(dsRNA) molecules between transcripts from the truncated LTR promoter and the U6 promoter. To rule this out, we inverted the orientation of the U6 cassette using a Gateway lentiviral construct for two of the *OAS1*-inducing U6 vectors (see **Supplementary Note** online). Both shRNAs still induced *OAS1*, indicating that the effect does not depend on the orientation of the Pol III cassette (data not shown).

The magnitude of *OAS1* induction was greater at higher multiplicities of infection with all of the vectors tested. Together with the observation that the *MORF4L2* vector was able to potentiate *OAS1* induction by pAB319 (**Fig. 1b**), the dependence on vector dose is consistent with a model in which shRNAs compete for processing to siRNA, and the accumulation of unprocessed or aberrantly processed Pol III transcripts triggers interferon expression. Northern blotting of cells expressing the pAB319 and pAB322 vectors detected only the presence of the correctly processed siRNAs (**Fig. 1g**), but this assay may not be sensitive enough to detect the small amounts of dsRNA that are sufficient to trigger an interferon response.

In conclusion, we show that a commonly used shRNA construct can induce an interferon response. Although this may not be a concern in initial screens with shRNA banks, we recommend testing for interferon induction before attributing a particular response to the gene targeted. One simple precaution to limit the risk of inducing an interferon response is to use the lowest effective dose of shRNA vector. Finally, we note that many commonly used tumor cells have a defective interferon response[6], which may explain why these effects have not previously been reported.

Accession numbers. The GEO accession numbers for the microarray data are GSM3891 and GSM3892.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank P. Gonczy for critical reading of the manuscript, T. Brummelkamp for supplying pSUPER, the Swiss Institute for Experimental Cancer Research microarray and bioinformatics core facilities for technical assistance and advice on data analysis and the Swiss National Science Foundation NCCR Molecular Oncology Programme and MEDIC Foundation for financial support.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

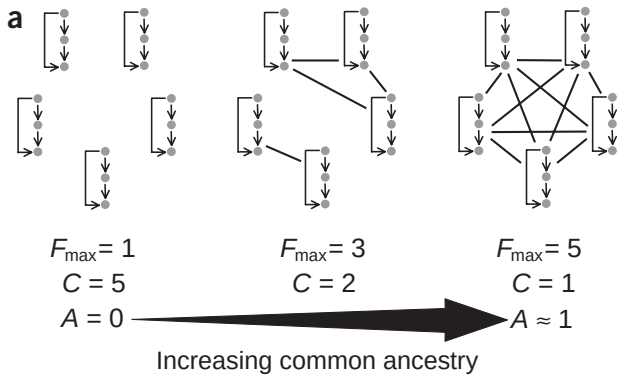
Received 7 February; accepted 12 May 2003
Published online 8 June 2003; doi:10.1038/ng1173

1. Bertram, M.J. *et al. Mol. Cell. Biol.* **19**, 1479–1485 (1999).
2. Yochum, G.S. & Ayer, D.E. *Mol. Cell. Biol.* **22**, 7868–7876 (2002).
3. Brummelkamp, T.R., Bernards, R. & Agami, R. *Science* **296**, 550–553 (2002).
4. Naldini, L. *et al. Science* **272**, 263–267 (1996).
5. Akusjarvi, G., Svensson, C. & Nygard, O. *Mol. Cell. Biol.* **7**, 549–551 (1987).
6. Stojdl, D.F. *et al. Nat. Med.* **6**, 821–825 (2000).

Convergent evolution of gene circuits

Gavin C Conant & Andreas Wagner

Convergent evolution is a potent indicator of optimal design. We show here that convergent evolution occurs in genetic networks. Specifically, we show that multiple types of transcriptional regulation circuitry in *Escherichia coli* and the yeast *Saccharomyces cerevisiae* have evolved independently and not by duplication of one or a few ancestral circuits.



b

	Circuit type	Number of circuits	Number of families (*C*)	Index of common ancestry (*A*)	Largest circuit family (F_{max})
Yeast	Feed-forward	48	44 (46.8 ± 1.9; *P* = 0.08)	0.082 (0.023 ± 0.035; *P* = 0.08)	5 (1.9 ± 1.4; *P* = 0.05)
	Bi-fan	542	435 (469.0 ± 37.7; *P* = 0.18)	0.197 (0.135 ± 0.070; *P* = 0.18)	49 (41.0 ± 31.1; *P* = 0.33)
	MIM-2	176	168 (164.5 ± 8.8; *P* = 0.60)	0.045 (0.065 ± 0.050; *P* =0.60)	5 (7.4 ± 6.2; *P* = 0.59)
	Reg. chain (3)	33	33	0	1
E. coli	Feed-forward	11	11	0	1
	Bi-fan	27	27	0	1

