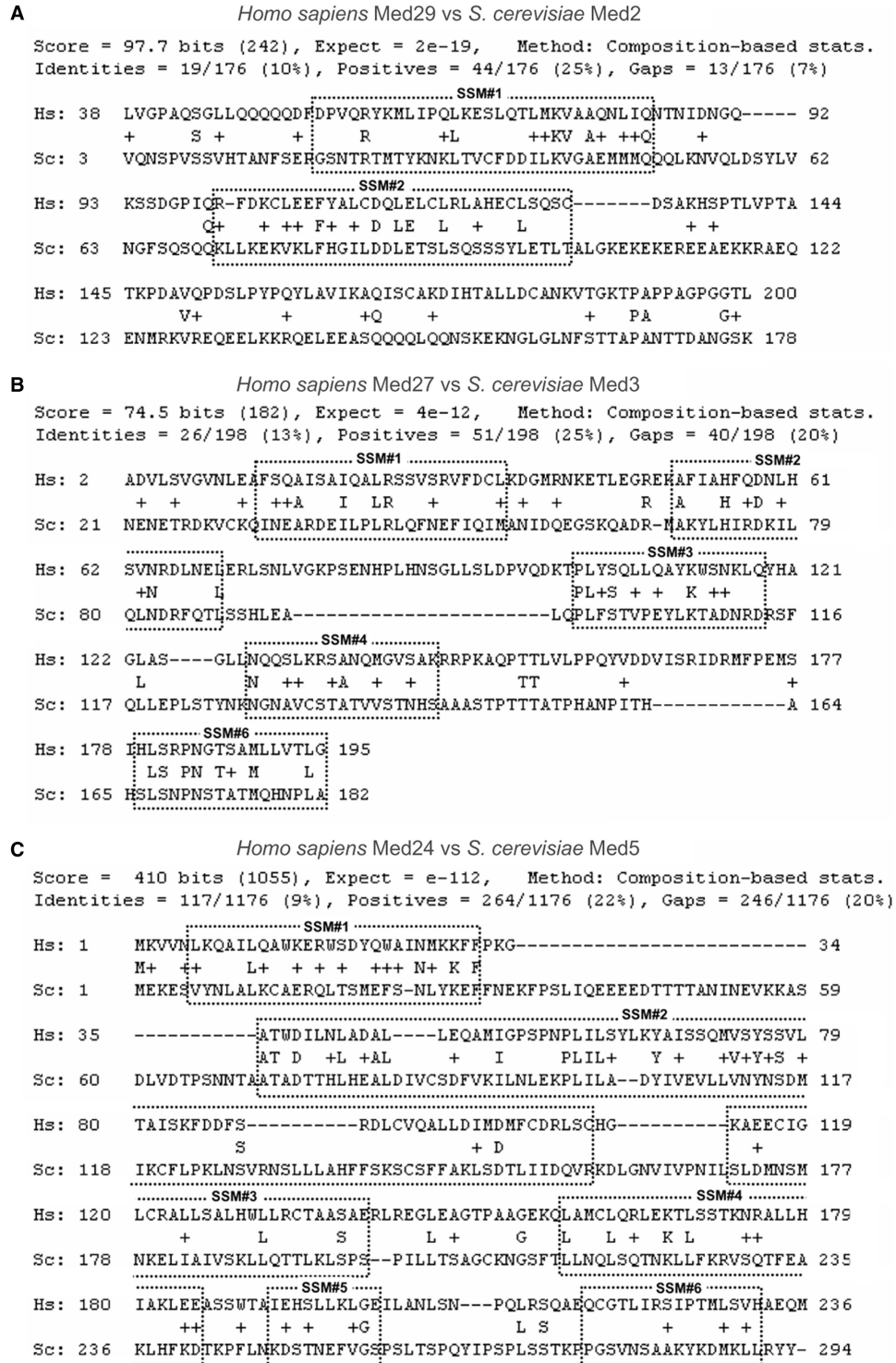


Figure 2. Fungal Med2, Med3 and Med5 are apparent homologs of metazoan Med29, Med27 and Med24, respectively. (A–C) PSI-blast alignments of human (Hs) Med29 (A), Med27 (B) and Med24 (C) with *S. cerevisiae* (Sc) Med2, Med3 and Med24 primary sequences, respectively, are shown. Numbers indicate amino-acid positions within human (top lines) and fungal (bottom lines) primary sequences. Portions corresponding to evolutionarily conserved SSMs are denoted by dashed-line boxes.

Med5/Med24 subunits were detected in all but two of the 40 opisthokonts investigated (the lone exceptions being *S. pombe* and the flat worm *Schistosoma mansoni*; Figure S5). Given that *S. cerevisiae* Med5/Nut1 exhibits a histone-acetyl transferase (HAT) activity (82), it remains

to be tested whether mammalian Med24/Trap100 also modifies histones *in vitro* and/or interacts with nucleosomes *in vivo*. The finding that the fungal Tail module subunit Med5 is homologous to the metazoan Med24 subunit fits nicely with the observation that Med24/

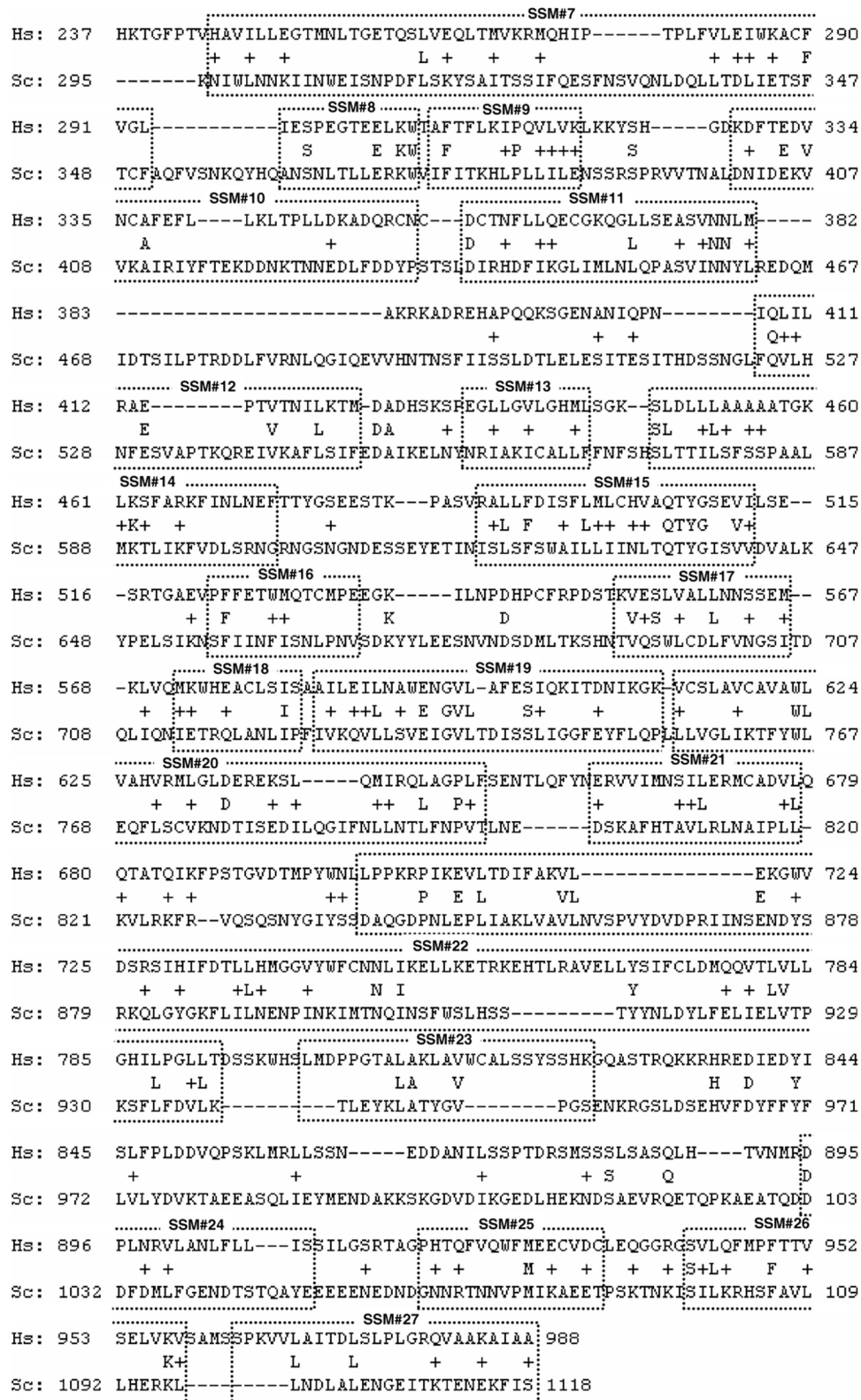


Figure 2. Continued.

