whereas two other studies reported biallelic expression starting from day four of development [23,24]. Large-scale identification of single nucleotide polymorphisms in transcribed regions of the chicken genome [25] will facilitate systematic allelic gene expression analyses of imprinted gene orthologues and provide new clues as to the evolution of imprinting mechanisms.

Acknowledgements

We thank H. Hameister for providing BAC HGMP-17-C4 and A. Emser for help in the statistical analysis. This study was supported by the German Research Foundation (HA 1374/6–2 and SCHM 484/21–2) and the Boehringer Ingelheim Foundation.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2005.07.004

References

1 Cattanach, B.M. and Kirk, M. (1985) Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* 315, 496–498
2 Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* 3, 662–673
3 Murphy, S.K. and Jirtle, R.L. (2003) Imprinting evolution and the price of silence. *BioEssays* 25, 577–588
4 Reik, W. and Walter, J. (2001) Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32
5 Walter, J. and Paulsen, M. (2003) The potential role of gene duplications in the evolution of imprinting mechanisms. *Hum. Mol. Genet.* 12, R215–220
6 Kitsberg, D. *et al*. (1993) Allele-specific replication timing of imprinted gene regions. *Nature* 364, 459–463
7 Knoll, J.H. *et al*. (1994) Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. *Nat. Genet.* 6, 41–46
8 Cerrato, F. *et al*. (2003) Paternal imprints can be established on the maternal Igf2-H19 locus without altering replication timing of DNA. *Hum. Mol. Genet.* 12, 3123–3132
9 International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716
10 McQueen, H.A. *et al*. (1998) Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Res.* 8, 621–630
11 Burt, D.W. (2004) The chicken genome and the developmental biologist. *Mech. Dev.* 121, 1129–1135
12 Onyango, P. *et al*. (2000) Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res.* 10, 1697–1710
13 Paulsen, M. *et al*. (2000) Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Hum. Mol. Genet.* 9, 1829–1841
14 Nicholls, R.D. and Knepper, J.L. (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu. Rev. Genomics Hum. Genet.* 2, 153–175
15 Greally, J.M. *et al*. (1998) The mouse H19 locus mediates a transition between imprinted and non-imprinted DNA replication patterns. *Hum. Mol. Genet.* 7, 91–95
16 Brunner, B. *et al*. (2000) Molecular cloning and characterization of the *Fugu rubripes* MEST/COPG2 imprinting cluster and chromosomal localization in *Fugu* and *Tetraodon* nigroviridis. *Chromosome Res.* 8, 465–476
17 Toder, R. *et al*. (1996) The human/mouse imprinted genes IGF2, H19, SNRPN and ZNF127 map to two conserved autosomal clusters in a marsupial. *Chromosome Res.* 4, 295–300
18 Yokomine, T. *et al*. (2005) Structural and functional analysis of a 0.5-Mb chicken region orthologous to the imprinted mammalian Ascl2/Mash2-Igf2-H19 region. *Genome Res.* 15, 154–165
19 Paulsen, M. *et al*. (2005) Evolution of the Beckwith-Wiedemann syndrome region in vertebrates. *Genome Res.* 15, 146–153
20 Schmid, M. *et al*. (1989) Chromosome banding and DNA replication patterns in bird karyotypes. *Cytogenet. Cell Genet.* 52, 139–146
21 Grützner, F. *et al*. (2001) Chicken microchromosomes are hypermethylated and can be identified by specific painting probes. *Cytogenet. Cell Genet.* 93, 265–269
22 Koski, L.B. *et al*. (2000) Monoallelic transcription of the insulin-like growth factor-II gene (Igf2) in chick embryos. *Mol. Reprod. Dev.* 56, 345–352
23 O'Neill, M.J. *et al*. (2000) Allelic expression of IGF2 in marsupials and birds. *Dev.Genes Evol.* 210, 18–20
24 Yokomine, T. *et al*. (2001) Sequence polymorphisms, allelic expression status and chromosome locations of the chicken IGF2 and MPR1 genes. *Cytogenet. Cell Genet.* 93, 109–113
25 International Chicken Polymorphism Map Consortium. (2004) A genetic variation map for chicken with 2.8 milllion single-nucleotide polymorphisms. *Nature* 432, 717–722

doi:10.1016/j.tig.2005.07.004

Network motifs are enriched with transcription factors whose transcripts have short half-lives

