Myc and Max: Molecular evolution of a family of proto-oncogene products and their dimerization partner

WILLIAM R. ATCHLEY* AND WALTER M. FITCH†

*Center for Quantitative Genetics, Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614; and †Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92717

Contributed by Walter M. Fitch, July 3, 1995

ABSTRACT The *myc* gene family encodes a group of transcription factors that regulate cell proliferation and differentiation. These genes are widely studied because of their importance as proto-oncogenes. Phylogenetic analyses are described here for 45 Myc protein sequences representing c-, N-, L-, S-, and B-*myc* genes. A gene duplication early in vertebrate evolution produced the c-*myc* lineage and another lineage that later gave rise to the N- and L-*myc* lineages by another gene duplication. Evolutionary divergence in the *myc* gene family corresponds closely to the known branching order of the major vertebrate groups. The patterns of sequence evolution are described for five separate highly conserved regions, and these analyses show that differential rates of sequence divergence (= mosaic evolution) have occurred among conserved motifs. Further, the closely related dimerization partner protein Max exhibits significantly less sequence variability than Myc. It is suggested that the reduced variability in *max* stems from natural selection acting to preserve dimerization capability with products of *myc* and related genes.

Cell proliferation and differentiation are fundamental to growth, development, and evolution. Virtually all multicellular organisms go through well-regulated sequences of cell growth and differentiation during ontogeny, and these sequences have been tightly choreographed by the action of natural selection. Indeed, evolutionary changes in morphology are often keyed to genetic changes in the control of patterns of cell proliferation and differentiation (1).

The *myc* family of proto-oncogenes comprises a group of transcriptional activators important in the regulation of normal cell proliferation and in the transformation and differentiation of many cell lineages (2–5). As a consequence, *myc* genes are an important model for understanding both the control of normal cellular activities and the evolution of essential developmental processes. Transcriptional activity in *myc* is apparently controlled through interaction with the related *max* gene (6–10). The *myc* genes have been found in a wide variety of vertebrates, including mammals, birds, amphibians, and fish; however, to date their occurrence among invertebrates is poorly documented (11). On the basis of its extensive distribution, *myc* has been estimated to be at least 600 million years old (5).

Control of cell proliferation is highly buffered against perturbation because of its fundamental relationship to development. However, control of cell proliferation can go awry if the *myc* gene becomes deregulated as the result of retroviral transduction and insertional mutagenesis, chromosomal translocations, or gene amplification. When deregulation occurs, there is increased cell proliferation, impaired differentiation, and induction of a potentially preneoplastic state (12). After this oncogenic activation, *myc* genes may become associated

with a variety of malignant neoplasms. Thus, in addition to being a paradigm for normal cell activities, *myc* genes are also an important model for understanding malignant diseases (2, 4, 5, 12–15).

An extensive literature exists about many aspects of the *myc* gene family. However, little is known about evolutionary aspects, including the extent to which the gene sequence has been conserved during at least 600 million years of evolution, the relationships of *myc* to various “partner” genes, and the relative rates of evolutionary divergence among the various genes. Such information is needed to better understand the patterns of sequence change and differences in gene expression and function. Herein, we examine the molecular evolution of the *myc* gene family. First, we examine the patterns of genetic divergence among 45 separate amino acid sequences. Second, we describe the evolution of several highly conserved functional components of *myc* genes. Third, we examine evolutionary divergence in Max, a dimerization partner of Myc.

Characteristics of the *myc* Gene Family. The *myc* gene family and various associated genes (*max*, *mad*, and *mxl1*) are all members of the basic helix–loop–helix (bHLH) family of transcription factors. Myc and associated proteins all recognize a basic DNA-binding motif (CANNTG) often referred to as an E box or enhancer box. Further, there are two highly conserved dimerization motifs. The bHLH motif located near the carboxyl terminus consists of about 50 amino acids and can be modeled as two amphipathic helices separated by a flexible “loop.” It is a dimerization motif permitting multimerization with other bHLH proteins. The second conserved dimerization motif is the leucine zipper (ZIP), which generally consists of a heptad repeat of leucine residues and constituting an amphipathic α -helix or coil. The ZIP component occurs between the bHLH domain and the carboxyl terminus (5). Simultaneous presence of both bHLH and ZIP motifs might suggest a bHLH/ZIP family of transcription factors. The evolutionary validity of such a classification is considered in another paper.