Trap100 is closely associated with the predicted mammalian Med16/Sin4 counterpart (i.e. Trap95) (73,74), as this has been recently reported for *S. cerevisiae* Med5/Nut1 and Med16/Sin4 (43). Furthermore, the identification of

likely Med16/Sin4/Trap95, Med23/Sur2 (above) and Med24/Trap100/Med5/Nut1 subunits in zygomycetes is consistent with the proposed role for mammalian Med23/Sur2 in assembling a submodule with Med24/Trap100

(see above for shared functions in nematodes) and Med16/Trap95 (73,74).

Lastly, together with data from previous cross-species comparisons (12,41,42), these new findings indicate that none of the *S. cerevisiae* MED subunit is specific to fungi: the full set has been conserved in animals. In forthcoming publications, I therefore propose to refer metazoan Med24, Med27 and Med29 to as Med24/Med5, Med27/Med3 and Med29/Med2, respectively, and conversely fungal Med2, Med3 and Med5, to as Med2/Med29, Med3/Med27 and Med5/Med24, respectively.

Identification of putative MED subunits in a large sample of non-opisthokont eukaryotes

During the course of this study, >90% completed (i.e. 5- to 10-fold coverage) genome sequences were released for 30 non-opisthokont species distributed among most major eukaryotic supergroups, taxa or phyla, including Microsporidia, Amoebozoa (amoebae), Viridiplantae (land plants), Chlorophytae (green algae), Rhodophytae (red algae), Heterokonta (oomycetes and diatoms), Apicomplexa, Ciliata, Kinetoplastida, Trichomonadida and Diplomonadida [see dataset S1 for species names, references and websites, and see ref. (83,84) and Figure 3A for eukaryotic phylogenetic trees]. To identify apparent MED subunits in these non-opisthokont eukaryotes, the following step-by-step approach was performed. First, entire primary sequences or selected portions encompassing one or several SSM(s) of *bona fide* or predicted metazoan/fungal MED subunits were used as queries in series of PSI-Blast, BlastP and TBlastN analyses. Candidate MED proteins were then selected by focusing not only on the highest scores but also on conservation of most SSMs defined above and of their spacing/distribution within the entire primary sequences. Regarding Med4 and Med19, the presence at their C-termini of a typical acidic or basic region, respectively (see dataset S2, pp. 11 and 83, respectively), was a further selection criterion. Among the initially selected candidate MED subunits, orphan proteins without any obvious similarity to other protein classes (as determined by InterPro domain analyses) were retained and one was eventually assigned as a likely MED subunit by 'jump-starting' PSI-Blast analyses (as detailed in Materials and methods section).

In fine, this approach allowed identification of at least one apparent MED subunit (i.e. Med31) in all but one of the examined non-opisthokont species (Figure 3B and S6; see dataset S5 for overall alignments). Significantly, the lone exceptions are the kinetoplastids (trypanosomes and *Leishmania major*) in which gene regulation occurs primarily at the post-transcriptional level (85). As the kinetoplastids also lack many GTFs (86), the available data support the notion that these atypical eukaryotes possess simplified PolIII initiation machineries and may thus represent early-diverging eukaryotes. Alternatively, ancestors of extant kinetoplastids may have deleted their MED subunits. Surprisingly, although structural homologs of Med31 (Soh1) could be readily detected in all but one of the investigated eukaryotic phyla, this subunit is

not critically required for cell viability in budding or fission yeast (87–89).

A conserved framework of core MED subunits from protists to animals

Regarding the structural conservation of MED from lower to higher eukaryotes, several outcomes raised by this comprehensive comparative genomics study deserve to be emphasized. First of all, it is worth noting that the predicted non-opisthokont MED subunits include most of the SSMs previously defined from primary sequence comparisons of animal/fungal subunits (dataset S5). For example, 13 out of 15 SSMs that characterize Med1, one of the most weakly conserved subunits among opisthokonts (human and *S. cerevisiae* primary sequences display only 9% identity and 22% similarity), are detected in the sequences of the social amoebae *D. discoideum*, the archamoebae *Entamoeba histolytica*, the red algae *Cyanidioschyzon merolae*, the trichomonad *T. vaginalis* and the diplomonad *G. lamblia* (*E*-values ranged from 7.0e-97 to 1.0e-107). Taken together with the fact that for most MED subunits the SSMs are distributed throughout their primary sequences (above, Figure 1B and C), these new comparative genomics data strengthen the view that their 3D structures individually, and by inference the overall architecture of the entire complex, have been conserved from protists to man.

Second, apart from kinetoplastids (above), the 10 core subunits critically required for cell viability in yeast have apparent homologs not only in all opisthokonts (above) but also in all or most of the examined eukaryotes (Figure S6). In fact, only counterparts of the Head subunits Med11 and Med22 remained undetectable in some non-opisthokont eukaryotes, possibly as a result of highly divergent primary sequences (or alternatively to release of uncompleted genomes). Significantly, apparent Med4, Med6-8, Med10, Med11, Med14, Med17 and Med21–22 subunits were detected in microsporidians (*E*-values 3.0e-28, 2.0e-55, 1.0e-34, 3.0e-39, 3.0e-28, 8.0e-18, 1.0e-113, 1.0e-113, 2.0e-21 and 1.0e-24, respectively, for *E. cuniculi* versus human primary sequences; see also dataset S8). These intracellular parasites are thought to be of fungal origin but with compacted and highly reduced genomes (~2000 genes) (90,91). The comparative genomics data thus suggest that these 10 subunits constitute a conserved framework from which other MED components are assembled. This view is in fact very well supported by a two-hybrid-based protein interaction map and *in vitro* module reconstitution experiments (24,25,31). In *S. cerevisiae*, the 10 essential core subunits are indeed linked to one another through direct physical interactions. Among these, Med4, Med7, Med8, Med11, Med14, Med17, Med21 and Med22 play critical scaffold roles for Tail, Middle and Head (sub)-module assembly (24,31). Crystallographic studies have shown that the *S. cerevisiae* Middle subunits Med7 and Med21 (Srb7) interact to form a flexible hinge thought to play a pivotal role within MED (92). Consistent with this, protein homology detection and structure predictions with the HHpred interactive server (93) indicated that apparent protistan Med7 and Med21

