Origin of multicellular eukaryotes – insights from proteome comparisons

L Aravind* and G Subramanian†

The complete genomes of the yeast *Saccharomyces cerevisiae* and the nematode worm *Caenorhabditis elegans* have recently become available allowing the comparison of the complete protein sets of a unicellular and multicellular eukaryote for the first time. These comparisons reveal some striking trends in terms of expansions or extensive shuffling of specific domains that are involved in regulatory functions and signaling. Similar comparisons with the available sequence data from the plant *Arabidopsis thaliana* produce consistent results. These observations have provided useful insights regarding the origin of multicellular organisms.

Addresses
*Department of Biology, Texas A&M University, College Station, Texas 77843, USA; e-mail: aravind@ncbi.nlm.nih.gov
†Laboratory of Parasitic Diseases, NIAID, NIH, Bethesda, Maryland 20894, USA; e-mail: GSubrama@niaid.nih.gov

Current Opinion in Genetics & Development 1999, **9**:688–694

0959-437X/99/$ – see front matter

Abbreviations
EGF epidermal growth factor
PSI-BLAST position-specific iterated basic local alignment search tool

Introduction

Ever since the discovery of microscopic lifeforms by van Leeuwenhoek in 1675, the tenuous line dividing single-celled and multicellular life forms has been appreciated. Organisms like *Volvox*, slime molds, fungi, choanoflagellates and sponges provide a gradation of organizational levels that suggests a possible evolutionary scenario for the origin of multicellular forms from the unicellular versions. Even within these simple multicellular forms a basic functional differentiation is seen, with some cells playing a reproductive role and others playing a 'structural' or 'nutritional' role. Thus, the issue of explaining the loss of reproductive abilities of certain cells for the sake of others within these multicellular units remains.

In the basic framework of the Darwinian evolutionary paradigm, it appears likely that, in a colony arising from the clonal expansion of a single cell, kin selection could function to result in a differential multicellular form. A set of cells in a colony would differentiated and thereby sacrifice their reproductive ability for that of their clones (kin) if this behavior favors a net increase in fitness. Thus, the fixation of multicellularity would proceed with the selection of genes that favor the increase of fitness of that state relative to the individually reproducing unicellular counterparts.