The *myc* family includes at least five different and genetically unlinked functional members: c-, N-, L-, S- and B-*myc*. The best known is c-*myc*, which was first identified as the cellular progenitor of the retroviral v-*myc* oncogene MC29 responsible for leukemia in chickens (16). Subsequently, a strong causal link between c-*myc* deregulation and malignancy was established (2, 13). c-*myc* is expressed in virtually all proliferating cells in embryonic and adult tissues. It has greatest expression in mesodermal and endodermal tissues and less in ectodermal tissues. c-*myc* is expressed in a wide variety of malignancies of different cell types. N- and L-*myc*, on the other hand, have much more restricted patterns of tissue and developmental stage expression and they are expressed in a restricted set of tumors. N-*myc* was isolated from human neuroblastomas and L-*myc* from small-cell lung carcinomas (17–19). Two other *myc* genes have been isolated from rats. S-*myc* is an intronless gene

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: bHLH, basic helix–loop–helix.

with significant similarity to the second and third exons of *N-myc* (20) and is believed to have tumor suppression abilities (5, 20). *B-myc* appears to possess only the *myc* second exon sequence and is most similar to *c-myc* (15). *B-myc* is expressed almost ubiquitously, but its function is unknown (4).

The *c-*, *N-*, and *L-myc* genes all have the same three-exon and two-intron structure which encodes multiple transcript species (15). The first exon appears to have a regulatory function and plays an important role in governing oncogenic potency, possibly through transcriptional control mechanisms. Removal of the first exon/intron is associated with a dramatic increase in oncogenic potential (21). Major polypeptide open reading frames occur in the second and third exons (4). The amino and carboxyl ends have sequences required for transformation, transactivation, nuclear localization, DNA binding, and protein oligomerization. The *myc* genes code for extremely short-lived nuclear phosphoproteins which exist in different forms (15). However, in spite of extensive research, the exact biochemical function of the Myc protein is unknown.

Max acts as a "partner" protein and dimerizes with *c-*, *N-*, and *L-Myc* to form a specific DNA-binding complex (12). Max has a length of only 160 amino acids, of which approximately 80 are in the DNA-binding/dimerization domain. Max may serve as a cofactor for Myc in transcriptional activation, forming a Myc/Max hetero-oligomer that binds to CACGTG

DNA sequences and activates transcription through the transcriptional activation domain contributed by Myc (7, 8). These Myc/Max complexes then regulate genes involved in cellular growth and differentiation processes (22). Myc apparently exerts its biological functions through association with Max, which acts as both activator and suppressor of Myc transcriptional activity (7, 23). The interactions between Myc and Max constitute a sensitive switch controlling entry and exit from the cell cycle in response to extracellular stimuli (23).

Phylogenetic Analyses. Amino acid sequences for 45 *myc* genes were initially aligned by using the CLUSTAL computer program (24) followed by manual improvement by eye. Genetic divergence among aligned sequences is expressed as pair-wise distances defined as the proportion of amino acids by which two sequences differ. The patterns of evolutionary change are described by a neighbor joining tree (25) that summarizes the interrelationships among sequences. The results were bootstrapped 500 times to provide confidence levels for the tree topology. The aligned sequences are available from W.R.A.

Evolutionary Relationships Within the *myc* Family. The gene tree in Fig. 1 describes the evolution of the *myc* gene family. A gene duplication producing two gene lineages in vertebrates probably occurred soon after the origin of the vertebrates. Because *myc* is found in fish, amphibians, birds,