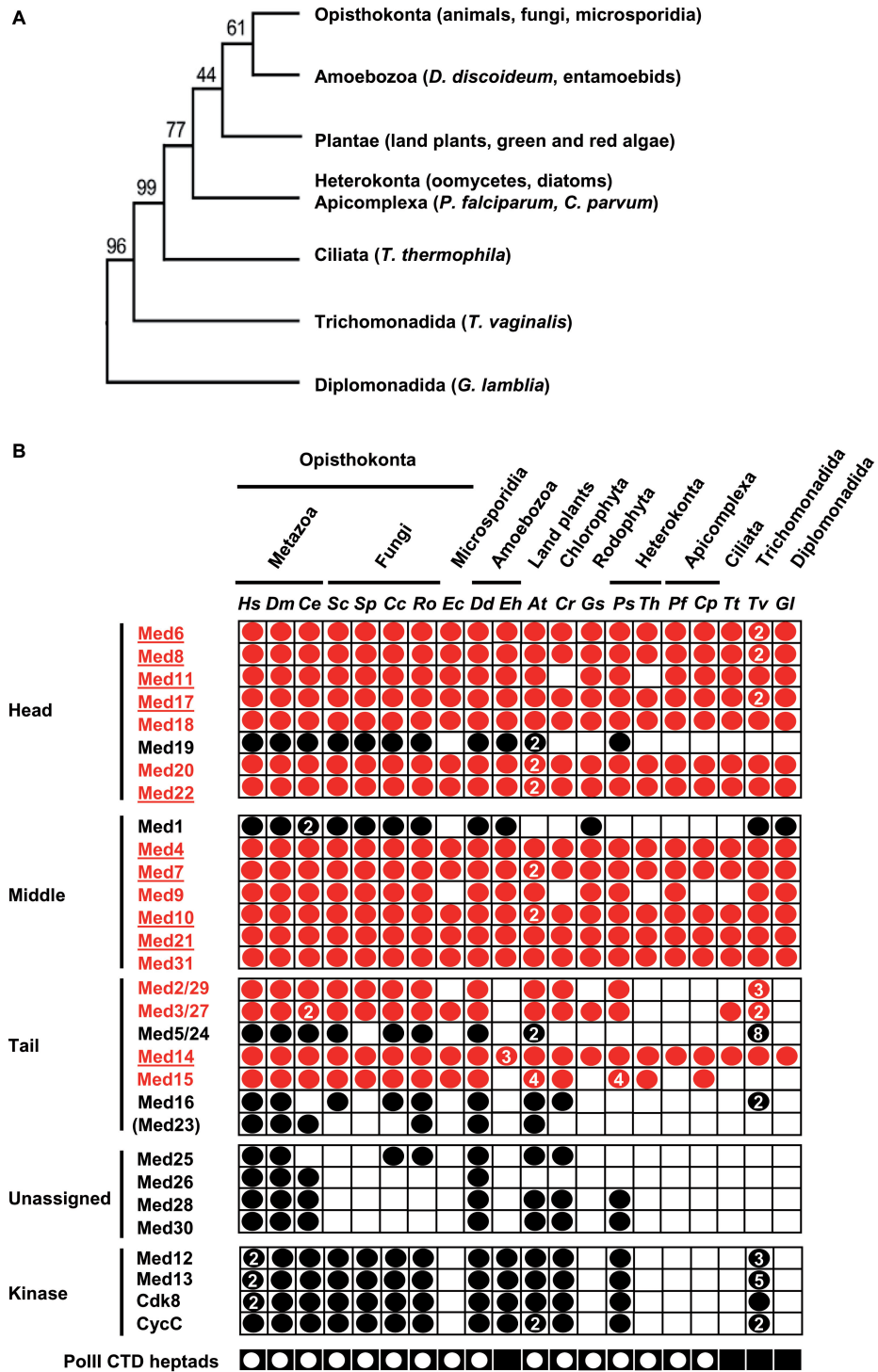


Figure 3. Mediator subunits conserved across the eukaryotic kingdom. (A) Schematic representation of the best tree from amino-acid sequence analyses of 22 evolutionarily conserved protein and two rRNA genes collected from species distributed among eight higher-order groups of eukaryotes. Bootstrap proportion (BP) values are shown on internal branches [after ref. (106)]. (B) Subunits are grouped to Head, Middle, Tail and Kinase. Subunits underlined being critically required for cell viability (Figure 1A). Except possibly Med23 (see text), the subunits only identified to date in purified metazoan MED complexes remain to be assigned to a specific module. Indicated species are representatives of the examined eukaryotic taxa (for a comprehensive analysis see Figure S6): *H. sapiens* (Hs), *D. melanogaster* (Dm), *C. elegans* (Ce), *S. cerevisiae* (Sc), *S. pombe* (Sp), *C. cinereus* (Cc), *R. oryzae* (Ro), *E. cuniculi* (Ec), *D. discoideum* (Dd), *E. histolytica* (Eh), *A. thaliana* (At), *C. reinhardtii* (Cr), *G. sulphuraria* (Gs), *P. sojae* (Ps), *T. pseudonana* (Th), *P. falciparum* (Pf), *C. parvum* (Cp), *T. thermophila* (Tt), *T. vaginalis* (Tv) and *G. lamblia* (Gl). Except for Med28 and Med29/Med2 for which no conserved domain signatures have been yet defined, HHpred analyses predicted with high statistical confidence (Prob >98%, with *E*-values >1.0e-7) that most if not all identified proteins are likely *bona fide* MED subunits (see dataset S8). Dots indicate that at least one apparent homolog is present in the investigated genomes. In the cases where several paralogs were detected, their number is indicated. Broadly conserved subunits, which include those essential for yeast viability, are depicted in red and may constitute a framework around which other subunits are assembled. White dots in the bottom row indicate the presence of repetitive heptads within the Rpb1 CTD.

subunits from *Thalassiosira pseudonana* (a diatom), *Tetrahymena thermophila* (a ciliate), *T. vaginalis* and *G. lamblia* are all prone to adopt spatial conformations highly similar to those of their budding yeast counterparts, with high-confidence parameters (Prob = 100% with *E*-values from 6.2e-22 to 1.2e-39; dataset S8).

Third, in addition to Med31 and the 10 essential framework/scaffold subunits, likely Med29/Med2, Med27/Med3, Med9, Med15, Med18 and Med20 subunits could be detected in species distributed among all or most eukaryotic phyla (Figure 3B and S6; see also datasets S5 and S8). Taken together with previous observations, these results are consistent with a widely conserved core MED consisting of at least 17 subunits: Med2–4, Med6–11, Med14–15, Med17–18, Med20–22 and Med31 (Figure 4). Significantly, apparent counterparts of *S. cerevisiae* Med18 (Srb5) and its direct molecular partner Med20 (Srb2) (32) could be detected in a broad range of species, again including microsporidians (with *E*-values of 4.0e-20 and 9.0e-22, respectively, for *E. cuculi* versus *S. cerevisiae* primary sequences). Significantly, the putative *T. vaginalis* (or *G. lamblia*) Head module is also predicted to include Med8, Med18 and Med20 subunits (*E*-values 1.0e-24, 3.0e-20 and 2.0e-30, respectively, for *T. vaginalis* versus *S. cerevisiae* primary sequences). Again, comparative modelling indicated that apparent *T. vaginalis* (or *G. lamblia*) Med18 and Med20 subunits are prone to adopt spatial conformations significantly similar to those determined for the yeast proteins (HHpred Prob = 100% with *E*-values from 4.4e-36 to 5.6e-45; dataset S8). These data provide evidence for an ancient evolutionarily conserved regulatory role of the Med8–Med18–Med20 triad. The *S. cerevisiae* Med18–Med20 pair is known to be required for stable PIC formation, efficient basal transcription and response to transcriptional activators *in vitro* (27,94). More recent data indicate that the Med8–Med18–Med20 triad constitutes a multipartite TBP-binding site (32). Together with these experimental data, comparative genomics thus provides compelling support for an evolutionarily conserved regulatory MED function in TBP recruitment to gene promoters and PIC assembly. Lastly, regarding the four remaining core subunits Med1, Med5/Med24, Med16 and Med19, whose counterparts are apparently lacking in many non-opisthokonts (Figure 3B and S6), it is significant that the *S. cerevisiae* Med5/Med24/Nut1 HAT directly interacts with both Med1 and Med16/Sin4 (25). An entire MED area, overlapping the Middle and Tail modules, may thus be lacking (or highly divergent) in some ‘lower eukaryotes’, as suggested above for some opisthokonts.