Edwin Wang and Enrico Purisima

Biotechnology Research Institute, National Research Council Canada, Montreal, Quebec, Canada H4P 2R2

Network motifs, the recurring regulatory structural patterns in networks, can self-organize to produce networks because of the large ratio of genes to transcription factors (TFs) in genomes. We find a common design principle of these motifs: the TFs whose transcripts have short half-lives are significantly enriched in motifs and hubs. This enrichment becomes one of the driving forces for the emergence of the network scale-free topology, enables the network to adapt quickly to environmental changes and mitigates gene expression fluctuations. Motifs are classified into subtypes that are preferentially used in different cellular conditions.

Introduction

Gene regulatory networks are viewed as directed graphs, in which nodes represent transcription factors (TFs) and

Corresponding author: Wang, E. (Edwin.Wang@cnrc-nrc.gc.ca).
Available online 18 July 2005

operons, whereas the regulatory relationships are represented by edges. In these networks, three major kinds of motifs are observed: single input (SIM), bi-fan and feedforward loop (FFL) [1–3] (Figure 1). Network motifs can be seen as functional and structural units, and the emergence of these motifs leads to the self-organization of the network [2,4]. By self-organization we mean that without the addition of extra connections, the links that are already present in the motifs define an extensive network that includes the majority of nodes in the entire network. We will illustrate this with the *Escherichia coli* gene regulatory network using the known gene regulation data from a literature-mined database, RegulonDB [5] and other sources [6,7]. An explanation of more network terms used in this article is included in supplementary data online.

In a network, a subset of links forms the network backbone that maintains the interconnections (directly and indirectly) between most TFs and thus maintains the integrity of the network. Without these backbone links, the graph would be fragmented into a collection of islands of smaller networks. How important are the motifs for maintaining the integrity of the network? If we remove all the motif links from the *E. coli* gene regulatory network, the network falls apart into disconnected islands (Figure 1 in the supplementary material online). Conversely, if we remove all links that are not part of motifs, we are left with a core network that preserves the backbone links (Figure 1 in supplementary material online).

It is known that bi-fans are essential to maintain the network backbone links [8]. Therefore, we examined whether other motifs are also essential for network integrity. Removal of all FFL links did not destroy network integrity, whereas removal of all SIM links resulted in network fragmentation as did removal of all bi-fan links (supplementary data online). This suggests that bi-fans or SIMs can self-organize to form networks. This begs the question as to how the motifs self-organize to form a network.

The large ratio of genes to TFs in genomes results in self-organization of motifs

When two motifs contain the same TF or gene, they self-assemble (i.e. they automatically form a networked pair of motifs; Figures 2 and 3 in supplementary material online).

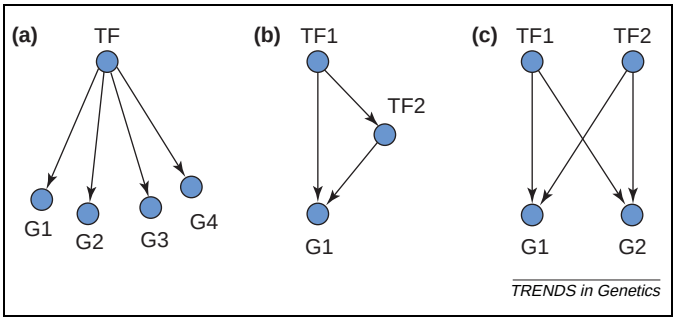

Figure 1. Network motifs in the *Escherichia coli* gene regulatory network. **(a)** Single input module (SIM): a transcription factor (TF) regulates a group of genes (G1, G2, G3 and G4). **(b)** Feedforward loop (FFL): a transcription factor (TF1) regulates the second transcription factor (TF2), both TF1 and TF2 regulate a target gene (G1). **(c)** Bi-fan: both transcription factors TF1 and TF2 regulate both target genes (G1 and G2). In the *E. coli* gene regulatory network, we identified 58 FFLs, 496 bi-fans and 13 SIMs.