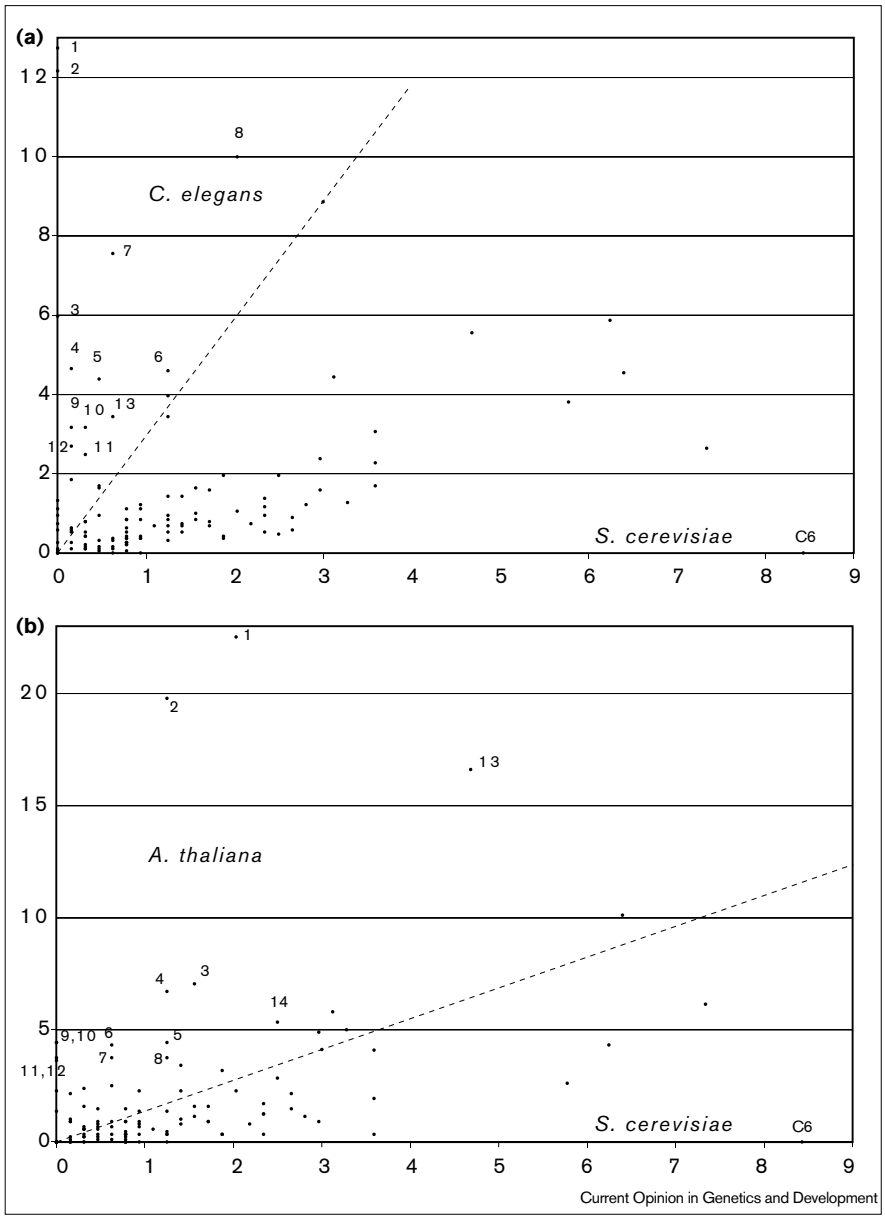
What are the genes that could favor the existence of the multicellular state? The first glimpses of this genetic machinery have emerged in the past year with the first determination of the complete sequence of a multicellular eukaryote — the nematode *Caenorhabditis elegans* [1••]. Concomitantly, considerable advances have been made in terms of sequencing the genome of another multicellular eukaryote — the plant *Arabidopsis thaliana* [2••]. These multicellular eukaryotes develop from a single cell (the zygote) that replicates to give rise to a spatially structured body with a number of differentiated cell types amongst which only a specialized subset participates in reproduction [3]. This can be contrasted with the yeast *Saccharomyces cerevisiae* which shows temporal differentiation in terms of the various gene expression states arising in response to environmental conditions such as presence of nutrients or pheromone and starvation but with very little spatial complexity or specialization [4•]. All these three organisms — the yeast, the animal and the plant — belong to the crown group of eukarya and share a relatively recent common ancestor with respect to the rest of the eukaryotes [5,6]. Thus in the comparison of multicellular genomes to that of yeast, one may expect the molecular basis for multicellularity to stand out over the general background of phylogenetic affinity seen between these organisms. We briefly review here the highlights of the findings obtained from the comparison of the predicted protein sequences of *S. cerevisiae* with that of *C. elegans* and *A. thaliana* and consider them in the context of the evolution of the differentiated state.

Taking snapshots of the worm and yeast proteomes

In silico studies on the predicted proteomes of organisms have been vastly aided by the recent development of sensitive sequence/profile/analysis methods that allow objective detection and statistical evaluation of subtle sequence similarities. Amongst these methods, PSI-BLAST (position-specific iterated basic local alignment search tool) that uses a position specific weight matrix derived from the alignments obtained from the primary gapped BLAST search to iteratively search the database is particularly powerful in terms of sensitivity and speed [7]. The other effective technique implemented in programs like HMMER searches the databases with user-supplied multiple alignments represented as a hidden Markov model [8••]. As a result of such searches, the protein domain families in a proteome can be comprehensively enumerated and trends in their distributions can be characterized.

Regulatory components of cellular processes such as signal transduction, transcription, cell–cell interactions and assembly of multimeric cellular components are most likely targets of selective forces in generating prominent differences between related organisms. Hence, to hone in on changes that accompanied the origin of multicellular