Evidence for a conserved Med2–Med3–Med15 triad, prone to adapt species-specific regulatory signals

In *S. cerevisiae*, the Tail module is thought to act as a dedicated sensor of incoming regulatory signals from diverse gene-specific transcription factors (95). Med14 (Rgr1) is the only Tail subunit for which structural homologs could be readily detected in all the examined eukaryotic species, with the notable exception of the kinetoplastids (above) (Figure 3B and S6; see also dataset

S5). These results are consistent with the proposal that *S. cerevisiae* Med14/Rgr1 anchors the Tail module to the universal Med4 and Med10 Middle module subunits (24,25). Note that some likely apicomplexan Med14 subunits harbour very large C-terminal extensions. Indeed, the *Plasmodium falciparum* and *Cryptosporidium parvum* Med14 homologs [known as CG2 for the malaria parasite *P. falciparum* (96)] include 2729 and 2806 amino-acid residues.

In addition to Med14, apparent counterparts of the *S. cerevisiae* Med2–Med3–Med15 triad are detected in *D. discoideum*, the three plants and the two oomycetes (Figure 3B and S6). Indeed, all investigated plants and oomycetes display three to five Med15 paralogs [for *A. thaliana*, see ref. (65)]. These observations offer the possibility of alternative Med2–Med3–Med15 triads which may regulate different gene sets. Most of the primary Med15 sequence is markedly divergent not only between phyla, or species sharing the same subphylum (dataset S5, pp. 68–72), but also between paralogs (data not shown). The highly divergent regions of the Med15 Tail subunit may thus correspond to dedicated surfaces interacting with diverse species-specific regulatory signals. Consistent with such an accommodation role, Med15/Arc105/Gal11 physically interacts with many unrelated gene-specific transcription factors both in metazoans and *S. cerevisiae* (67,95,97). Additionally, likely Med27/Med3 and Med15 counterparts are detected even in the tiny microsporidian proteomes [Figure 3B and S6; see also dataset S5 and (65)]. Taken together the available functional data and the apparent widespread conservation of the Med2/Med29, Med3/Med27, Med14 and Med15 subunits most likely reflect their key importance in the reception of species-specific regulatory signals from lower to higher eukaryotes.

A common set of ~30 MED subunits in diversified eukaryotic phyla

In ‘higher’ eukaryotes, it has been proposed that MED has incorporated a set of novel subunits interacting with metazoan-specific transcriptional regulators (11). In marked contrast with this view, the present work indicates that mammalian Med23–Med30 subunits possess structural counterparts not only in fungi (for Med24, Med27 and Med29, above) but also in amoebae, plants, red algae and/or oomycetes (Figure 3B and S6; see also dataset S5 and below). In the three plants examined, homologs of all but two mammalian MED subunits (i.e. Med1 and Med26) were detected. Regarding Med23 and Med25, these findings are consistent with their prior detection in zygomycetes (for both Med23 and Med25) and basidiomycetes (for Med25) (above). A recent independent biochemical study published while this article was in preparation revealed that the *A. thaliana* subunits, including a Med27/Med3 equivalent, predicted here are all *bona fide* MED components (98). Among some additional MED components apparently specific to plants, Med32 and two related large subunits termed Med33a/b are in fact homologous to Med29/Med2 and Med24/Med5, respectively (see dataset S5). Indeed, *A. thaliana* Med32

and human Med29/Intersex (or *S. cerevisiae* Med2) primary sequences display 11% identity and 31% similarity (10/29%) over 119 (158) amino acids (*E*-values 9.0e-31 and 2.0e-16, respectively). Similarly, human Med24/Trap100 and *A. thaliana* Med33a (or Med33b) display 9% identity and 23% similarity (9/20%) over their entire length (*E*-values 1.0e-147 and 1.0e-134, respectively). These data thus indicate that not only metazoans (above) but also plants have apparent homologs of all fungal Tail subunits, providing further compelling support for evolutionarily conserved roles in the reception and accommodation of regulatory signals (above). Also, note that Med1 and Med26 are the only metazoan subunits without detected plant equivalents. In contrast with this observation, the social amoeba *D. discoideum* is predicted to possess single counterparts of the entire set of known human MED subunits (Figure 3B and datasets S5 and S8), including not only of Med1/Trap220 (above) but also of Med24/Med5 and Med26 (Crsp70) (*E*-values 5.0e-37 and 2.0e-88, respectively). These data thus provide support for the recent proposal that Amoebozoa and Opisthokonta are more closely related taxa than previously thought (99,100).

It should also be emphasized here that a large set of 44 apparent MED subunits could be detected in the flagellated parasite *T. vaginalis* (Figure 3B and datasets S5 and S8). The trichomonad PolII initiation machinery appears in fact more metazoan than protistan (101). *Trichomonas vaginalis* is predicted to possess complete or near complete Middle and Head modules, respectively (see also below). Regarding the Tail module equivalent, although two Med16, eight Med24/Med5, two Med27/Med3 and three Med29/Med2 paralogs could be retrieved in PSI-Blast analyses (*E*-values ranged from 6.0e-20 to 1.0e-149 for *T. vaginalis* versus human primary sequences), trichomonad Med15 homologs remained undetectable, presumably owing to lack of a KIX domain (as for basidiomycetes, above) or to highly divergent SSMs. *Trichomonas vaginalis* is predicted to possess several paralogs not only of Med16, Med24/Med5, Med27/Med3 and Med29/Med2 but also of Med6 (two copies), Med8 (2), Med12 (3), Med13 (5), Med17 (2) and CycC (2) (Figure 3B and dataset S5; see also below). In Trichomonadida, alternative subunit pairings may thus lead to the assembly of distinct MED complexes, notably Head, Tail and Cdk8 modules (see below). Taken together these comparative genomics data and recent biochemical studies from *A. thaliana* (98), strongly argue for the emergence of a large modular MED-like complex containing at least 26 subunits, early on during eukaryotic evolution (Figure 4). Further, I speculate that the ‘versatility’ of this ancestral complex might have largely contributed to the diversification of genetic circuitries (see also below).

Putative MED Cdk8 module subunits are detected in organisms lacking PolII CTD heptads

The phosphorylation status of Rpb1 CTD heptads is thought to play a critical role in orchestrating the reversible interaction between PolII and MED (15) and this process may be fine-tuned by the Cdk8 kinase module

(19,35). Owing to detection of many likely MED subunits in most eukaryotes, it is somewhat unexpected that PolII lacks CTD heptads in *E. histolytica*, *P. tetraurelia*, *T. vaginalis* and *G. lamblia* (Figure 3B, bottom rows) (102,103). Even more surprising, *T. vaginalis* is predicted to possess a single Cdk8-type Cdk kinase (*E*-value 1.0e-146), two C-type cyclins (*E*-values 2.0e-81 and 7.0e-80), three Med12 (*E*-values 1.0e-163 to 0.0) and five Med13 paralogs (*E*-values 1.0e-88 to 1.0e-119), whereas the archamoebae *E. histolytica* apparently displays single likely Cdk8, CycC, Med12 and Med13 subunits (*E*-values 1.0e-136, 3.0e-89, 1.0e-127 and 1.0e-125, respectively) (Figure 3B and dataset S5). Conversely, in agreement with previous phylogenetic studies (104,105), putative Cdk8 module members remained undetected in all the examined apicomplexans and microsporidians, even though these species do possess PolII CTD heptads (Figure 3B and S6).