The large gene regulatory network arises from the self-organization of motifs that share common TFs or genes. We tested this concept by simulating 496 bi-fans by randomly sampling the *E. coli* genome pool containing 116 TFs and 321 operons (supplementary data). This is the same number of bi-fans found in the original network. Without adding any extra links, the motifs self-organized into a network. The self-organization occurred because of the limited number of distinct TFs relative to the number of bi-fans constructed in the simulation. We find that if the number of randomly generated bi-fans is larger than or equal to the number of TFs used to generate the bi-fans, these bi-fans can self-organize to form a network (supplementary data). This is the situation in cells, where one TF often regulates many target genes. All genomes encode a limited number of TFs but a large number of regulated genes. Therefore, it is common for a genome to have more bi-fans than TFs, consequently leading to the self-organization of motifs into a large network.

Aside from being self-organized, the *E. coli* gene regulatory network is also scale-free [1,9]. Visually, scale-free networks are characterized by the presence of hubs in the network (i.e. nodes that are directly connected to a large number of other nodes). However, the randomly generated bi-fans, although they self-organize, do not form a scale-free topology owing to the even distribution of TFs among the bi-fans. In the real gene regulatory network, the TFs in the bi-fans are unevenly distributed. Randomization tests showed that the TF combinations in bi-fans are significantly different from chance expectation (supplementary material online). The same unevenness in TF pairs is also observed for the FFLs.

Preferential usage of short transcript half-life transcription factors in hubs and network motifs

The uneven TF combinations in FFLs and bi-fans are intriguing and it is tempting to speculate as to its origin. In an attempt to gain some insight into this, we first collected the details of the half-lives of *E. coli* transcripts (THL) from Bernstein and coworkers [10] and mapped them onto the TFs in the network. Among the 116 TFs in the network, 107 of them have THLs that are mapped. They range from 14.4 to 1.9 min with a median of 5.5 min. Taking 5.2 min or less as a short THL, we calculated the percentage of short THL TFs in FFLs and bi-fans. Surprisingly, we found that 60% of the FFL and bi-fan TFs have short THLs. By contrast, 43% of the TFs in the network have short THLs (Table 1). Furthermore, 94.5% of the FFL TF pairs and 92% of the bi-fan TF pairs contain at least one short THL TF (Tables 2 and 3 in the supplementary material online). We also confirmed that this phenomenon does not occur by chance (supplementary data). Extending this analysis to SIMs, we found ~70% of SIM TFs have short THLs (Table 1). These data show that there is a preferential selection of short THL TFs in SIMs, FFLs and bi-fans. This preferential selection will lead to an uneven number of target genes regulated by each TF and therefore become one of the driving forces to generate a scale-free network topology (i.e. it leads to the formation of hubs in the network). In fact, ~70% of hub

Table 1. The enrichment of transcription factors whose transcripts have short half-lives in hubs and network motifs

	TF[a]	Natural rate (%)[b]	Random rate (%)[c]	*P* value[d]
Hub	11, 10	70.0	43.1	<0.05
SIM	13, 10	70.0	43.2	<0.05
FFL	29, 21	61.9	42.9	<0.04
Bi-fan	38, 28	60.1	43.1	<0.03

[a]The first number represents the total number of transcription factors (TFs) in each component (hub, SIM, FFL and bi-fan). The second number represents the number of TFs from the first that have a mapped half-life of a transcript (THL).
[b]The natural rate represents the observed fraction of TFs having short transcript half-lives in the THL-mapped TFs. When the THL of a TF is equal to or shorter than 5.2 min, we say this TF is a short THL TF.
[c]The random rate represents the fraction of the short THL TFs in the randomly sampled TFs.
[d]*P* value represents the probability. Details in statistical analysis are included in supplementary data online.

nodes are TFs with short THLs ($P<0.05$, Table 1). It has been previously reported that FFLs and bi-fans are naturally selected [11]. Here we find that short THL TFs become a selection trait in hubs and in all motifs (SIM, FFL and bi-fan).