Figure 1 Circuit duplication is rare in yeast and *E. coli*. (**a**) Two indicators of common ancestry for gene circuits. Each of $n = 5$ circuits of a given type (a feed-forward loop for illustration) is represented as a node in a circuit graph. Nodes are connected if they are derived from a common ancestor, that is, if all k pairs of genes in the two circuits are pairs of duplicate genes. $A = 0$ if no circuits share a common ancestor (the graph has n isolated vertices); $A \approx 1$ if all circuits share one common ancestor (the graph is fully connected). The number C of connected components indicates the number of common ancestors (two in the middle panel) from which the n circuits derive. F_{max} is the size of the largest family of circuits with a single common ancestor (the graph's largest component). (**b**) Little common ancestry in six circuit types. We considered two circuits to be related by common ancestry if each pair of genes at corresponding positions in the circuit had significant sequence similarity. Each row of the table shows values of C, A and F_{max} for a given circuit type, followed in parentheses by their average values ± standard deviations and P values, as defined by a permutation test described in **Supplementary Methods** online.

Department of Biology, 167 Castetter Hall, The University of New Mexico, Albuquerque, New Mexico 87131, USA. Correspondence should be addressed to A.W. (wagnera@unm.edu).

Convergent evolution occurs on all levels of biological organization, from organ systems to proteins. For instance, eyes and wings have evolved independently multiple times, and many aquatic vertebrates share a streamlined shape, despite their independent evolutionary origins[1]. On the smaller scale of proteins, lysozymes have been recruited independently for foregut fermentation in bovids, colubine monkeys and a bird[2,3]. Antifreeze glycoproteins in antarctic notothenioids and northern cod (living at opposite ends of the globe) have independently evolved similar amino acid sequences[4].

Recent studies have identified abundant genetic circuit motifs in transcriptional regulation networks of the yeast *S. cerevisiae*[5,6] and the bacterium *E. coli*[6,7]. These circuit motifs include regulatory chains, feed-forward circuits and a 'bi-fan' (**Fig. 1**). Such motifs may have had two principal evolutionary origins. First, they may have come about through the random duplication and subsequent diversification of a few ancestral circuits. Given the high frequency at which genes and genomes undergo duplication[8], this is a plausible scenario. It is equally possible, however, that these circuits arose independently by recruitment of unrelated genes. If such convergent circuit evolution is prevalent, then these circuits owe their abundance to the action of natural selection.

To determine the evolutionary origin of transcriptional regulation circuits, we defined two indicators of common circuit ancestry, A and F_{max}. Consider a genome containing n regulatory circuits, each with k genes and identical topology (for details see **Supplementary Methods** online). A pair of circuits shares a common ancestor if all k gene pairs in the circuit pair are gene duplicates. We next defined a 'circuit graph' whose n nodes represent the n circuits and where an edge connects two nodes (circuits) if the circuits have a common ancestor. Our first indicator, A, of common circuit ancestry, is equal to $A = 1 - (C/n)$, where C is the number of components in the graph (**Fig. 1a**). The greater A is, the greater is the fraction of circuits sharing a common ancestor. Our second indicator is F_{max}, the size of the largest family of circuits with common ancestry (**Fig. 1a**).

We identified duplicate genes using BLASTP[9] at a significance threshold of $E \leq 10^{-5}$ (E values between 10^{-3} and 10^{-11} yield the same results). Using this criterion, neither of two circuit types in *E. coli* showed evidence of common ancestry ($A = 0$ and $F_{max} = 1$ for both; **Fig. 1b**). We also studied 18 yeast circuit types, and only three (feed-forward loops, multi-input modules of size 2 and bi-fans) showed evidence of common ancestry ($A > 0$ and $F_{max} > 1$; **Fig. 1b**). This may be due to chance alone, however, simply because duplicate genes are abundant in the yeast genome. Therefore, we used permutation tests (described in **Supplementary Methods** online) to assess the statistical evidence of A and F_{max}. For no circuit type was A significantly different from the chance expectation. For example, yeast contains 542 bi-fan motifs with $A = 0.197$. The probability of observing $A = 0.197$ by chance is $P = 0.18$: too large to reject the null hypothesis. We observed a marginally significant value of $F_{max} = 5$ for feed-forward loops ($P = 0.05$). Even for this circuit type, however, most circuits (43 of 48) showed independent ancestry.

Our analysis of yeast circuits rests on genome-scale chromatin precipitation experiments that use a statistical error threshold (P_e) to identify true regulatory interactions[5]. The results reported in **Figure 1b** are based on $P_e = 10^{-3}$, but we found the same results when varying P_e between 10^{-2} and 10^{-5}. As above, only feed-forward loops yielded a marginally significant value of $A = 0.11$ ($P = 0.03$) and $F_{max} = 3$ ($P = 0.03$) at $P_e \leq 10^{-4}$. Lowering P_e further to $P_e = 10^{-5}$ yielded $A = 0$ and $F_{max} = 1$.

We also asked whether members of one gene family preferentially occurred in one type of gene circuit. This would be expected if many circuits originated through duplication. Specifically, we asked whether the likelihood of a gene occurring in a given circuit type increases if one of its duplicates occurs in that type. The answer is no (**Table 1**).