Taken together these observations indicate that during evolution (i) Rpb1 CTD repetitive heptads were not initially required for assembling a PolII–MED holoenzyme complex *in vivo* and (ii) Cdk8 was not co-opted within MED as a dedicated CTD Ser2/Ser5-directed kinase. Consistent with these hypotheses, it has been shown in *S. cerevisiae* that Rpb1 is not the sole point of contact between the PolII polymerase and MED (23). Furthermore, Cdk8 phosphorylates many transcriptional regulators (34,36–40). Finally, these data are amenable to a separable Cdk8 module that appeared early on during eukaryotic evolution, but was subsequently lost in some phyla or groups of species, as suggested above for parts of the Tail domain.

CONCLUSIONS

Discovered in fungi then in metazoans, MED is a versatile modular interface conveying specific regulatory information from diverse DNA-bound transcription factors to the basal PolII initiation machinery. Despite its key importance in fine-tuned gene regulation, the underlying molecular mechanisms remain poorly understood. Extending previous analyses, this comprehensive comparative genomics study has revealed an astonishingly widespread evolutionary conservation of most if not all core MED subunits initially identified in *S. cerevisiae*, across a vast spectrum of eukaryotes. Surprisingly, among the subunits thus far attributed specifically to mammalian complexes, none are apparently truly animal-specific. Indeed, all the 33 known human MED subunits have single structural counterparts in the cellular slime mold *D. discoideum* and all but two are also detected in plants. Significantly, all the predicted *A. thaliana* subunits have now been identified in purified MED preparation, suggesting that most if not all proteins identified throughout this work will prove to be genuine MED components.

Together with the identification of evolutionarily conserved motifs (i.e. SSMs) distributed throughout the primary sequences of most subunits, the comparative genomic analyses reported here provide compelling evidence that the MED architecture has been conserved

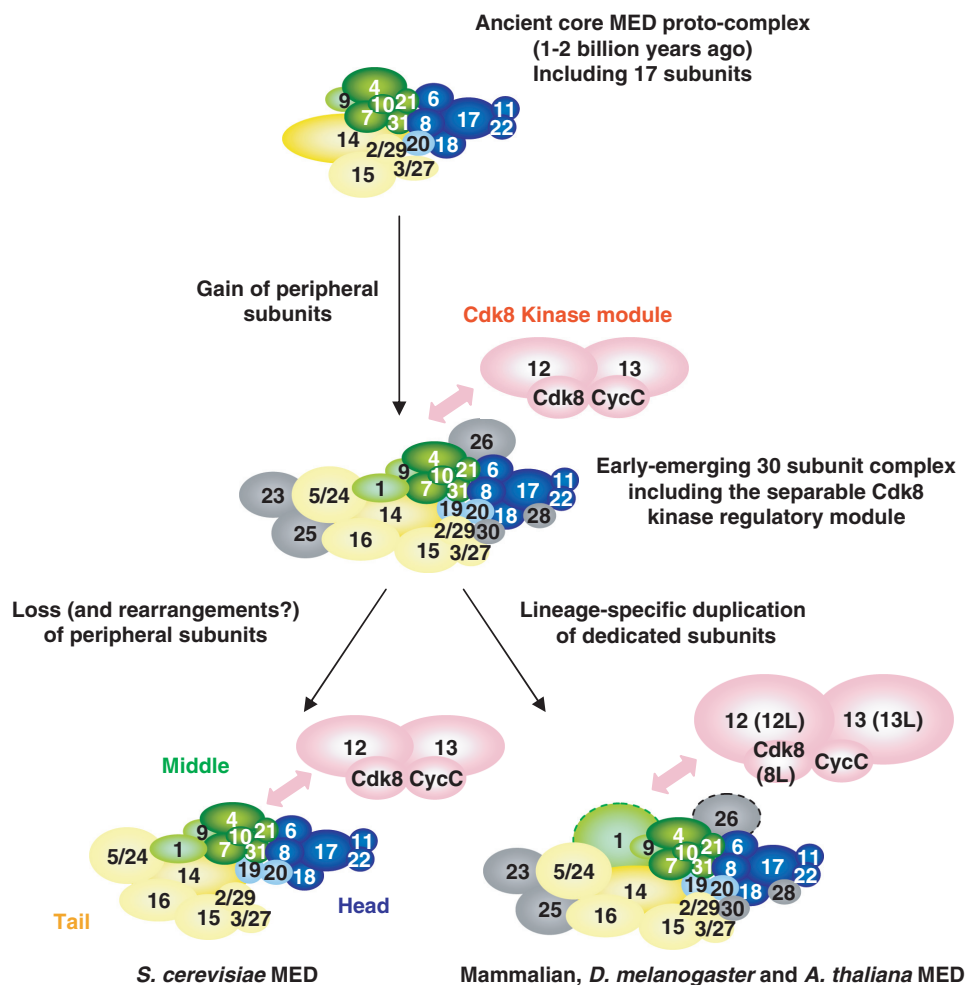


Figure 4. A model for MED architectural organization during eukaryotic evolution, from a primitive scaffold to current complexes. MED subunits have been assigned to Head, Middle, Tail and Cdk8 kinase modules according to a connection map from *S. cerevisiae* (25), as shown in Figure 1A. Subunits initially identified in purified metazoan complexes and without homologs in budding yeast MED are indicated in grey. Hypothetical structural organizations of the present-day mammalian, *Drosophila* and *Arabidopsis* MED complexes as inferred from the current yeast model, orthology assignments and other functional studies. Broken circles outline that Med1 and Med26 subunits are apparently absent in plants (98). Structural rearrangements, notably the location of the Med2/29–Med3/27–Med15 subunits (see text), may have occurred during evolution. Note that subunits of the separable Cdk8 module, which may also have appeared early on during eukaryotic evolution, have detectable homologs (and in a few cases paralogs) only in some extant taxa. It is also noteworthy that during metazoan diversification Med1, Med12 and Med13 have considerably increased their size and have been duplicated in some (sub)-phylum.

throughout eukaryotic evolution, from protists to man. A common set of about 30 MED subunits appearing early on during eukaryotic diversification could thus have accommodated the tremendous complexity of the genetic circuitries, such as those typically found in present-day multi-celled organisms. Further, I speculate that the accommodation process may have been facilitated through specific changes on non-conserved regions, prone to be located on the external complex surface. In other words, one set of ~30 subunits would have sufficed to (i) ‘interpret’ species-specific regulatory signals impinging on dedicated but highly divergent subunits and (ii) ‘transmit’ proper transcriptional instructions to the basal PolIII initiation machinery, possibly through the 17 framework subunits defined here. However, as suggested here for the Med2–Med3–Med15 triad, it remains to be seen whether subunit rearrangements occurred within MED during evolution, contributing to its versatile

flexibility. Regarding these issues, it will undoubtedly be of great interest to determine which SSMs defined here correspond to functional domains required for physical interactions between the widely conserved MED subunits and specific PolIII subunits or GTFs. Lastly, this comprehensive work now paves the way for detailed structure–function analyses of MED activity in transcriptional regulation in lower as well as higher eukaryotes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