Short THL TFs can alter their transcript concentrations quickly, which will facilitate the adaptation of the motif to rapid condition changes [12]. Short THL TFs also mitigate gene expression fluctuations, or internal noise [13,14], which can scramble cell signals and corrupt circadian clocks [15]. Taken together, the network generated by self-organization of these motifs has evolved to be more robust and adaptable to the cellular condition changes. The preferential usage of short THF TFs in hubs and network motifs enables gene expression to turn on and off quickly, which represents a common design principle of these motifs and the network. The frequent occurrence of FFL and bi-fan TF pairs containing one short THL TF can be seen as a criterion for self-organizing FFLs and bi-fans.

Motif and hub subtypes and their usages in different cellular conditions

To explore the relationship between the motifs and THL of TFs, we classified the motifs and hubs into two types based on the THLs of their TFs (Table 2). The classification and the characteristics of these subtypes are discussed in supplementary material. To get a dynamic view of the motif-subtype usages in different cellular conditions, we reconstructed three sub-networks using 22 microarray experimental data of the *E. coli* grown in these conditions: logarithmic growth phase, diauxic shift and the stationary phase (supplementary data).

Table 2 summarizes the dynamic representation of the networks and the motif subtype usages. A more detailed discussion of the results is included in supplementary data. Briefly, the frequencies of Type I FFLs in the three sub-networks are similar, suggesting that FFLs are used as buffers to maintain some biological processes. Type I hubs, bi-fans and SIMs are favored by an active growth condition in which many biological processes are coordinated and quickly respond to the inducing conditions. However, Type II bi-fans are favored by the cellular conditions such as the stationary and diauxic shift, which significantly reduces the level of the biosyntheses of DNA and protein and inhibits aerobic metabolism as reported previously.

Concluding remarks

In conclusion, the large ratio of genes to TFs in genomes leads to a sharing of TFs or genes by motifs and is sufficient to result in their self-organization. We find a common design principle of these motifs: the TFs whose transcripts have short half-lives are significantly enriched in motifs and hubs. This enrichment enables the network to adapt quickly to environmental changes and mitigates gene expression fluctuations, or internal noise. Furthermore, it becomes one of the driving forces for the emergence of the network scale-free topology. Most FFLs and bi-fans contain at least one short THL TF, which can be seen as another criterion for self-assembly of these motifs. We have classified the motifs according to their short THL TF content. We show that the percentage of the different motif subtypes is dependent on the cellular conditions.

Acknowledgements

We thank J. Collado-Vides and his colleagues for providing the data of the RegulonDB database. We thank M.M. Babu at LMB-MRC for many useful suggestions regarding the manuscript that led to the interpretation of hubs having transcripts with short half-lives. We also thank M. Whiteway, P. Lau, H. Hogues, Z. Yu, C. Wu and A. Nantel for discussions. This work is supported by Genome and Health Initiative (GHI), Canada. NRC publication No. 46238.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2005.06.013

Table 2. Dynamic representation of gene regulatory networks and the frequencies of motif types in different cellular conditions[a]

		Static	Aerobic	Diauxie	Stationary
Size	Number of transcription factors	116	61	84	35
	Number of target operons	321	201	229	96
	Number of regulatory interactions	567	300	372	147
Type I hubs	Hub	11 (70.0%)	6 (100.0%)	8 (71.4%)	3 (66.7%)
Type I motif	SIM	13 (70.0%)	11 (91.7%)	9 (69.2%)	5 (71.4%)
	FFL	14 (77.8%)	4 (80.0%)	7 (77.8%)	3 (100.0%)
	Bi-fan	20 (40.0%)	11 (68.8%)	8 (38.1%)	3 (37.5%)

[a]Changes in the gene regulatory networks are tabulated for the static, logarithmic growth, diauxic shift (diauxie) and stationary phases. The last four rows show the number and fraction of hubs and motifs active in the various cellular conditions. Type I FFLs in which the half-life of the transcript (THL) of the first transcription factor (TF) is shorter than the that of the second TF, and Type II FFLs in which the THL of the first TF is longer than that of the second TF. For bi-fans, hubs and SIMs, if the THL of a TF is 5.2 min or less we define it as a short THL TF. A bi-fan is defined as Type I if both TFs have short THLs and as Type II if one of the TF pair has a long THL. Hubs or SIMs are defined as Type I if they have short THL TFs and Type II if they long THL TFs.