Figure 1



Scatter plots of the number of proteins with a given regulatory domain found in **(a)** *C. elegans* and **(b)** *A. thaliana* with respect to those in yeast. The protein numbers have been divided by the size of proteome in terms of number of proteins. The lines shown in each graph have their slope equal to the ratio of the proteome sizes of the organisms compared. This allows easy visualization of the most prominent domain expansions. A selected set of domains that show prominent proliferation are indicated by numbers next to the points. In **(a)**, the numbers are for the following domains: (1) nuclear receptor finger; (2) C-lectin; (3) EGF; (4) MATH; (5) PTPase; (6) homeodomain; (7) POZ; (8) F-box; (9) PDZ; (10) cation channels; (11) FNIII; (12) SH2; (13) von Willebrand A. In **(b)**, the numbers represent: (1) F-box; (2) LRR; (3) bHLH; (4) Kelch; (5) homeodomain; (6) MADS; (7) POZ; (8) EF-Hands; (9) Viviparous-1; (10) AP-ATPase; (11) AP-2; (12) TIR; (13) Ring; (14) Myb. The C6 shown on the *S. cerevisiae* axis represents the yeast-specific C6 binuclear zinc cluster expansion. The protein kinases have not been shown as their numbers are out of the range of most of the families shown here. Complete data for the domains shown here are available as supplementary material.

states, a comparison between the yeast and nematode proteomes was conducted using domains most frequently found in regulatory proteins. Using databases of such regulatory domains (such as SMART [simple modular architecture research tool] [9••]) profiles were constructed for each of the domain families and the numbers of proteins containing them in each of the two proteomes were determined [10••]. This revealed that there were several substantially larger regulatory domain families in *C. elegans* that were entirely absent in yeast and even in the case of those that were shared, *C. elegans* often possessed larger numbers of proteins with these domains. To account for the ~3-fold increase in the number of proteins encoded by *C. elegans* with respect to *S. cerevisiae*, the above protein counts were normalized by the number of proteins present in each of the proteomes. Even in terms of normalized counts, specific domain families in *C. elegans* show a prominent expansion relative to their numbers in yeast (Figure 1a). An examination of these expanded families provided some immediate clues regarding animal adaptations that include multicellularity.

Lineage-specific domain expansions

The large *C. elegans*-specific domain families that were entirely lacking in yeast included domains mediating extracellular adhesion or those binding DNA and predicted to regulate transcription. These novel extracellular adhesion modules often contain cysteines that form disulfide bonds such as the

Figure 2 legend

A schematic showing the domain organization of proteins involved in small GTPase regulation. Those architectures present only in yeast are in the top panel, those only in *C. elegans* are in the middle panel and those common to both are in the panel at bottom. Only multi-domain proteins are shown and the each domain is indicated by a filled shape as it occurs in the primary sequence. The GAPs are the GTPase-activating proteins for Rho, Ras and Arf whereas the GEFs are the guanine nucleotide exchange factors. FCH, an α-helical domain interaction; LIM, a cysteine-rich interaction domain; PX, PhoX homologous domain; C1/C2, protein kinase C conserved region 1 and 2; PH, plecstrin homology domain; Spec, spectrin domain; RA, Ras-association domain; Ig, immunoglobulin domain; FN3, fibronectin type3 domain; Sec14, Sec14 homologous domain; ANK, ankyrin repeat; CH, calponin homology domain; UBA, ubiquitin-associated domain; OP, octicosapeptide repeat; PLC, phospholipase C; IPPase, inositol phosphatase; Myo, myosin head group; WW, WW domain; cNMP, cyclic nucleotide binding domain; CNH, citron/Nik-1 homologous domain; DEP, Dishevelled, Egl-10, and pleckstrin domain. **(b)** The first bar in this graph is the ratio of the number of the above mentioned GTPase regulatory proteins in *C. elegans* to yeast. The second bar is a similar ratio of the total number of domains found in these proteins as can be identified by sequence analysis. The third represents a similar ratio for the number of domains occurring per proteins while the fourth bar is the *C. elegans*: yeast ratio for the total number of individual domain types found in these proteins. Note that while the total number of domains per protein has only slightly increased in *C. elegans* the number of different domain types found in these proteins has nearly doubled.

epidermal growth factor (EGF) and C-type lectin domains (Figure 1a). The novel DNA-binding domains belong to several distinct families that include the metal chelating forms such as the nuclear hormone receptor family [11•], the T-box [12•] and the paired box [13•]. In the case of families that were shared by these organisms but expanded in *C. elegans*, similar trends were observed with certain transcription regulation related domain families such as the homeodomain and the POZ domain displaying prominent expansions (Figure 1a) [14••,15•]. Further, modules such as the immunoglobulin domain, the von Willebrand A domain (TFIIH subunit p44), the fibronectin III (CHS5) domains and the leucine rich repeat that were found in intracellular contexts in the yeast [16•,17] were expanded in the worm and predominantly present in the extracellular segments of proteins. The domains participating in phosphotyrosine signal transduction such as the tyrosine kinases, SH2 domains, the phosphotyrosine phosphatases and the phosphotyrosine binding domain were strikingly expanded in *C. elegans* [18,19]. Similar prominent expansions are observed in the PDZ domains, cyclic nucleotide cyclases, calcium binding EF hands, cation channels and the F-Box that targets the ubiquitin-mediated degradation pathway [20•,21•,22,23•].