I apologize to my colleagues whose work could not be cited due to space limitations. I would like to especially

acknowledge Nicolas Loncle, Muriel Boube and David L. Cribbs for their support throughout this work and their critical readings of the article. I am grateful to Michel Werner for his interest in this project and his timely critical advice on the article. I am indebted to anonymous referees for their helpful comments and suggestions. I also thank Laurent Joulia, Serge Plaza, Gaylord Darras, Benjamin Guglielmi, Christian Marck and Pierre Thuriaux for their encouragement to accomplish this monumental task. Preliminary sequence data for *B. malayi*, *T. gondii*, *C. posadasii*, *E. histolytica*, *E. invadens* and *T. vaginalis* were obtained from the Institute for Genome Research (TIGR) with support from the National Institute of Allergy and Infectious Diseases (NIAID). I acknowledge the DOE Joint Genome Institute (JGI), the University of California and the US Department of Energy (DOE) for availability of *X. tropicalis*, *C. reinhardtii* and *P. tricornutum* sequences. Preliminary sequence data for *T. castaneum* and *A. mellifera* were obtained from the Baylor College of Medicine (BCM) and the National Human Genome Research Institute (NHGRI). I acknowledge the Broad Institute, the Harvard University, the Massachusetts Institute of Technology (MIT) and the Whitehead Institute for Biomedical Research (WIBR) for availability of *C. savignyi*, *A. egypti*, *C. cinerea* and *R. oryzae* sequences. I thank the Michigan State University for the availability of unpublished *G. sulphuraria* genome sequences. Lastly, I acknowledge the *A. locustae* genome project, Marine Biological Laboratory (MBL) at Woods Hole, MA, funded by NSF award number 0135272. This work benefited from the ongoing support of the 'Centre National de la Recherche Scientifique' (CNRS), and grants from the 'Association pour la Recherche contre le Cancer' (ARC) and the 'Agence Nationale pour la Recherche' (ANR). Funding to pay the Open Access publication charges for this article was provided by ARC.

Conflict of interest statement. None declared.

REFERENCES

- Davidson,E.H., Rast,J.P., Oliveri,P., Ransick,A., Caletani,C., Yuh,C.H., Minokawa,T., Amore,G., Hinman,V., Arenas-Mena,C. *et al.* (2002) A genomic regulatory network for development. *Science*, **295**, 1669–1678.
- Levine,M. and Tjian,R. (2003) Transcription regulation and animal diversity. *Nature*, **424**, 147–151.
- Lemon,B. and Tjian,R. (2000) Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.*, **14**, 2551–2569.
- Kornberg,R.D. (2001) The eukaryotic gene transcription machinery. *Biol. Chem.*, **382**, 1103–1107.
- Hahn,S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.*, **11**, 394–403.
- Woychik,N.A. and Hampsey,M. (2002) The RNA polymerase II machinery: structure illuminates function. *Cell*, **108**, 453–463.
- Roeder,R.G. (1998) Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 201–218.
- Asturias,F.J. (2004) RNA polymerase II structure, and organization of the preinitiation complex. *Curr. Opin. Struct. Biol.*, **14**, 121–129.
- Kornberg,R.D. (2005) Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.*, **30**, 235–239.
- Kim,Y.J. and Lis,J.T. (2005) Interactions between subunits of Drosophila mediator and activator proteins. *Trends Biochem. Sci.*, **30**, 245–249.
- Malik,S. and Roeder,R.G. (2005) Dynamic regulation of pol II transcription by the mammalian mediator complex. *Trends Biochem. Sci.*, **30**, 256–263.
- Conaway,R.C., Sato,S., Tomomori-Sato,C., Yao,T. and Conaway,J.W. (2005) The mammalian mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.*, **30**, 250–255.
- Bjorklund,S. and Gustafsson,C.M. (2005) The yeast mediator complex and its regulation. *Trends Biochem. Sci.*, **30**, 240–244.
- Chadick,J.Z. and Asturias,F.J. (2005) Structure of eukaryotic mediator complexes. *Trends Biochem. Sci.*, **30**, 264–271.
- Myers,L.C., Gustafsson,C.M., Bushnell,D.A., Lui,M., Erdjument-Bromage,H., Tempst,P. and Kornberg,R.D. (1998) The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.*, **12**, 45–54.
- Max,T., Sogaard,M. and Svejstrup,J.Q. (2007) Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.*, **282**, 14113–14120.
- Nonet,M., Sweetser,D. and Young,R.A. (1987) Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell*, **50**, 909–915.
- Svejstrup,J.Q., Li,Y., Fellows,J., Gnatt,A., Bjorklund,S. and Kornberg,R.D. (1997) Evidence for a mediator cycle at the initiation of transcription. *Proc. Natl Acad. Sci. USA*, **94**, 6075–6078.
- Liu,Y., Kung,C., Fishburn,J., Ansari,A.Z., Shokat,K.M. and Hahn,S. (2004) Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol. Cell. Biol.*, **24**, 1721–1735.
- Yudkovsky,N., Ranish,J.A. and Hahn,S. (2000) A transcription reinitiation intermediate that is stabilized by activator. *Nature*, **408**, 225–229.
- Asturias,F.J., Jiang,Y.W., Myers,L.C., Gustafsson,C.M. and Kornberg,R.D. (1999) Conserved structures of mediator and RNA polymerase II holoenzyme. *Science*, **283**, 985–987.
- Dotson,M.R., Yuan,C.X., Roeder,R.G., Myers,L.C., Gustafsson,C.M., Jiang,Y.W., Li,Y., Kornberg,R.D. and Asturias,F.J. (2000) Structural organization of yeast and mammalian mediator complexes. *Proc. Natl Acad. Sci. USA*, **97**, 14307–14310.
- Davis,J.A., Takagi,Y., Kornberg,R.D. and Asturias,F.A. (2002) Structure of the yeast RNA polymerase II holoenzyme: mediator conformation and polymerase interaction. *Mol. Cell.*, **10**, 409–415.
- Kang,J.S., Kim,S.H., Hwang,M.S., Han,S.J., Lee,Y.C. and Kim,Y.J. (2001) The structural and functional organization of the yeast mediator complex. *J. Biol. Chem.*, **276**, 42003–42010.
- Guglielmi,B., van Berkum,N.L., Klapholz,B., Bijma,T., Boube,M., Boschiero,C., Bourbon,H.M., Holstege,F.C. and Werner,M. (2004) A high resolution protein interaction map of the yeast mediator complex. *Nucleic Acids Res.*, **32**, 5379–5391.
- Park,J.M., Kim,H.S., Han,S.J., Hwang,M.S., Lee,Y.C. and Kim,Y.J. (2000) In vivo requirement of activator-specific binding targets of mediator. *Mol. Cell Biol.*, **20**, 8709–8719.
- Lee,Y.C., Park,J.M., Min,S., Han,S.J. and Kim,Y.J. (1999) An activator binding module of yeast RNA polymerase II holoenzyme. *Mol. Cell Biol.*, **19**, 2967–2976.
- Myers,L.C., Gustafsson,C.M., Hayashibara,K.C., Brown,P.O. and Kornberg,R.D. (1999) Mediator protein mutations that selectively abolish activated transcription. *Proc. Natl Acad. Sci. USA*, **96**, 67–72.
- Han,S.J., Lee,J.S., Kang,J.S. and Kim,Y.J. (2001) Med9/Cse2 and Gal11 modules of mediator are required for transcriptional repression of distinct group of genes. *J. Biol. Chem.*, **276**, 37020–37026.
- Sakurai,H. and Fukasawa,T. (2000) Functional connections between mediator components and general transcription factors of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **275**, 37251–37256.
- Takagi,Y., Calero,G., Komori,H., Brown,J.A., Ehrensberger,A.H., Hudmon,A., Asturias,F. and Kornberg,R.D. (2006) Head module control of mediator interactions. *Mol. Cell.*, **23**, 355–364.
- Lariviere,L., Geiger,S., Hoepfner,S., Rother,S., Strasser,K. and Cramer,P. (2006) Structure and TBP binding of the mediator head subcomplex Med8-Med18-Med20. *Nat. Struct. Mol. Biol.*, **10**, 10.

