

- 5 Riedel, C.G. *et al.* (2006) Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature* 441, 53–61
- 6 Kitajima, T.S. *et al.* (2006) Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* 441, 46–52
- 7 Tang, Z. *et al.* (2006) PP2A is required for centromeric localization of Sgo1 and proper chromosome segregation. *Dev. Cell* 10, 575–585
- 8 Mailhes, J.B. *et al.* (2003) Okadaic acid, an inhibitor of protein phosphatase 1 and 2A, induces premature separation of sister chromatids during meiosis I and aneuploidy in mouse oocytes *in vitro*. *Chromosome Res.* 11, 619–631
- 9 Gregan, J. *et al.* (2007) What makes centromeric cohesion resistant to separate cleavage during meiosis I but not during meiosis II? *Cell Cycle* 7, 151–153
- 10 McGuinness, B.E. *et al.* (2005) Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. *PLoS Biol.* 3, e86
- 11 Huang, H. *et al.* (2007) Tripin/hSgo2 recruits MCAK to the inner centromere to correct defective kinetochore attachments. *J. Cell Biol.* 177, 413–424
- 12 Kouznetsova, A. *et al.* (2007) Bi-orientation of achiasmatic chromosomes in meiosis I oocytes contributes to aneuploidy in mice. *Nat. Genet.* 39, 966–968
- 13 Hauf, S. *et al.* (2007) Aurora controls sister kinetochore mono-orientation and homolog bi-orientation in meiosis-I. *EMBO J.* 26, 4475–4486
- 14 Petronczki, M. *et al.* (2006) Monopolar attachment of sister kinetochores at meiosis I requires casein kinase 1. *Cell* 126, 1049–1064
- 15 Kitajima, T.S. *et al.* (2004) The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis. *Nature* 427, 510–517

0168-9525/\$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tig.2008.02.001 Available online 2 April 2008

Genome Analysis

Conservation of divergent transcription in fungi

Philip R. Kensche, Martin Oti, Bas E. Dutilh and Martijn A. Huynen

Center for Molecular and Biomolecular Informatics and Nijmegen Center for Molecular Life Sciences, Radboud University Medical Center. 6500 HB Nijmegen, The Netherlands

The comparison of fully sequenced genomes enables the study of selective constraints that determine genome organisation. We show that, in fungi, adjacent divergently transcribed ($\leftarrow\rightarrow$) genes are more conserved in orientation than convergent ($\rightarrow\leftarrow$) or co-oriented ($\rightarrow\rightarrow$) gene pairs. Furthermore, the time divergent orientation of two genes is conserved correlates with the degree of their co-expression and with the likelihood of them being functionally related. The functional interactions of the proteins encoded by the conserved divergent gene pairs indicate a potential for protein function prediction in eukaryotes.

Background

In prokaryotes, conservation of co-oriented ($\rightarrow\rightarrow$) and divergent gene pairs ($\leftarrow\rightarrow$) correlates with the presence of operons and bi-directional promoters, respectively, and is used to predict functional links between genes [1,2]. In eukaryotes, operons are a common genomic feature in only a few clades, such as Nematodes [3] and Urochordates [4]. By contrast, bi-directional promoters are a universal mechanism for co-regulation: Adjacent genes organised in divergent orientation and separated by short intergenic spacers tend to be co-regulated in flies [5] and vertebrates [6]. Furthermore, in vertebrates, the distance between divergently transcribed genes tends to be conserved in different species [6,7] and divergent gene pairs are overrepresented among gene pairs that are conserved in orientation [7]. Among gene pairs whose orientations are conserved in the two fungi *Saccharomyces cerevisiae* and *Candida albicans*, divergent gene pairs are enriched only if they are highly co-expressed [8,9]. However, divergent gene pairs did not

retain their orientation with an especially high rate [8]. We investigated the relationship between conservation of gene orientation, the class of orientation and intergenic distances. Furthermore, we asked whether the conservation of divergent gene orientation can be used for protein function prediction in eukaryotes. We based our study on a set of 19 fungi with a phylogenetic tree whose branches sum up to a total of 3850 million years (My; Supplementary Figure S1).