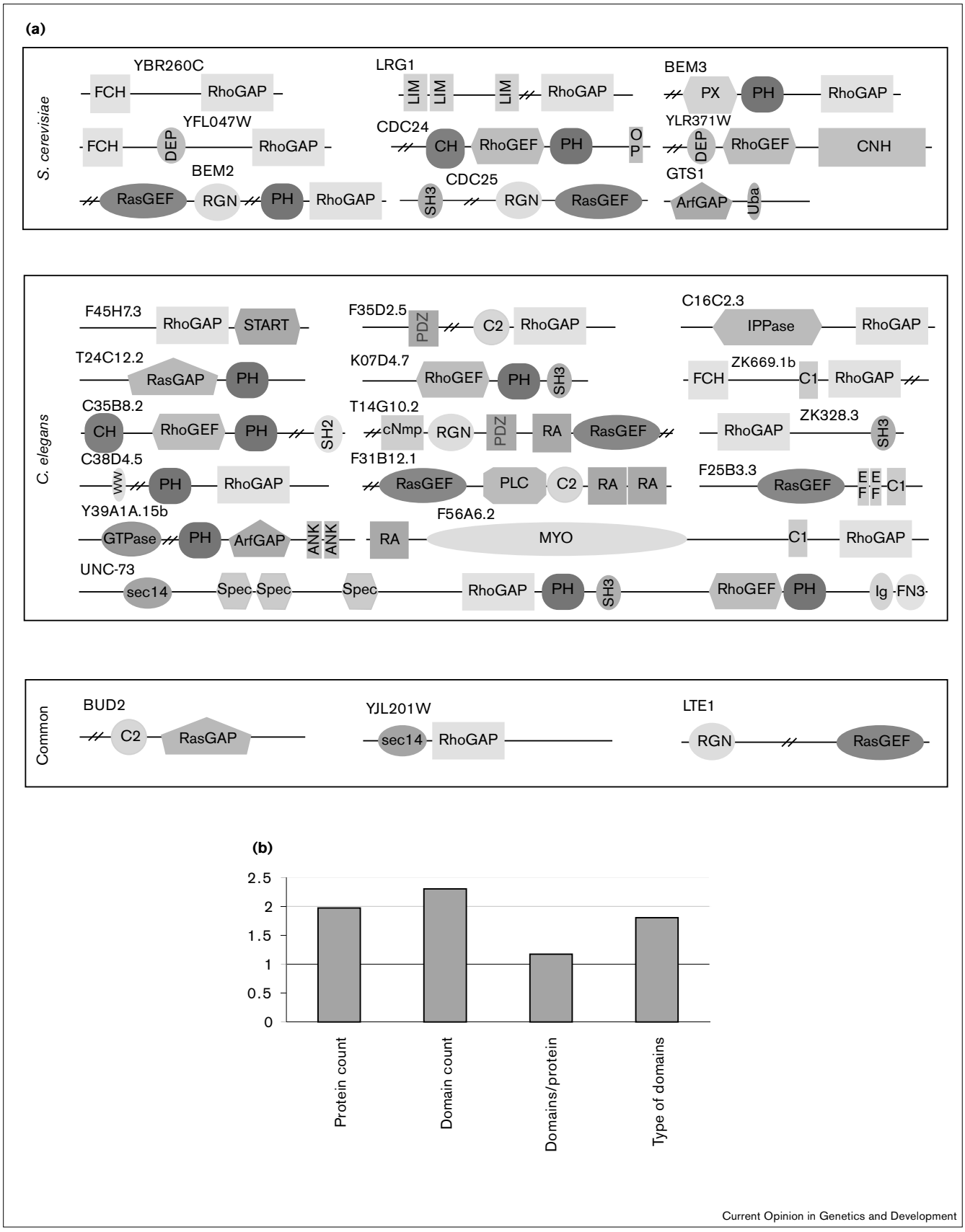
The invention and proliferation of specific transcription factor families in animals correlates well with the increase in drastically different cell types and spatial complexity that is achieved through a developmental process. The increase and recruitment of extracellular-interaction domains not only correlates with the need for adhesion in multicellular organisms but also with the need for specific intercellular contacts required for tissue development. Related to this intercellular communication is the rise of intracellular signaling domains that facilitate response to external stimuli. Of particular interest in this group is the proliferation of the PDZ domains that bind to carboxy-terminal peptides of membrane proteins and thereby act as adaptors for intracellular cascades (Figure 1a) [21•]. Thus, it appears that the rise of animal multicellularity can be correlated to the selection of protein families that allowed differentiation (the transcription factors) and cellular communication (the signaling and adhesion specific domains). In contrast to these expansions, the only one unique to yeast is that of the C6 zinc cluster DNA binding domain. Several members of this family regulate metabolic processes such as galactose utilization, leucine synthesis and the like [24] suggesting that it is an adaptation consistent with the temporal spectrum of cellular states shown by the yeast.