FIG. 1. A neighbor joining tree of the protein sequences for the *myc* family of cell proliferation transcription factors. Leg lengths given along branches refer to the number of amino acid replacements per 1000 residues, while bootstrap values greater than 95% are given at the nodes as asterisks. Distances were calculated ignoring gapped positions. The numbers for each taxon refer to the following GenBank accession numbers and species names: 1, L11362, *Xenopus laevis*; 2, L11362, *X. laevis*; 3, P10166, *Mus musculus*; 4, P12524, *Homo sapiens*; 5, J03069, *H. sapiens*; 6, P18444, *Gallus gallus*; 7, P26014, *Serinus canaria*; 8, P04198, *H. sapiens*; 9, S11558, *Marmota monax*; 10, P03966, *Mus musculus*; 11, M36277, *Mus musculus*; 12, S20073, *Rattus norvegicus*; 13, P22555, *Marmota monax*; 14, P24793, *X. laevis*; 15, P23999, *R. norvegicus*; 16, L11710, *Brachydanio rerio*; 17, P06646, *Oncorhynchus mykiss*; 18, P06171, *X. laevis*; 19, M26285, *X. laevis*; 20, S22574, *X. laevis*; 21, S07556, *X. laevis*; 22, P15171, *X. laevis*; 23, S22545, *X. laevis*; 24, K02082, *G. gallus*; 25, P06647, avian retrovirus; 26, M16529, avian proviral retrovirus myc; 27, M74581, avian retrovirus Myc protein; 28, P01109, *G. gallus*; 29, P12523, Myc transforming protein; 30, A93947, avian myelocytomatosis virus MC29; 31, P06295, Myc transforming protein; 32, P01110, avian myelocytomatosis virus MC29; 33, A01108, *Mus musculus*; 34, P09416, *R. norvegicus*; 35, P06877, *Felis catus*; 36, M15204, feline leukemia virus v-Myc; 37, M25762, feline proviral (FTT) v-Myc gene; 38, A01350, Burkitt lymphoma, *H. sapiens*; 39, P01106, *H. sapiens*; 40, P23583, *Pan troglodytes*; 41, M88116, *Hylobates lar*; 42, M88115, *Calithrix jacchus*; 43, S03325, *Marmota monax*; 44, A31198, *R. norvegicus*; and 45, M80364, *Asterias vulgaris*.

and mammals, this gene family in vertebrates is at least 400 million years old (26). The separate N-*myc* and L-*myc* sequences arose from a second gene duplication. Indeed, there are a large number of instances of paralogy (gene homology by duplication) in this tree even allowing for the fact that some pairs of sequences might be allelic. Various paralogues of c-, L-, and N-*myc* occur independently in *Xenopus*, mouse, woodchuck, and human. Additionally, B-*myc*, which possesses only a single exon, was derived very early from the lineage that gave rise to the mammalian c-*myc* sequences. S-*myc*, on the other hand, arose early in tetrapod evolution from the lineage that gave rise to N-*myc*.

The statistical reliability of the overall phylogenetic results is quite high as shown by the bootstrap values, particularly in the L- and N-*myc* genes and at the major nodes in c-*myc*. The c-*myc* values are lower in some instances toward the tips of the branches where the sequences become very similar and there is little genetic variation.

The various members of the *myc* gene family have differing characteristics in terms of tissue and developmental stage expression, relative oncogenicity, and retroviral interactions (2, 4, 5, 15). Fig. 1 indicates that three separate lineages (c-, L-, and N-*myc*) arose very early in the evolution of this gene family and exhibit considerable genetic differentiation.

The branching order within the c-*myc* lineage corresponds to the known evolutionary branching patterns of the major vertebrate groups (26): fish divergence occurs first, followed by amphibians, and finally birds and mammals. Table 1 provides a series of comparisons among major vertebrate groups, their approximate ages, and the extent of genetic divergence. Table 2 provides the regression of amino acid differences against the age in millions of years for each comparison given in Table 1. All regression lines are fitted through the origin and involve both the number of amino acid differences and the proportion of amino acid differences (differences scaled by the length of the aligned sequence). Regression of the c-Myc data gives the equation of Y (c-Myc amino acid differences) = $0.37 (\pm 0.007) X$ (age of origin in millions of years) ($P < 0.001$). Thus, there is 0.37 amino acid difference per million years for c-Myc.

Within primates, the order of branching of the taxa is also evolutionarily correct. Two human sequences are represented here—i.e., the normal c-*myc* gene and a sequence that has been oncogenically activated by a translocation involved in the genesis of Burkitt lymphoma (28). The amino acid sequence derived from the translocated c-*myc* protein differs from the normal human c-Myc by 15 amino acids. However, the neighbor-joining algorithm correctly placed the two human se-

quences at the same node and different from chimp. The unweighted pair-group method of averages (UPGMA) algorithm on the same genetic distance matrix placed the Burkitt lymphoma sequence between marmoset and woodchuck.