33. Borggreffe, T., Davis, R., Erdjument-Bromage, H., Tempst, P. and Kornberg, R.D. (2002) A complex of the *srb8*, -9, -10, and -11 transcriptional regulatory proteins from yeast. *J. Biol. Chem.*, **277**, 44202–44207.
34. van de Peppel, J., Kettelar, N., van Bakel, H., Kockelkorn, T.T., van Leenen, D. and Holstege, F.C. (2005) Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell.*, **19**, 511–522.
35. Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S. and Young, R.A. (1998) Temporal regulation of RNA polymerase II by *Srb10* and *Kin28* cyclin-dependent kinases. *Mol. Cell.*, **2**, 43–53.
36. Nelson, C., Goto, S., Lund, K., Hung, W. and Sadowski, I. (2003) *Srb10/Cdk8* regulates yeast filamentous growth by phosphorylating the transcription factor *Ste12*. *Nature*, **421**, 187–190.
37. Chi, Y., Huddleston, M.J., Zhang, X., Young, R.A., Annan, R.S., Carr, S.A. and Deshaies, R.J. (2001) Negative regulation of *Gcn4* and *Msn2* transcription factors by *Srb10* cyclin-dependent kinase. *Genes Dev.*, **15**, 1078–1092.
38. Vincent, O., Kuchin, S., Hong, S.P., Townley, R., Vyas, V.K. and Carlson, M. (2001) Interaction of the *Srb10* kinase with *Sip4*, a transcriptional activator of gluconeogenic genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **21**, 5790–5796.
39. Hallberg, M., Polozkov, G.V., Hu, G.Z., Beve, J., Gustafsson, C.M., Ronne, H. and Bjorklund, S. (2004) Site-specific *Srb10*-dependent phosphorylation of the yeast mediator subunit *Med2* regulates gene expression from the 2-microm plasmid. *Proc. Natl Acad. Sci. USA*, **101**, 3370–3375.
40. Hirst, M., Kobor, M.S., Kuriakose, N., Greenblatt, J. and Sadowski, I. (1999) *GAL4* is regulated by the RNA polymerase II holoenzyme-associated cyclin-dependent protein kinase *SRB10/CDK8*. *Mol. Cell.*, **3**, 673–678.
41. Boubé, M., Joulia, L., Cribbs, D.L. and Bourbon, H.M. (2002) Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell*, **110**, 143–151.
42. Bourbon, H.M., Aguilera, A., Ansari, A.Z., Asturias, F.J., Berk, A.J., Bjorklund, S., Blackwell, T.K., Borggreffe, T., Carey, M., Carlson, M. et al. (2004) A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell.*, **14**, 553–557.
43. Beve, J., Hu, G.Z., Myers, L.C., Balciunas, D., Werngren, O., Hultenby, K., Wibom, R., Ronne, H. and Gustafsson, C.M. (2005) The structural and functional role of *Med5* in the yeast mediator tail module. *J. Biol. Chem.*, **280**, 41366–41372.
44. Samuelsen, C.O., Baraznenok, V., Khorosjutina, O., Spahr, H., Kieselbach, T., Holmberg, S. and Gustafsson, C.M. (2003) *TRAP230/ARC240* and *TRAP240/ARC250* mediator subunits are functionally conserved through evolution. *Proc. Natl Acad. Sci. USA*, **100**, 6422–6427.
45. Cavalier-Smith, T. and Chao, E.E. (1995) The opalozoan *Apusomonas* is related to the common ancestor of animals, fungi and choanoflagellates. *Proc. Roy. Soc. Lond. B Biol. Sci.*, **261**, 1–6.
46. Steenkamp, E.T., Wright, J. and Baldauf, S.L. (2006) The protistan origins of animals and fungi. *Mol. Biol. Evol.*, **23**, 93–106.
47. Hedges, S.B. (2002) The origin and evolution of model organisms. *Nat. Rev. Genet.*, **3**, 838–849.
48. Hedges, S.B., Blair, J.E., Venturi, M.L. and Shoe, J.L. (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.*, **4**, 2.
49. Stiller, J.W. and Hall, B.D. (2002) Evolution of the RNA polymerase II C-terminal domain. *Proc. Natl Acad. Sci. USA*, **99**, 6091–6096.
50. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
51. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
52. Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059–3066.
53. Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.*, **33**, 511–518.
54. Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
55. Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. and Kornberg, R.D. (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, **77**, 599–608.
56. Kwon, J.Y., Park, J.M., Gim, B.S., Han, S.J., Lee, J. and Kim, Y.J. (1999) *Caenorhabditis elegans* mediator complexes are required for developmental-specific transcriptional activation. *Proc. Natl Acad. Sci. USA*, **96**, 14990–14995.
57. Park, J.M., Gim, B.S., Kim, J.M., Yoon, J.H., Kim, H.S., Kang, J.G. and Kim, Y.J. (2001) *Drosophila* mediator complex is broadly utilized by diverse gene-specific transcription factors at different types of core promoters. *Mol. Cell. Biol.*, **21**, 2312–2323.
58. Spahr, H., Beve, J., Larsson, T., Bergstrom, J., Karlsson, K.A. and Gustafsson, C.M. (2000) Purification and characterization of RNA polymerase II holoenzyme from *Schizosaccharomyces pombe*. *J. Biol. Chem.*, **275**, 1351–1356.
59. Rachez, C. and Freedman, L.P. (2001) Mediator complexes and transcription. *Curr. Opin. Cell. Biol.*, **13**, 274–280.
60. Malik, S. and Roeder, R.G. (2000) Transcriptional regulation through mediator-like coactivators in yeast and metazoan cells. *Trends Biochem. Sci.*, **25**, 277–283.
61. Boubé, M., Faucher, C., Joulia, L., Cribbs, D.L. and Bourbon, H.M. (2000) *Drosophila* homologs of transcriptional mediator complex subunits are required for adult cell and segment identity specification. *Genes Dev.*, **14**, 2906–2917.
62. Gustafsson, C.M. and Samuelsson, T. (2001) Mediator – a universal complex in transcriptional regulation. *Mol. Microbiol.*, **41**, 1–8.
63. Sato, S., Tomomori-Sato, C., Banks, C.A., Sorokina, I., Parmely, T.J., Kong, S.E., Jin, J., Cai, Y., Lane, W.S., Brower, C.S. et al. (2003) Identification of mammalian mediator subunits with similarities to yeast mediator subunits *Srb5*, *Srb6*, *Med11*, and *Rox3*. *J. Biol. Chem.*, **278**, 15123–15127.
64. Tomomori-Sato, C., Sato, S., Parmely, T.J., Banks, C.A., Sorokina, I., Florens, L., Zybailov, B., Washburn, M.P., Brower, C.S., Conaway, R.C. et al. (2004) A mammalian mediator subunit that shares properties with *Saccharomyces cerevisiae* mediator subunit *Cse2*. *J. Biol. Chem.*, **279**, 5846–5851.
65. Novatchkova, M. and Eisenhaber, F. (2004) Linking transcriptional mediators via the GACKIX domain super family. *Curr. Biol.*, **14**, R54–R55.
66. Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Buillard, V., Cerutti, L., Copley, R. et al. (2007) New developments in the InterPro database. *Nucleic Acids Res.*, **35**, D224–D228.
67. Yang, F., Vought, B.W., Satterlee, J.S., Walker, A.K., Jim Sun, Z.Y., Watts, J.L., DeBeaumont, R., Saito, R.M., Hyberts, S.G., Yang, S. et al. (2006) An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature*, **442**, 700–704.
68. Mittler, G., Stuhler, T., Santolin, L., Uhlmann, T., Kremmer, E., Lottspeich, F., Berti, L. and Meisterernst, M. (2003) A novel docking site on mediator is critical for activation by VP16 in mammalian cells. *EMBO J.*, **22**, 6494–6504.
69. Booth, V., Koth, C.M., Edwards, A.M. and Arrowsmith, C.H. (2000) Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J. Biol. Chem.*, **275**, 31266–31268.
70. Ling, Y., Smith, A.J. and Morgan, G.T. (2006) A sequence motif conserved in diverse nuclear proteins identifies a protein interaction domain utilised for nuclear targeting by human TFIIS. *Nucleic Acids Res.*, **34**, 2219–2229.
71. Hoepfner, S., Baumli, S. and Cramer, P. (2005) Structure of the mediator subunit cyclin C and its implications for CDK8 function. *J. Mol. Biol.*, **350**, 833–842.
72. Nilsson, L., Li, X., Tiensuu, T., Auty, R., Greenwald, I. and Tuck, S. (1998) *Caenorhabditis elegans* lin-25: cellular focus, protein expression and requirement for *sur-2* during induction of vulval fates. *Development*, **125**, 4809–4819.
73. Ito, M., Okano, H.J., Darnell, R.B. and Roeder, R.G. (2002) The TRAP100 component of the TRAP/mediator complex is essential in broad transcriptional events and development. *EMBO J.*, **21**, 3464–3475.