References

1 Shen-Orr, S.S. *et al*. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31, 64–68
2 Milo, R. *et al*. (2002) Network motifs: simple building blocks of complex networks. *Science* 298, 824–827
3 Lee, T.I. *et al*. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804
4 Babu, M.M. *et al*. (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* 14, 283–291
5 Salgado, H. *et al*. (2004) RegulonDB (version 4.0): Transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res.* 32, D303–D306
6 Babu, M.M. and Teichmann, S.A. (2003) Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* 31, 1234–1244
7 Grkovic, S. *et al*. (2002) Regulation of bacterial drug export systems. *Microbiol. Mol. Biol. Rev.* 66, 671–701
8 Dobrin, R. *et al*. (2004) Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network. *BMC Bioinformatics* 5, 10
9 Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113
10 Bernstein, J.A. *et al*. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9697–9702
11 Conant, G.C. and Wagner, A. (2003) Convergent evolution of gene circuits. *Nat. Genet.* 34, 264–266
12 Yang, E. *et al*. (2003) Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res.* 13, 1863–1872
13 Hasty, J. *et al*. (2002) Engineered gene circuits. *Nature* 420, 224–230
14 Ozbudak, E.M. *et al*. (2002) Regulation of noise in the expression of a single gene. *Nat. Genet.* 31, 69–73
15 Barkai, N. and Leibler, S. (2000) Circadian clocks limited by noise. *Nature* 403, 267–268

0168-9525/$ - see front matter
doi:10.1016/j.tig.2005.06.013

Intergenomic conflict revealed by patterns of sex-biased gene expression

Tim Connallon and L. Lacey Knowles

Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079, USA

Intergenomic conflict can affect the distribution of genes across eukaryotic genomes. Because the phenotypic optima of males and females often differ, the fitness consequences of newly arisen alleles might not be concordant between the sexes and can be sexually antagonistic – genetic variants favored in one sex are deleterious in the other. In this article, we demonstrate that previously unexplained patterns of sex-biased gene expression in *Drosophila melanogaster* might have evolved by sexual antagonism, and that the majority of sex-biased expression is due to adaptive changes in males, implying that males experience stronger selection than females.

Introduction

The ultimate fate of new mutations is inextricably linked to the evolutionary forces acting on them. Selective pressures commonly differ between males and females, and might actually operate in opposing directions [1]. The evolution of sex-limited gene-expression is expected as a result of this 'sexually antagonistic' genomic tug of war, leaving a characteristic genetic footprint that depends on the degree of dominance on an allele and which of the two sexes benefits [2–4]. The X chromosome is predicted to be a haven for sexually antagonistic variation because dominant alleles that benefit females are under positive selection during two-thirds of their evolution, and rare recessive male-benefiting alleles are masked in females [3].

Because changes in gene regulation, as opposed to coding sequence, are important generators of phenotypic diversity [5], and coding sequence and regulatory evolution are often decoupled [6,7], microarrays (i.e. gene expression profiles) offer a window into the evolutionary processes shaping the genome. Studies have revealed that an amazing proportion of the genome is sex-biased in expression (i.e. expression level is sexually dimorphic) [8–13] and that a significant number of genes show newly derived sex-limitation between recently diverged species [11]. Male- and female-biased genes also exhibit idiosyncratic X-linkage patterns: male-biased genes are strongly deficient, whereas female-biased genes are slightly overabundant [10,13]. However, inferences about the processes driving patterns of sex-biased expression have been constrained by the lack of a conceptual framework in which to (i) identify sex-specific adaptive change; and (ii) translate gene expression data into the classic population genetic constructs of recessive and dominant mutations. Both are fundamental to testing whether patterns of genomic divergence reflect sexually-antagonistic selection.

In this article, we consider published cDNA microarray data [11] under alternative models of adaptive change, with attention to whether shifts in expression correspond to recessive or dominant substitutions. Within this framework, we analyzed thousands of sex-biased genes,

Corresponding author: Knowles, L.L. (knowlesl@umich.edu).
Available online 20 July 2005