Trends in domains with lower relative numerical differences between the proteomes

There are other families the normalized counts of which show a comparable value in the nematode and yeast. A careful analysis of these families shows that, even here, diversifying trends are noticeable in the domain architectures of these proteins. To clarify this point, the regulators of small GTPase signaling namely the GAPs (GTPase activating proteins) and GEFs (guanine nucleotide exchange factors) for Rho, Ras and Arf family of proteins were investigated [25•,26,27] (Figure 2a). The absolute numbers of these proteins and the domains found in them showed an expansion in *C. elegans*, whereas the total number of domains found per protein was comparable in the two organisms (Figure 2b). The number of different types of domains found in these proteins, however, showed a two-fold increase in *C. elegans* with very few domain architectures being conserved between the two taxa (Figure 2). This suggests that there have been several novel domain juxtapositions during the origin of multicellular animals in addition to the duplications, probably corresponding to the increased intracellular signaling. The C_2H_2 finger containing proteins are present in large numbers in both proteomes [28•] and show comparable normalized counts. A careful examination of their domain architectures and sequence similarity reveals that almost none of them are common to the two proteomes with a greater number of finger domains per protein seen in *C. elegans*, suggesting that binding to larger regulatory elements in the multicellular genomes selected for the independent expansion of multidomain C_2H_2 fingers in animals.

The process of differentiation in animals involves asymmetric division and lateral inhibition to designate certain

Figure 2



Current Opinion in Genetics and Development

members of a cellular field for a given differentiation state [29•]. A key signaling pathway that participates in this process is the Notch pathway that includes the receptor protein Notch, its ligands and modifier, as well as intracellular proteins such as the transcription factor Su(H)/CBF (suppressor of hairless/C promoter binding factor) that transmits the signal to the nucleus [30•]. It is seen that whereas the receptor Notch and its ligand may be unique to animals, the extracellular modifier (the glycosyltransferase Fringe [31]) and the intracellular components (DNA-binding domain Su[H] and chromatinic ATPase Strawberry Notch) are most likely ancient components as their homologs can be found in plants and *Schizosaccharomyces* (L Aravind, unpublished data). This suggests that the pathway did not arise all at once with the rise of animal multicellularity but rather by recruiting pre-existing proteins and adding a few new inventions. Similarly, recruitment followed by divergence could have given rise to the novel domain inventions in animals such as the Paired domain HTHs (helix turn helices) from tranposase HTH domains [32], the Hedgehog like proteins from selfish elements like inteins [33] and the NHR (nuclear hormone receptor) zinc finger from the Lim domains [34].

Opposed to the prominent differences in the regulatory proteins is the relative conservatism of domains found in certain general chromatinic factors such as the SWI/SNF ATPases and the cell cycle regulatory proteins [35]. Nevertheless, detailed investigation of these systems reveals some rather interesting distinctions between unicellular fungi and animals. The structural chromosomal proteins that are involved in positive and negative regulation of chromatin structure show a detectable increase in diversity of domain architectures and domain numbers per protein in animals. Selection for the establishment of multiple distinct chromatin states in the multicellular organisms would have resulted in this diversification of the structural components while maintaining a relatively constant set of catalytic activities in the form of the ATPases. The core cell cycle components such as the MCMs (minicromosome maintenance proteins) and the CDKs (cyclin-dependent kinases) are universally found in the eukaryotes but the regulatory components such as the retinoblastoma protein, E2F, the KIP/WAF-type CDK inhibitor, Cyclin D, BRCA1 and BRCA2 are only found in plants and animals to the exclusion of the yeast (L Aravind, unpublished data). These proteins may have been selected for in multicellular eukaryotes specifically to regulate proliferation as has been suggested by studies on these proteins in animals [36•,37•].