74. Stevens, J.L., Cantin, G.T., Wang, G., Shevchenko, A. and Berk, A.J. (2002) Transcription control by E1A and MAP kinase pathway via Sur2 mediator subunit. *Science*, **296**, 755–758.
75. Bjorklund, S., Buzaite, O. and Hallberg, M. (2001) The yeast mediator. *Mol. Cells*, **11**, 129–136.
76. Zhang, F., Sumibcay, L., Hinnebusch, A.G. and Swanson, M.J. (2004) A triad of subunits from the Gal11/tail domain of Srb mediator is an in vivo target of transcriptional activator Gen4p. *Mol. Cell. Biol.*, **24**, 6871–6886.
77. Moran, G., Stokes, C., Thewes, S., Hube, B., Coleman, D.C. and Sullivan, D. (2004) Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence-associated genes in *Candida dubliniensis*. *Microbiology*, **150**, 3363–3382.
78. Kaiser, B., Munder, T., Saluz, H.P., Kunkel, W. and Eck, R. (1999) Identification of a gene encoding the pyruvate decarboxylase gene regulator CaPdc2p from *Candida albicans*. *Yeast*, **15**, 585–591.
79. Linder, T., Rasmussen, N.N., Samuelsen, C.O., Chatzidaki, E., Baraznenok, V., Beve, J., Henriksen, P., Gustafsson, C.M. and Holmberg, S. (2008) Two conserved modules of *Schizosaccharomyces pombe* mediator regulate distinct cellular pathways. *Nucleic Acids Res.*, **29**, 29.
80. Sato, S., Tomomori-Sato, C., Banks, C.A., Parmely, T.J., Sorokina, I., Brower, C.S., Conaway, R.C. and Conaway, J.W. (2003) A mammalian homolog of *Drosophila melanogaster* transcriptional coactivator intersex is a subunit of the mammalian mediator complex. *J. Biol. Chem.*, **278**, 49671–49674.
81. Baidooobonso, S.M., Guidi, B.W. and Myers, L.C. (2007) Med19 (Rox3) regulates intermodule interactions in the *Saccharomyces cerevisiae* mediator complex. *J. Biol. Chem.*, **282**, 5551–5559.
82. Lorch, Y., Beve, J., Gustafsson, C.M., Myers, L.C. and Kornberg, R.D. (2000) Mediator-nucleosome interaction. *Mol. Cell.*, **6**, 197–201.
83. Baldauf, S.L., Roger, A.J., Wenk-Siefert, I. and Doolittle, W.F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, **290**, 972–977.
84. Baldauf, S.L. (2003) The deep roots of eukaryotes. *Science*, **300**, 1703–1706.
85. Palenchar, J.B. and Bellofatto, V. (2006) Gene transcription in trypanosomes. *Mol. Biochem. Parasitol.*, **146**, 135–141.
86. Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.A., Adlem, E., Aert, R. *et al.* (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, **309**, 436–442.
87. Fan, H.Y. and Klein, H.L. (1994) Characterization of mutations that suppress the temperature-sensitive growth of the hpr1 delta mutant of *Saccharomyces cerevisiae*. *Genetics*, **137**, 945–956.
88. Linder, T. and Gustafsson, C.M. (2004) The Soh1/MED31 protein is an ancient component of *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* mediator. *J. Biol. Chem.*, **279**, 49455–49459.
89. Szilagyi, Z., Grallert, A., Nemeth, N. and Sipiczki, M. (2002) The *Schizosaccharomyces pombe* genes sep10 and sep11 encode putative general transcriptional regulators involved in multiple cellular processes. *Mol. Genet. Genomics*, **268**, 553–562.
90. Katinka, M.D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretailade, E., Brottier, P., Wincker, P. *et al.* (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, **414**, 450–453.
91. Slamovits, C.H., Fast, N.M., Law, J.S. and Keeling, P.J. (2004) Genome compaction and stability in microsporidian intracellular parasites. *Curr. Biol.*, **14**, 891–896.
92. Baumli, S., Hoepfner, S. and Cramer, P. (2005) A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer. *J. Biol. Chem.*, **280**, 18171–18178.
93. Soding, J., Biegert, A. and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244–W248.
94. Thompson, C.M., Koleske, A.J., Chao, D.M. and Young, R.A. (1993) A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell*, **73**, 1361–1375.
95. Myers, L.C. and Kornberg, R.D. (2000) Mediator of transcriptional regulation. *Annu. Rev. Biochem.*, **69**, 729–749.
96. Su, X., Kirkman, L.A., Fujioka, H. and Wellem, T.E. (1997) Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant *P. falciparum* in Southeast Asia and Africa. *Cell*, **91**, 593–603.
97. Kato, Y., Habas, R., Katsuyama, Y., Naar, A.M. and He, X. (2002) A component of the ARC/Mediator complex required for TGFbeta/Nodal signalling. *Nature*, **418**, 641–646.
98. Backstrom, S., Elfving, N., Nilsson, R., Wingsle, G. and Bjorklund, S. (2007) Purification of a plant mediator from *Arabidopsis thaliana* identifies PFT1 as the Med25 subunit. *Mol. Cell*, **26**, 717–729.
99. Baldauf, S.L. and Doolittle, W.F. (1997) Origin and evolution of the slime molds (Mycetozoa). *Proc. Natl Acad. Sci. USA*, **94**, 12007–12012.
100. Eichinger, L., Pacheb, J.A., Glockner, G., Rajandream, M.A., Suckang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q. *et al.* (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature*, **435**, 43–57.
101. Carlton, J.M., Hirt, R.P., Silva, J.C., Delcher, A.L., Schatz, M., Zhao, Q., Wortman, J.R., Bidwell, S.L., Alsmark, U.C.M., Besteiro, S. *et al.* (2007) Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*, **315**, 207–212.
102. Quon, D.V., Delgadillo, M.G. and Johnson, P.J. (1996) Transcription in the early diverging eukaryote *Trichomonas vaginalis*: an unusual RNA polymerase II and alpha-amanitin-resistant transcription of protein-coding genes. *J. Mol. Evol.*, **43**, 253–262.
103. Seshadri, V., McArthur, A.G., Sogin, M.L. and Adam, R.D. (2003) *Giardia lamblia* RNA polymerase II: amanitin-resistant transcription. *J. Biol. Chem.*, **278**, 27804–27810.
104. Guo, Z. and Stiller, J.W. (2004) Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. *BMC Genomics*, **5**, 69.
105. Guo, Z. and Stiller, J.W. (2005) Comparative genomics and evolution of proteins associated with RNA polymerase II C-terminal domain. *Mol. Biol. Evol.*, **22**, 2166–2178.
106. Arisue, N., Hasegawa, M. and Hashimoto, T. (2005) Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Mol. Biol. Evol.*, **22**, 409–420.