Divergent gene orientation is highly conserved

We calculated the fraction of adjacent gene pairs that are conserved in orientation between every possible pair of species and plotted it against the species' divergence times (Figure 1a; see Supplementary Material 1). At large evolutionary distances of 400 My and more, for instance, between Euscomycota and Hemiascomycota, divergent gene pairs are more conserved in orientation than convergent ($\rightarrow\leftarrow$) or co-oriented gene pairs. In addition, the more recently separated species among both the Euscomycota and Basidiomycota show a pronounced conservation of divergent orientation. We further quantified the conservation of orientation of individual orthologous group pairs by how long their orientation has been maintained in fungal evolution. To this end, for each pair of orthologous groups, we inferred with a Dollo parsimony model [10] where in fungal evolution they gained and lost their orientation (Figure 1b). A comparison of the resulting conservation times shows that divergent gene pairs are on average (geometric means) maintained for 61 My longer than co-oriented gene pairs (one-sided Mann-Whitney test, $P < 1.2 \times 10^{-47}$) and 32 My longer than convergent gene pairs ($P < 2.5 \times 10^{-13}$). The enrichment of divergent gene pairs over other orientations increases with the conservation time. If genes are located on either strand

Corresponding author: Huynen, M.A. (M.Huynen@cmbi.kun.nl).