Table 1 Gene families are not over-represented in circuit types

Organism	Circuit type	P_{motif}[a]	$P_{motif\|duplicate}$[b]	P[c]
S. cerevisiae	Bi-fan	0.82	0.80	NA
	Feed-forward	0.38	0.42	0.21
	Multi-input motif	0.77	0.76	NA
	Regulator chains	0.64	0.67	0.30
E. coli	Bi-fan	0.50	0.67	0.11
	Feed-forward	0.82	0.67	NA

[a]Probability of a randomly chosen regulatory gene occurring in a given circuit type. [b]Probability of a regulatory gene occurring in a circuit type given that one of its duplicates occurs in that circuit type (see **Supplementary Methods** online). [c]P value for one-sided exact binomial test of the null hypothesis $P_{motif} = P_{motif\ |\ duplicate}$. NA indicates that a test has not been carried because $P_{motif} > P_{motif\ |\ duplicate}$. The number of transcriptional regulators was $n = 112$ and $n = 22$ for the yeast and *E. coli* analyses, respectively.

In sum, we found no common ancestry among the *E. coli* circuit types, the yeast regulatory chains or the yeast multi-input motifs with more than two regulators. Of the remaining three yeast circuit types, two showed common ancestry indistinguishable from that expected by chance. Only feed-forward loops showed marginally significant values of either A or F_{max}, but this finding is not statistically robust. Moreover, most (43 of 48) feed-forward loops have clearly independent origins. We also note that the probability of falsely identifying a pair of circuits as duplicates decreases with increasing circuit size. The larger a circuit is, the less evidence of duplication it shows in our analysis.

Multiple lines of evidence indicate that duplicate genes diverge rapidly in function[10–12]. Our findings that gene circuits do not share common ancestry and that duplicate regulatory genes are randomly distributed across gene circuit types underscore this point, because they imply that duplicate transcriptional regulators can readily evolve new interactions. The short DNA binding sites of transcriptional regulators account for much of this plasticity. In microbes like yeast and *E. coli*, new regulatory interactions can arise rapidly[13], even on the time scale of laboratory evolution experiments[14]. Transcriptional regulation circuits are thus ideal systems for studying convergent evolution, because natural selection has much raw material (variation in regulatory interactions) to shape such circuits.

The finding that gene circuits have evolved repeatedly makes a strong case for their optimal design. For example, the design of a feed-forward loop may serve to activate the regulated (downstream) genes only if the farthest-upstream regulator is persistently activated. Moreover, the same design rapidly deactivates genes once this regulator is shut off[7]. Our results also suggest that convergent evolution, probably rare in protein sequences, may have an important role in the higher organizational level of gene circuits. Stephen Jay Gould famously asked what would be conserved if life's tape, its evolutionary history, was replayed[15]. Transcriptional regulation circuits, it seems, might come out just about the same.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

G.C.C. is supported by the US Department of Energy's Computational Sciences Graduate Fellowship program, administered by the Krell Institute. A.W. would like to thank the US National Institutes of Health and the Santa Fe Institute for their support.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 28 March; accepted 21 May 2003
Published online 22 June; doi:10.1038/ng1181

1. Futuyma, D.J. *Evolutionary Biology* 3rd edn. (Sinauer Associates, Sunderland, Massachusetts, 1998).
2. Stewart, C.-B., Schilling, J.W. & Wilson, A.C. *Nature* **330**, 401–404 (1987).
3. Kornegay, J.R., Schilling, J.W. & Wilson, A.C. *Mol. Biol. Evol.* **11**, 921–928 (1994).
4. Chen, L., DeVries, A.L. & Cheng, C.-H.C. *Proc. Natl. Acad. Sci. USA* **94**, 3817–3822 (1997).
5. Lee, T.I. *et al. Science* **298**, 799–804 (2002).
6. Milo, R. *et al. Science* **298**, 824–827 (2002).
7. Shen-Orr, S.S., Milo, R., Mangan, S. & Alon, U. *Nat. Genet.* **31**, 64–68 (2002).
8. Lynch, M. & Conery, J.S. *Science* **290**, 1151–1155 (2000).
9. Altschul, S.F. *et al. Nucleic Acids Res.* **25**, 3389–3402 (1997).
10. Wagner, A. *Mol. Biol. Evol.* **18**, 1283–1292 (2001).
11. Dermitzakis, E.T. & Clark, A.G. *Mol. Biol. Evol.* **19**, 1114–1121 (2002).
12. Wagner, A. *Proc. Natl. Acad. Sci. USA* **97**, 6579–6584 (2000).
13. Stone, J.R. & Wray, G.A. *Mol. Biol. Evol.* **18**, 1764–1770 (2001).
14. Ferea, T.L., Botstein, D., Brown, P.O. & Rosenzweig, R.F. *Proc. Natl. Acad. Sci. USA* **96**, 9721–9726 (1999).
15. Gould, S.J. *Wonderful Life: The Burgess Shale and the Nature of History* (W.W. Norton, New York, 1989).