Two distinct c-*myc* genes (c-*myc1* and c-*myc2*) occur in *Xenopus*, one active in oocytes and the other active in both oocytes and postgastrula embryos (29, 30). This additional gene is hypothesized to have arisen when the genome of the ancestor of the *Xenopus laevis* group was duplicated (31). The duplication of the *Xenopus* genome to a tetraploid level means that these organisms have two pairs of loci for each function. The amino acid sequences for the two groups of c-*myc* genes in *Xenopus* used in these analyses differ by an average of 25.6 (± 0.9) residues. On the basis of albumin data, Bisbee *et al.* (31) estimated that the *Xenopus* genome was duplicated approximately 30 million years ago. However, extrapolating from regression of c-Myc divergence onto the age of origin on the vertebrate groups divergence suggests that duplication of the *Xenopus* genome occurred approximately 78 million years ago. The standard error of this estimated age of origin (± 1.31) is unreasonably small and exists because the c-Myc data (Table 1) consist essentially of two groups of data points. The multiple correlation (r^2) associated with this regression analysis is 0.99.

Transduction of a cellular protooncogene by a retrovirus produces an oncogene capable of inducing sarcomas, leukemias, and carcinomas. The transition of a normal c-*myc* allele to a transduced mutant v-*myc* allele with oncogenic function is usually accompanied by various amounts of change in DNA sequence and the control of gene expression. Five transforming avian retroviruses (v-*myc*) are included in these analyses (MC29, MH2, OK10, HBI, and FH3) together with a normal (non-transduced) c-*myc* from the chicken.

The gene tree shows three groups of avian Myc proteins (Fig. 1). The chicken c-Myc sequence is most similar to FH3 (32) and OK10 and differs from FH3 by a single amino acid difference and from OK10 by two amino acids. HBI, which was derived from MC29 (33), differs from MC29 by an average of 5 amino acids (34, 35). The three MH2 v-Myc sequences differ from chicken c-Myc by an average of 31.6 amino acids. Within the mammalian clade, two v-Myc sequences from the cat involving the feline leukemia virus (FeLV) are included (36) together with the normal cat c-Myc. Cat c-Myc differs from these two v-Myc sequences by an average of four amino acids.

As in the case of c-*myc*, there are also two L-*myc* genes in *Xenopus*—i.e., L-*myc1* and L-*myc2* genes (37). These genes, which encode proteins that differ by 32 amino acid residues, also possibly arose with the early genomic doubling in *Xenopus*

Table 1. Amino acid difference values for major taxa for c-Myc and Max together with the age at which the evolutionary split between the taxa is assumed to have occurred

Comparison	Age, Myr	Difference value (changes per site)			
		Myc	Max	Myc motif	Max motif
Human-chimp	7.5	2		0	
Human-gibbon	17.5	10		2	
Chimp-gibbon	17.5	10		2	
Human-marmoset	40	16		2	
Chimp-marmoset	40	16		2	
Gibbon-marmoset	40	14		2	
Mammal-bird	300	125 (0.232)	9 (0.056)	11.7 (0.154)	0.67 (0.008)
Mammal-amphibian	360	138 (0.256)	36 (0.225)	22.6 (0.297)	1.83 (0.023)
Bird-amphibian	360	137 (0.255)	36 (0.225)	17.0 (0.224)	1.50 (0.018)
Mammal-fish	420	148 (0.275)	22 (0.138)	22.2 (0.224)	1.33 (0.016)
Bird-fish	420	143 (0.266)	21 (0.131)	15.5 (0.204)	1.00 (0.012)
Amphibian-fish	420	156 (0.290)	35 (0.219)	24.8 (0.326)	1.50 (0.018)
Aligned length, aa		538	160	76	81

Vertebrate ages (in millions of years) are from Benton (26) and primate values are from Hayasaka *et al.* (27). Values in parentheses are the average amino acid changes per site. Values are given for the full sequence (Myc and Max) and for the bHLH and ZIP motifs (Myc motif).