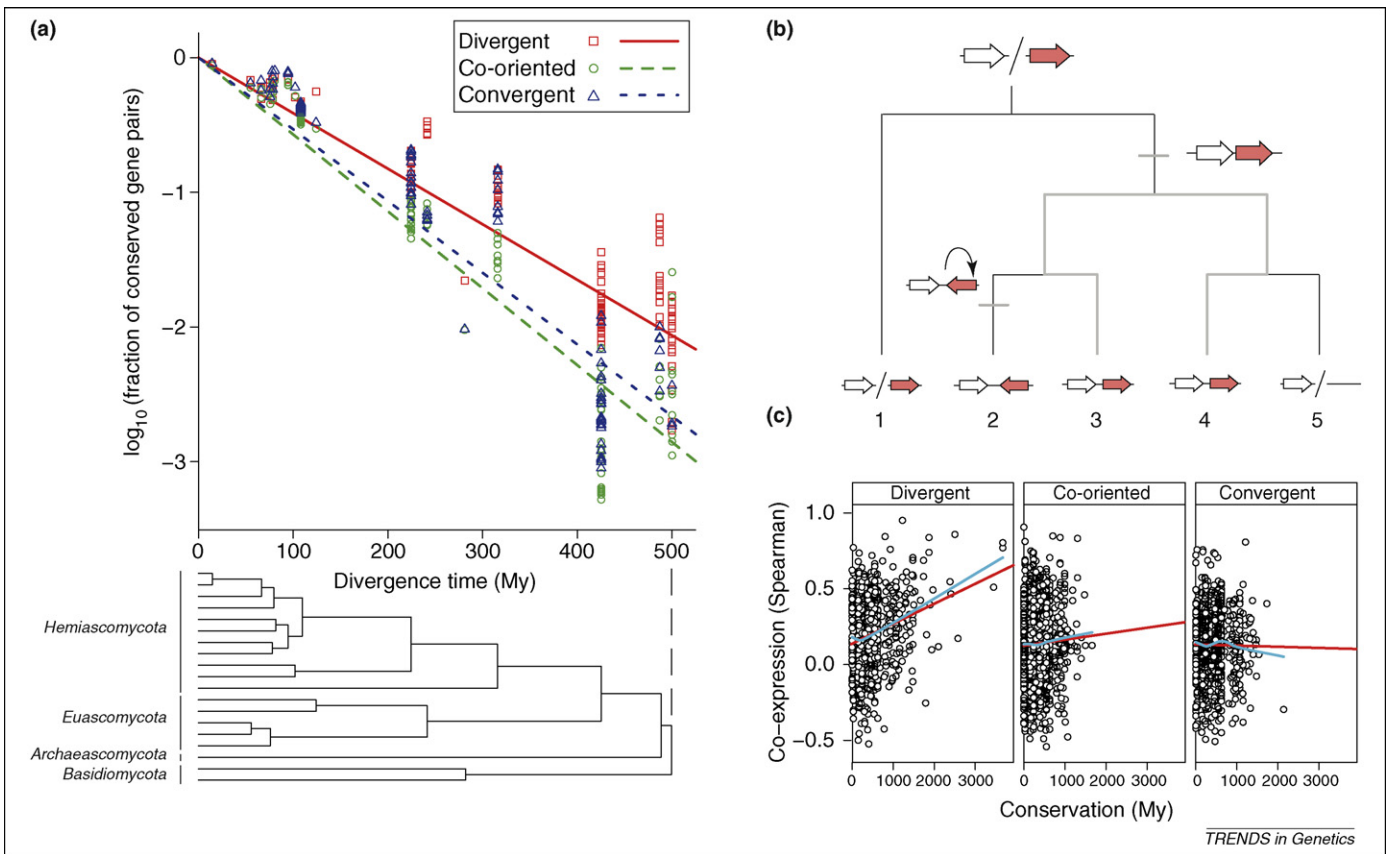


Figure 1. (a) Fractions of divergent, convergent and co-oriented gene pairs conserved between each pair of species plotted against the species' divergence times. The divergence times were estimated based on the assumption that Basidiomycota and Ascomycota diverged 500 million years ago [20]. For details on methods see Supplementary File 1 online. The average rates with which the fractions of conserved pairs decay are as follows; divergent: $r = 4.12 \times 10^{-3}$; convergent: $r = 5.32 \times 10^{-3}$; co-oriented: $r = 5.71 \times 10^{-3}$; adjacent (data not shown): $r = 4.63 \times 10^{-3}$. (b) Conservation of orientation of orthologous group pairs. Members of two orthologous groups (white and gray) are co-oriented in species 3 and 4. In other species, the two genes are not adjacent (1), not co-oriented (2) or missing (5). Dollo parsimony [10] generally assumes that a trait was gained only once in the evolution of an organism group, namely in the most recent common ancestor (MRCA) of all species that have this trait. In this example, it thus infers a single gain of orientation, directly before the MRCA of species 3 and 4, and two subsequent losses. The conservation time for this co-oriented orthologous group pair is calculated as the sum of the lengths of the four branches highlighted in gray. (c) Correlation of co-expression in *Saccharomyces cerevisiae* and conservation. Red: linear model fitted with least squares (divergent: Pearson $r = 0.25$, $P = 2.4 \times 10^{-15}$; co-oriented: $r = 0.046$, $P = 0.047$; convergent: $r = -0.011$, $P = 0.71$). Blue: local regression with loess [21].

of the DNA with the same probability, for structural reasons, we expect to find 50% co-oriented, 25% convergent and 25% divergent gene pairs. Indeed, divergent pairs represent only $\sim 30\%$ of the total set of conserved orthologous group pairs (3346 divergent, 3283 convergent, 5131 co-oriented). Nevertheless, above a conservation cut-off of 1000 My, divergent pairs (169) are about two or three times more frequent than convergent (95) and co-oriented pairs (56) pairs, respectively. Finally, if we consider the top 25 conserved divergent gene pairs, we find only one (3.5%) convergent and three co-oriented (10.4%) gene pairs that have been conserved over a similar evolutionary distance.

The neutral evolutionary model

The differences in the conservation of different gene orientations could be explained by a neutral evolutionary model in which the probability for a linkage break between two genes is proportional to the distance between the genes on the chromosome [8,11]. Given the conservation of divergent gene pairs, this model predicts that genes with divergent orientation should be separated by shorter intergenic spacers than convergent or co-oriented genes. However, divergently oriented transcripts in *S. cerevisiae* are

separated by longer spacers (median 401 bp, one-sided Mann-Whitney test, both with and without controlling for conservation, $P < 1.2 \times 10^{-49}$; Supplementary Figure S4) than co-oriented (median 180 bp) or convergent (median -84 bp) transcripts. This result applies to all 19 fungi if the distances between open reading frames (ORFs) are used: spacers between divergent genes were always longer than those between convergent genes ($P < 2.4 \times 10^{-39}$) and, for the majority of species (12/19), longer than those between co-oriented genes ($P < 0.05$). Nevertheless, among gene pairs with the same orientation, the conservation negatively correlates with the length of the spacer between transcripts (divergent $\tau = -0.2$, convergent $\tau = -0.13$, co-oriented $\tau = -0.11$, $P < 1.3 \times 10^{-9}$), which is consistent with previous findings in Hemiascomycota [11] and with the neutral evolutionary model. These seemingly contradictory results can be explained by a refined neutral model: Divergent gene pairs may be more conserved in orientation because the length of the region in which a selectively neutral disruption can happen is shorter than the intergenic region itself. For instance, one would expect that removing an important transcription factor binding site from its core promoter and gene by a rearrangement has a negative fitness effect.