The story of the weed – *Arabidopsis thaliana*

These comparisons of the complete proteomes of *C. elegans* and *S. cerevisiae* suggest that specific domain families may have undergone expansion and combination with other domains in response to the forces selecting for the maintenance of multicellularity. The availability of significant amounts of *A. thaliana* sequence data [2••] allows a similar independent comparison between a distinct multicellular organism and yeast to test if there are any general trends in the evolution of multicellularity. The results of such a comparison are very consistent with the *C. elegans*/*S. cerevisiae* yeast comparison (Figure 1b). The plants show the invention and expansion of two novel transcription factor families, AP2 (Apetala-2 domain) and viviparous1 [38,39], and also a proliferation of shared transcription factors such as MYB, bHLH and the MADS domain (Figure 1b).

The plants also show the recruitment of distinct extracellular adhesion modules such as the bulb lectin and an EGF-like cysteine rich module and LRR just as in animals. In terms of increases in the counts of intracellular signaling domains, the protein kinases, PP2C phosphatases, EF hands and the F-boxes figure prominently. Interestingly, there is no expansion of the PDZ domain suggesting that this feature may be linked to the animal-specific development of neural tissue and associated proliferation of the cation channels. The plants also draw extensively from their cyanobacterial inheritance to develop bacterial type signaling systems in the form of multiple two-component system proteins — kinases and receiver domain proteins [40]. Although similarities in the expansion trends are observed between animals and plants, even the common domain expansions such as that of the homeodomain, the F-box, LRR and the protein kinases appear to have occurred independently. In all these cases, analogous rather than orthologous domain organizations are seen, for example animal receptor tyrosine kinases compared with the plant receptor kinases with LRR and bulb lectin domains. This provides strong evidence that evolution has worked in the crown group of eukaryotes to produce rather similar solutions for multicellularity on independent occasions.

Conclusion

These consistent results for *C. elegans* and *A. thaliana* allow the reconstruction of the possible evolutionary events that lead to multicellularity. The majority of the regulatory domains were present in the common ancestor of the crown group but the general lack of conservation of domain architectures suggests that they were independently recruited and expanded along the different lineages. This is particularly striking in the case of the transcription factors where there have been additional specific domains in each lineage. Further, it is possible that specific transcriptional regulation developed under rather distinct selective forces in different eukaryotic lineages: in yeast for regulation of temporal states and in animals and plants for spatial and developmental diversification of cell states.

In reconstructing the origin of multicellularity, one can imagine a ground state in the form of a colony of identical cells. Certain changes that restricted the expression of ancestral transcription factors to distinct subsets of cells in this colony could have triggered off spatial differentiation. Selection of organisms with such systems could have led to the fixation of interacting protein networks and led to the origin of forms with an increased number of cell types.

Such a process would have provided a selective opportunity for the expansion of transcription factors (for differentiation), signaling molecules (to respond to intercelluar communication) and adhesion molecules (to hold the cells together). Thus, we see these molecules as prominent components of the proteomes of complex multicellular eukaryotes such as *C. elegans* and *A. thaliana*. Although the issue of primary derivation of these novel regulatory functions like adhesion and intercellular signaling could be raised, as mentioned above, recruitment of pre- existing domains appears to have played the a major part. Those domains that are entirely new are largely based on stabilizing disulfide bonds or chelated metal ions, suggesting that they could have arisen relatively easily *de novo*. Thus, complex as multicellular eukaryotes may appear, a set of rather simple steps could have lead to their origins.

The *Drosophila* genome is expected to be fully sequenced by the end of this year, with those of *Homo sapiens* and *Dictyostelium discoideum* to follow. Analysis of these genomes are likely to provide exciting insights regarding the relative evolutionary time in which the regulators of multicellular development and patterning emerged.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. *C. elegans* Genome Sequencing Consortium: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282**:2012-2018.

•• The primary publication of the Genome Sequencing Consortium describing the approach to obtaining the complete sequence and associated caveats. It also discusses the strategies used to discover how genes annotate the large proteome and makes some observations regarding the striking features that can be inferred from the proteome.

2. Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M: ***Arabidopsis thaliana*: a model plant for genome analysis.** *Science* 1998, **282**:662,679-682.

•• A report on the progress of the *Arabidopsis* genome sequencing projects with respect to generation of primary sequence, its assembly and annotation.

3. Gerhart J, Kirschner MW: **Cells, embryos, and evolution: toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability.** *Blackwell Science Inc* 1997.

4. Dickinson JR, Schweizer ME: *The Metabolism and Molecular Physiology of Saccharomyces cerevisiae*. London: Taylor & Francis; 1998.

• A recent book providing details regarding the physiological states of the yeast *Saccharomyces cerevisiae*.

5. Doolittle RF, Feng DF, Tsang S, Cho G, Little E: **Determining divergence times of the major kingdoms of living organisms with a protein clock [see comments].** *Science* 1996, **271**:470-477.

6. Sogin ML, Morrison HG, Hinkle G, Silberman JD: **Ancestral relationships of the major eukaryotic lineages.** *Microbiologia* 1996, **12**:17-28.

7. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.

8. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14**:755-763.

•• Provides a very useful review of the use of HMM in biological sequence analysis. The methods described here are implemented in the 'easy to use' HMMER program that can be downloaded from the World Wide Web at: http://hmmer.wustl.edu/

9. Ponting CP, Schultz J, Milpetz F, Bork P: **SMART: identification and annotation of domains from signalling and extracellular protein sequences.** *Nucleic Acids Res* 1999, **27**:229-232.

•• Excellent system for the annotation of signaling domains in proteins based on HMMs derived from high-quality alignments. Can be accessed on the World Wide Web at: http://coot.embl-heidelberg.de/SMART/

10. Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T *et al.*: **Comparison of the complete protein sets of worm and yeast: orthology and divergence.** *Science* 1998, **282**:2022-2028.

•• The first comparison of the complete proteomes of two eukaryotes. The authors describe the lineage-specific domain expansions, conservation of the core functions and the use of PSI-BLAST sequence profiles in establishing counts for proteins containing particular domains.

11. Sluder AE, Mathews SW, Hough D, Yin VP, Maina CV: **The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes.** *Genome Res* 1999, **9**:103-120.

• Detailed analysis of the largest transcription factor family – the nuclear hormone receptors in the *C. elegans* proteome. The analysis identifies the ancient families as well as the nematode-specific expansion that has occurred within this superfamily.

12. Papaioannou VE, Silver LM: **The T-box gene family.** *Bioessays* 1998, **20**:9-19.

• Describes the diversity of the T-Box transcription factors in the animals.

13. Hobert O, Ruvkun G: **Pax genes in *Caenorhabditis elegans*: a new twist.** *Trends Genet* 1999, **15**:214-216.

• Describes the interesting case of the alternative domain organizations of two HTH domains comprising the Paired Box. The carboxy- and amino-terminal HTH domains have been shown to occur independently of each other as well as in the form of fusions as in the classic Paired domains.

14. Ruvkun G, Hobert O: **The taxonomy of developmental control in *Caenorhabditis elegans.*** *Science* 1998, **282**:2033-2041.

•• Provides a detailed summary of the conservation of developmental pathways in the nematode. Some of the conclusions reported here need to be reconsidered due to use of an outdated sequence set.

15. Aravind L, Koonin EV: **Fold prediction and evolutionary analysis of the POZ domain: structural and evolutionary relationship with the potassium channel tetramerization domain.** *J Mol Biol* 1999, **285**:1353-1361.

• This paper describes the provenance of the 'eukaryotic' signaling domains. The intracellular occurrence of von Willebrand factor A domains in the TFIIH p44 subunit is reported for the first time and enumeration of these domains in complete genomes is provided.

16. Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV: **Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer.** *J Mol Biol* 1999, **289**:729-745.

• Describes the intracellular occurrence of von Willebrand factor A domains in the TFIIH subunit p44 subunit.

17. Bateman A, Chothia C: **Fibronectin type III domains in yeast detected by a hidden Markov model.** *Curr Biol* 1996, **6**:1544-1547.

18. Pawson T, Scott JD: **Signaling through scaffold, anchoring, and adaptor proteins.** *Science* 1997, **278**:2075-2080.

19. Shoelson SE: **SH2 and PTB domain interactions in tyrosine kinase signal transduction.** *Curr Opin Chem Biol* 1997, **1**:227-234.