Table 2. Regression analyses of relative rates of genetic divergence for c-Myc and Max protein sequences given in Table 1

Protein	Data	Regression coefficient
Complete data (df = 11)		
c-Myc	Raw data	0.369 $\pm$ 0.007
Data for major taxa only (df = 5)		
c-Myc	Raw data	0.369 $\pm$ 0.010
Max	Raw data	0.070 $\pm$ 0.001
c-Myc	Proportions	0.000686 $\pm$ 0.000018
Max	Proportions	0.000437 $\pm$ 0.000069
c-Myc	Motif, raw data	0.050 $\pm$ 0.004
Max	Motif, raw data	0.003 $\pm$ 0.0004

Regression equations are fitted through the origin and protein sequence divergence is the dependent (Y) variable, while vertebrate age in millions of years is the independent (X) variable. The regression equation is of the form Y (sequence divergence) = a [X (age of origin)]. Regression of the number of amino acid differences is described as "Raw data," while the other comparisons are for data expressed as proportions of the entire sequence to correct for differences in sequence lengths. Proportions were computed as the number of amino acid differences divided by the total length of the sequence. Analyses of only the bHLH and ZIP motifs are described as "Motif, raw data." Regression coefficients are presented $\pm$ SD. A t statistic testing that the null hypothesis that the regression coefficient (a) is zero was rejected in each instance at $P = 0.001$.

referred to above. There is also a second L-*myc* gene in humans (L-*myc2*), found on the X chromosome, whose amino acid sequence differs from L-*Myc1* by 126 residues. Human L-*myc2* is thought to be a processed pseudogene and whether it is a functional gene is open to question (38).

Evolution of Conserved Sequences. These analyses indicate at least five discrete, functionally independent domains have been highly conserved among c-, L-, and N-*myc* lineages. These conserved domains are separated by blocks of amino acids exhibiting considerably less similarity. Such an arrangement is consistent with the concept of a transcription factor containing separable transcriptional activation and specific DNA-binding domains (39). Deletion of any of these conserved motifs seriously alters *myc* function.

Because these various highly conserved *myc* motifs have quite different functions, they are probably subject to different selection pressures. This raises at least two questions, including (i) do these different motifs within the same gene exhibit differing rates of evolutionary change (i.e., mosaic evolution), and (ii) do these motifs contain equivalent amounts of phylogenetic information and give equivalent estimates of phylogenetic relationships? To resolve these questions, separate analyses were carried out on the individual conserved components.

Two conserved regions are the bHLH and ZIP motifs discussed earlier. The remaining conserved motifs consist of three "Myc boxes" located in the second exon. The molecular role of the Myc boxes is obscure (40), and even the number and composition of the boxes are not well established (5, 11, 40). These sequences are thought to be independent sites of transcriptional activation and may be required for the transforming activity of the protein (11, 39, 41, 42). Myc box B could be involved in transcriptional repression through interaction with factor TF-III (43). Experimental manipulation of these Myc boxes suggests that they are involved in neoplastic transformation. Deletions spanning Myc box A disable but do not eliminate transforming activity (40). The compositions of these three Myc boxes in human c-Myc (11) are as follows (consensus sequence in vertebrate c-, N-, and L-Myc used in these analyses underlined: box A = PSEDIWKKFELLTPPLSPSRRS; box B = KNIIIQDCMWSGFSAAAKL; and box C = SECIDPSVVEPYPLNDSS (box C not found in L-Myc).

Separate neighbor joining trees were produced for the bHLH motif (52 amino acids), the ZIP domain (24 amino acids), and three Myc boxes (63 amino acids). Several things

are evident from these trees. First, the bootstrap values indicate that the major gene lineages described in Fig. 1 persist. Bootstrap values are low for some of the finer structure of the trees, which is to be expected when the sequence lengths are short and exhibit little if any genetic variation. Second, there are a number of instances where identical sequences exist across several species for significant portions of the gene. For example, c-Myc in all primates (except marmoset) together with rat and cat have identical bHLH sequences, while humans, chicken, and woodchuck have identical bHLH sequences for N-Myc. Further, for ZIP, all of the *Xenopus* c-Myc sequences are identical. Finally, all mammals (except the human Burkitt lymphoma sequence) are identical for the three c-Myc boxes. These results indicate that, as expected, the rate of evolutionary change is quite different in these conserved components relative to the remainder of the gene. On the other hand, there has been considerable change in the Myc boxes in v-Myc MH2 sequences relative to other components of the gene. Recall that some of the Myc boxes are thought to be involved with neoplastic transformation.

Additional insights about the evolution of the *myc* gene family can be obtained by examining the functions of these components of the gene and their relationships with other gene families. The bHLH and ZIP motifs are rather common mechanisms for DNA binding and dimerization. The bHLH motif is not restricted to the *myc* family but rather is used to characterize a rather large collection of transcription factors involved with cell growth, differentiation, and transcriptional control, including the *MyoD*, *USF*, *twist*, and *achaete-scute* gene families. Likewise, the ZIP motif occurs in a variety of unrelated gene families, including non-bHLH genes like the oncogenes *jun* and *fos* as well as the CREB/ATF family of transcription factors (44). Indeed, because of its simple structure, one could speculate that the ZIP motif has arisen independently in multiple unrelated genes or has evolved by modular evolution through a process similar to exon shuffling.