Conservation of divergent orientation correlates with co-expression

The refined neutral model still leaves us with the question why the regions in which rearrangements are selectively

neutral should be shorter between divergent gene pairs than between convergent or co-oriented pairs. One explanation could be that for some divergent gene pairs the region is reduced to zero length by mechanisms of

Table 1. Top 25 conserved divergent orthologous group pairs

Conservation (My) ^a	ORF 1 ^{b,c}	Function 2	ORF 2	Function 2	Relation ^d
3672 (95.1%)	YDR224c YBL002w (FOG_2510)	Histone H2B	YDR225w YBL003c (FOG_2543)	Histone H2A	Nucleosome
3446 (89.5%)	YBR009c YNL030w (FOG_2256)	Histone H4	YBR010w YNL031c (FOG_2521)	Histone H3	Nucleosome
2577 (66.9%)	YLR410w (FOG_232)	Physically interacts with rRNA processing proteins	YLR409c (FOG_3415)	Involved in maturation of pre-18S rRNA	Ribosome assembly
2504 (65.0%)	YOL039w (FOG_1240)	Ribosomal protein P2 α	YOL040c (FOG_657)	Component of the small (40S) ribosomal subunit	Ribosome
2490 (64.7%)	(FOG_1877) YBR207w YER145c)	High-affinity iron transporter	(FOG_2004) YFL041w YMR058w)	Multi-copper oxidase	Iron import [22,23]
2406 (62.5%)	YHR066w (FOG_1786)	Constituent of 66S pre-ribosomal particles	YHR065c (FOG_637)	Maturation of the 35S rRNA	Ribosome assembly
2387 (62.0%)	YIL020c (FOG_2727)	Histidine biosynthesis	YIL019w (FOG_897)	Pre-rRNA processing; ribosome assembly	Unexplained
2294 (59.6%)	(FOG_832)	Mlo2p (<i>S. pombe</i>); involved in chromosome segregation [15]	(FOG_946) YBL023c)	Mcm2p; implicated in DNA synthesis [14]	DNA-replication, chromosome segregation
2247 (58.4%)	(FOG_3048) YEL051w)	Vacuolar ATPase	(FOG_3753) YDR416w)	Involved in pre-mRNA splicing	Unexplained
2101 (53.6%)	YPR191w (FOG_3052)	Subunit of the ubiquinol cytochrome-c reductase complex	YPR190c (FOG_670)	RNA polymerase III subunit C82	Unexplained
2023 (52.6%)	(FOG_2704)	MFS sugar transporter, xylose transporter	(FOG_4642)	α -L-arabinofuranosidase/ β xylosidase	Xylose metabolism
1971 (51.7%)	YOR160w (FOG_1988)	Npl3p-dependent mRNA export from the nucleus	YOR159c (FOG_2119)	Involved in spliceosomal mRNA splicing	mRNA processing
1940 (50.4%)	YPR187w (FOG_3400)	RNA polymerase subunit ABC23	YPR186c (FOG_4075)	Transcription factor IIIA (TFIIIA)	RNA polymerase
1921 (49.4%)	(FOG_1391) YLL057c)	Involved in obtaining SO ₃ from taurine [12]	(FOG_5857) YIL166c)	MFS, upregulated by sulfur depletion [13]	Sulfur homeostasis
1900 (49.3%)	YBR020w (FOG_3740)	Galactokinase	YBR019c (FOG_3946)	UDP galactose-4-epimerase	Galactose metabolism
1895 (49.2%)	(FOG_1156) YPL078c)	Mitochondrial ATPase subunit ATP4	(FOG_638) YPR144c)	Small ribosomal subunits biogenesis and assembly	Unexplained
1871 (48.6%)	YMR142c YDL082w (FOG_1745)	Component of the large ribosomal subunit	YMR143w YDL083c (FOG_4503)	Component of the small ribosomal subunit	Ribosome
1863 (48.4%)	YDR448w (FOG_2360)	Transcription coactivator	YDR447c (FOG_399)	Ribosomal protein 51 of the small subunit	Gene expression
1849 (48.0%)	(FOG_2177) YHR172w)	γ -tubulin complex subunit Alp4	(FOG_4086) YGR253c)	Proteasome component Pup2	Cell division
1820 (47.3%)	(FOG_2739) YGL016w)	Karyopherin beta; nuclear import	(FOG_427, YLL029c)	Mutant defective in spore formation	Hypothetical
1810 (47.0%)	YBR171w (FOG_1245)	Member of Sec63 complex; protein targeting and import into the endoplasmic reticulum (ER)	YBR170c (FOG_5151)	Npl4p; ER and nuclear membrane; genetic interaction with Sec63 ([18,24], PPI ID: 1107)	ER import and export
1787 (46.4%)	YGR289c (FOG_2439)	Maltose permease	YGR292w (FOG_824)	Maltase	Maltose catabolism
1787 (46.4%)	YER072w (FOG_2806)	Vacuolar transporter chaperon (VTC)	YER071c (FOG_511)	Protein of unknown function	Hypothetical
1777 (46.2%)	(FOG_1072) YBR191w YPL079w)	60S ribosomal protein L21-A	(FOG_2007) YBR189w YPL081w)	40S ribosomal protein S9	Ribosome
1777 (46.2%)	(FOG_1954) YDL030w)	Spliceosome assembly	(FOG_600) YDR021w YKR059w YJL138c)	RNA helicases; nucleolar maturation of rRNAs	Unexplained