20. Hurley JH: **The adenylyl and guanylyl cyclase superfamily.** *Curr Opin Struct Biol* 1998, **8**:770-777.

• Provides an introduction to the recent developments regarding the cyclic nucleotide generating enzymes.

21. Ranganathan R, Ross EM: **PDZ domain proteins: scaffolds for signaling complexes.** *Curr Biol* 1997, **7**:R770-R773.

• Reviews the role of PDZ domains in acting as the sub-membrane scaffolds in animal signaling systems.

22. Ikura M: **Calcium binding and conformational response in EF-hand proteins.** *Trends Biochem Sci* 1996, **21**:14-17.

23. Patton EE, Willems AR, Tyers M: **Combinatorial control in ubiquitin dependent proteolysis: don't Skp the F-box hypothesis.** *Trends Genet* 1998, **14**:236-243.

• Provides a good introduction to the significance of the F-Box in ubiquitin-dependent signaling. The role of the F-box modules in targeting the ubiquitin-conjugating system is described.

24. Schjerling P, Holmberg S: **Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators.** *Nucleic Acids Res* 1996, **24**:4599-4607.

25. Goldberg J: **Structural and functional analysis of the ARF1**
• **ARFGAP complex reveals a role for coatomer in GTP hydrolysis.** *Cell* 1999, **96**:893-902.
A primary paper describing the molecular mechanism of the ARF-like GTPase-activating proteins. These GAPs have a structure similar to the GATA zinc fingers and unlike other GAPs contribute no residues to the GTPase active site.

26. McCormick F: **Activators and effectors of ras p21 proteins.** *Curr Opin Genet Dev* 1994, **4**:71-76.

27. Van Aelst L: **Rho GTPases and signaling networks.** *Genes Dev* 1997, **11**:2295-2322.

28. Clarke ND, Berg JM: **Zinc fingers in *Caenorhabditis elegans*:**
• **finding families and probing pathways.** *Science* 1998, **282**:2018-2022.
Describes the elaboration of different zinc chelating domains in *C. elegans*.

29. Hawkins N, Garriga G: **Asymmetric cell division: from A to Z.** *Genes*
• *Dev* 1998, **12**:3625-3638.
Reviews the role of asymmetric divisions in patterning and differentiation.

30. Artavanis-Tsakonas S, Rand MD, Lake RJ: **Notch signaling: cell**
• **fate control and signal integration in development.** *Science* 1999, **284**:770-776.
An update review of the animal-specific Notch signaling pathway describing the role of the recently uncovered components of this pathway.

31. Yuan YP, Schultz J, Mlodzik M, Bork P: **Secreted fringe-like signaling molecules may be glycosyltransferases.** *Cell* 1997, **88**:9-11.

32. Franz G, Loukeris TG, Dialektaki G, Thompson CR, Savakis C: **Mobile Minos elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain.** *Proc Natl Acad Sci USA* 1994, **91**:4746-4750.

33. Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ: **Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins.** *Cell* 1997, **91**:85-97.

34. Perez-Alvarado GC, Miles C, Michelsen JW, Louis HA, Winge DR, Beckerle MC, Summers MF: **Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP.** *Nat Struct Biol* 1994, **1**:388-398.

35. Eisen JA, Sweder KS, Hanawalt PC: **Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions.** *Nucleic Acids Res* 1995, **23**:2715-2723.

36. Irminger-Finger I, Siegel BD, Leung WC: **The functions of breast**
• **cancer susceptibility gene 1 (*BRCA1*) product and its associated proteins.** *Biol Chem* 1999, **380**:117-128.
A summing up of the present understanding of the role the BRCA proteins in the regulation of cellular proliferation.

37. Macleod K: **pRb and E2f-1 in mouse development and**
• **tumorigenesis.** *Curr Opin Genet Dev* 1999, **9**:31-39.
A discussion of the role of the E2F/Rb system in regulating cell proliferation in animals.

38. Suzuki M, Kao CY, McCarty DR: **The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity.** *Plant Cell* 1997, **9**:799-807.

39. Riechmann JL, Meyerowitz EM: **The AP2/EREBP family of plant transcription factors.** *Biol Chem* 1998, **379**:633-646.

40. Yeh KC, Lagarias JC: **Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry.** *Proc Natl Acad Sci USA* 1998, **95**:13976-13981.