See full table in Supplementary Material online.

^aPercentages are conservation scores relative to total sum of branch lengths in the tree (3850 My).

^bOrthologs in *S. cerevisiae* that reside in the conserved orientation. Otherwise orthologous group identifiers (see Supplementary Material online) and *Saccharomyces cerevisiae* orthologs are given in parentheses.

^cORF, open reading frame.

^dType of interaction, common process or physical complex. 'Hypothetical' relations involve proteins of unknown function.

co-regulation, such as bi-directional promoters. Indeed, divergent gene pairs in budding yeast have higher co-expression values (median = 0.201) than convergent (median = 0.140) and co-oriented pairs (median = 0.143; one-sided Mann-Whitney test, $P < 3.5 \times 10^{-10}$). Furthermore, among divergently oriented gene pairs, conservation correlates with co-expression, explaining 6.1% of its variation (Figure 1c; Kendall's $\tau = 0.10$; $P = 6.0 \times 10^{-4}$). This correlation is not caused by confounding effects of the length of the intergenic spacer on both conservation and co-expression because it remains significant when controlling for distance between transcripts (partial Kendall's $\tau = 0.08$ $P = 4.5 \times 10^{-3}$). By contrast, for co-oriented and convergent gene pairs, co-expression and conservation do not correlate ($\tau < 0.028$, $P > 0.14$).

Conservation of gene orientation for protein function prediction

Of the 25 most conserved divergent gene pairs, a significant proportion are obviously functionally related (Table 1). Like in prokaryotes, conserved divergently oriented gene pairs show a wide array of functional interactions, including shared complex membership and involvement in the same metabolic pathway. In some cases, the gene order conservation can aid in predicting a specific protein function. An interesting case is a gene pair that is conserved over ~50% of the total phylogenetic distance in the tree and that involves the *S. cerevisiae* genes *JLP1* (YLL057c) and *YIL166c*. *Jlp1p* is a dioxygenase that releases sulfite from sulfonates, such as taurine [12]. *Yil166cp* is a member of the Major Facilitator Superfamily (MFS) that has no ortholog with a functional annotation but that is specifically upregulated under sulfur depletion [13]. Although taurine is a known sulfur source, no transporter for taurine has yet been identified in *S. cerevisiae*. The conservation of *Yil166cp* with *Jlp1p* suggests *Yil166cp* as the missing taurine transporter. Another interesting example of the potential for protein function prediction comes from the gene pair encoding *Mlo2p* and *Mcm2p*, which is divergently oriented in *Schizosaccharomyces pombe* and most of the Euascomycota. *Mcm2p* is presumed to be part of the replicative helicase [14], whereas *Mlo2p* was identified as leading to incorrect chromosome segregation when overexpressed [15]. The conserved orientation of *Mcm2p* and *Mlo2p* in the fungi indicates that *Mlo2p* may have an earlier role in DNA replication than in the chromosome segregation process itself.

The many gene pairs with a functional relation in Table 1 suggest that the conservation could be used to predict functional links between genes. We quantified the evidence for a functional relation between conserved gene pairs with receiver operator characteristic (ROC) curves, using various databases of protein complexes and pathway information [16–18] [area under the curve (AUC); Supplementary Figure S5]. Divergently transcribed gene pairs that are functionally linked are more conserved in orientation than those that are not functionally related (AUC = 0.63; one-sided Mann-Whitney test; $P < 1.65 \times 10^{-2}$). The predictive value in fungi is similar to that of the Hamming distances of the phylogenetic profiles [19]

(AUC = 0.64). By contrast, for functionally related co-oriented and convergent gene pairs the conservation scores were not higher than for unrelated pairs (AUC ~ 0.5; $P > 0.33$).

Concluding remarks

We have shown that divergently transcribed gene pairs are strongest conserved among adjacent gene pairs in fungi and that the level of conservation correlates with the level of co-expression and the likelihood of being functionally related. Our results, thus, link the functional signals of co-expression and functional similarity specifically to divergent gene pairs rather than adjacency in general [11]. The pronounced conservation of divergent gene pairs suggests that many of them could be regulated by bi-directional promoters, as is the case in humans [6]. Our approach to quantify the degree of conservation of individual orthologous group pairs could serve to identify such bi-directional promoters. Furthermore, we showed how the conservation of divergent orientation can be used for protein function prediction in eukaryotes. The value of conserved divergent orientation has already been shown in prokaryotes, although there its coverage is low compared to the conservation of co-oriented gene pairs in operons [2]. Nevertheless, with the large number of genome sequences expected to become available, the comparative approach will become increasingly important to provide a new source of information about protein function in eukaryotes that is independent for genomics data like co-expression or physical interaction.

Acknowledgements

We thank Michael Egmont-Peterson and Fiona Nielsen for valuable discussions on statistical procedures. This work is part of the BioRange program of the Netherlands Bioinformatics Centre (NBIC), which is supported by the Netherlands Genomics Initiative (NGI) and by the European Union's Sixth Framework Programme EPITEM (CT-2005-019067).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2008.02.003.

References

- Dandekar, T. *et al.* (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23, 324–328
- Korbel, J.O. *et al.* (2004) Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat. Biotechnol.* 22, 911–917
- Spieth, J. *et al.* (1993) Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell* 73, 521–532
- Ganot, P. *et al.* (2004) Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome. *Mol. Cell Biol.* 24, 7795–7805
- Herr, D.R. and Harris, G.L. (2004) Close head-to-head juxtaposition of genes favors their coordinate regulation in *Drosophila melanogaster*. *FEBS Lett.* 572, 147–153
- Trinklein, N.D. *et al.* (2004) An abundance of bidirectional promoters in the human genome. *Genome Res.* 14, 62–66
- Li, Y-Y. *et al.* (2006) Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. *PLoS Comput. Biol.* 2, e74
- Hurst, L.D. *et al.* (2002) Natural selection promotes the conservation of linkage of co-expressed genes. *Trends Genet.* 18, 604–606

- 9 Huynen, M.A. and Snel, B. (2003) Exploiting the variations in the genomic associations of genes to predict pathways and reconstruct their evolution. In *Frontiers in Computational Genomics* (Galperin, M.Y. et al., eds), pp. 145–166, Caister Academic Press
- 10 Farris, J.S. (1977) Phylogenetic analysis under Dollo's law. *Syst. Zool.* 26, 77–88
- 11 Poyatos, J.F. and Hurst, L.D. (2007) The determinants of gene order conservation in yeasts. *Genome Biol.* 8, R233
- 12 Hogan, D.A. et al. (1999) Cloning and characterization of a sulfonate/alpha-ketoglutarate dioxygenase from *Saccharomyces cerevisiae*. *J. Bacteriol.* 181, 5876–5879
- 13 Boer, V.M. et al. (2003) The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J. Biol. Chem.* 278, 3265–3274
- 14 Maiorano, D. et al. (2006) MCM proteins and DNA replication. *Curr. Opin. Cell Biol.* 18, 130–136
- 15 Javerzat, J.P. et al. (1996) Fission yeast genes which disrupt mitotic chromosome segregation when overexpressed. *Nucleic Acids Res.* 24, 4676–4683
- 16 Stark, C. et al. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539
- 17 Kanehisa, M. et al. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280
- 18 Mewes, H.W. et al. (2004) MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* 32, D41–D44
- 19 Pellegrini, M. et al. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4285–4288
- 20 Berbee, M.L. and Taylor, J.W. (2001) Systematics and evolution, part B. In *The Mycota: A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research* (Vol. 7) (McLaughlin, D.J. et al., eds), In pp. 229–245, Springer
- 21 Cleveland, W.S. (1993) *Visualizing Data*, Hobart Press
- 22 Kwok, E.Y. et al. (2006) Evidence for iron channeling in the Fet3p-Ftr1p high-affinity iron uptake complex in the yeast plasma membrane. *Biochemistry* 45, 6317–6327
- 23 Singh, A. et al. (2006) Assembly, activation, and trafficking of the Fet3p.Ftr1p high affinity iron permease complex in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 281, 13355–13364
- 24 DeHoratius, C. and Silver, P.A. (1996) Nuclear transport defects and nuclear envelope alterations are associated with mutation of the *Saccharomyces cerevisiae* NPL4 gene. *Mol. Biol. Cell* 7, 1835–1855

0168-9525/\$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tig.2008.02.003 Available online 28 March 2008

Genome Analysis

Early evolution of eukaryotic DNA-dependent RNA polymerases

Marta Kwapisz, Frédéric Beckouët and Pierre Thuriaux

CEA, iBiTec-S, Service de Biologie Intégrative et Génétique Moléculaire, Gif-sur-Yvette F-91191, France

Eukaryotic DNA-dependent RNA polymerases (Pol I–III) share a conserved core of 12 subunits, which is closely related to archaeal RNA polymerases. Rpb8, a subunit found in Pol I, II and III, was thought to be restricted to eukaryotes. We show here that Rpb8 closely resembles an archaeal protein called G, found only in Crenarchaea, which identifies a last missing link between the core structure of archaeal and eukaryotic RNA polymerases.

Archaeal and eukaryotic transcription: two closely related systems

Transcription is catalysed by two unrelated classes of DNA-dependent RNA polymerases. The first contains the single-subunit RNA polymerase of bacteriophage T7. Moreover, related enzymes transcribe the mitochondrial genome and contribute to chloroplast transcription [1]. The second, and more widespread, class of multi-subunit RNA polymerases (Table 1) includes the eukaryotic (nuclear) enzymes (Pol I, Pol II and Pol III), their archaeal, bacterial and chloroplast counterparts, several polymerases acting on DNA viruses, and the recently discovered Pol IV, which is involved in the silencing of plant heterochromatin [2]. Among them, eukaryotic RNA polymerases are characterised by their complex multi-subunit composition, with up to 17 different subunits in the case of yeast Pol III [3].

The chromatin of eukaryotes imposes a major constraint on transcription by preventing the direct targeting of RNA polymerases to DNA promoters, which probably explains that Pol I, II and III are first directed to pre-initiation complexes before starting transcription. These complexes are minimally defined by the TATAA box binding protein (TBP), which is common to all three transcription systems, by the related Pol II- and Pol III-associated initiation factors TFIIB and Brf1, and by the Pol II-specific TFIIE factor [4]. Bacteria have no chromatin, and their RNA polymerases directly bind to promoters by specialized sigma subunits. By contrast, many Archaea have small DNA-binding proteins that are closely related to histones H3 and H4, and form tetrasomes that evoke the eukaryotic nucleosomes [5,6].

Yeast Pol I, II and III are by far the best characterised eukaryotic RNA polymerases [3,7,8]. They share a conserved core of 12 subunits, made of five polypeptides common to all three polymerases (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12), and of seven distinct but paralogous components present in Pol I, II and III (Table 1). This triplicated core structure is present, with some variations, in all eukaryotic lineages and thus presumably appeared at an early step of eukaryotic evolution [9]. Moreover, the subunit composition of this Pol I, II and III core structure is closely related to the one of the single RNA polymerases operating in Archaea [10,11], as recently illustrated by the electron microscopy and X-ray structure of two archaeal

Corresponding author: Thuriaux, P. (pierre.thuriaux@cea.fr).