The three Myc boxes, on the other hand, appear to be unique. These motifs are assumed to be involved in transcriptional activation and neoplastic transformation. Employing the National Center for Biotechnology Information computer program BLAST, we used the sequence for each of the boxes as a "probe" of the nonredundant GenBank protein data base. We found no sequences closely related to any of the three Myc boxes. This suggests that these highly conserved motifs are unique to the *myc* gene family and are involved with some specific *myc* function.

Evolutionary Aspects of *myc* and *max*. The intimate association among the *myc* and *max* genes raises interesting evolutionary questions about the evolution of *myc* and *max* genes and their patterns of genetic variation. How closely related are *myc*, *max*, *mad*, and *mxl1*? Our unpublished results on sequence evolution in the entire bHLH family of transcription factors suggest that the *myc* and *max* gene families share a common ancestor, as do *mad* and *mxl1*. However, the two pairs of genes are more distantly related.

Within the *max* family, there are several instances of paralogy (Fig. 2). One occurs where *Xenopus* branches off before zebrafish and out of the accepted evolutionary branching pattern for vertebrates (26). A second instance is found in *Xenopus*, where two separate sequences occur that may have arisen with the previously described genome duplication. The *Xenopus* sequences differ at several sites, including a rather extensive set of insertions that distinguishes them from other vertebrate *max* sequences. Finally, a second human sequence is known that has a nine-residue insertion that has arisen from a single gene by alternative splicing (12). Inclusion of the latter sequence does not enhance the phylogenetic analyses because the additional residues simply appear as an autapomorphy.

Table 2 gives the regression of amino acid difference for the Max protein on time, Y (amino acid differences) = $0.070 (\pm 0.011)$

FIG. 2. Neighbor joining tree for Max protein sequences. The GenBank accession numbers and species names are as follows: 1, S67549, *Rattus norvegicus*; 2, M63903, *Mus musculus*; 3, A38431, *Homo sapiens*; 4, L12469, *Gallus gallus*; 5, L11711, *Brachydanio rerio*; and 6, L09739, *Xenopus laevis*. See Fig. 1 for additional details.

X (age in millions of years) ($P < 0.01$). Hence, there has been 0.07 amino acid change per million years for Max.

Because of the intimate association in transcriptional regulation between *myc* and *max* it is important to ask whether *myc* and *max* are evolving at the same rate. Because the proteins are quite different in length, the average number of substitutions *per site* for c-Myc and Max was computed (Table 1). These values are consistently less for each value in Max compared with c-Myc. A Mann-Whitney two-sample rank test (45) indicates that the medians of the two samples differ at $P < 0.01$. If the comparison is restricted to only the highly conserved bHLH and ZIP motifs, the differences in rates of evolution are more striking. In each instance, Max shows at least an order of magnitude less variability than Myc.

These results persist when the regression slopes for the c-Myc and Max protein divergence over time are compared (Table 2). Results of statistical tests for homogeneity of slope (45) are as follows: (i) test of homogeneity in rate of divergence over time in the entire sequence in c-Myc versus Max for the number of amino acid differences ($P < 0.001$) and for differences scaled for aligned sequence length ($P < 0.01$), and (ii) rate of divergence in the conserved bHLH/ZIP motif over time for the number of differences ($P < 0.001$). Thus, Max exhibits a significantly lower rate of sequence divergence over time than c-Myc for both the entire sequence and the conserved motifs.

The greatly reduced variability in Max sequences may stem from its binding and dimerization with at least three other genes. Natural selection may oppose variation that would interfere with the rather broad interactive activities between Max and Myc, Mad and Mxi1. Including Mad and Mxi1 does not clarify the analyses of relative variability at present because there are so few sequences available for these two genes.

We are grateful to Helene Van for her technical assistance. We are indebted to Robert Eisenman, Stephanie Curtis, Nicole Schreiber-Agus, and Steven Spiker for their helpful comments on the manuscript. W.R.A. was supported by the Alfred P. Sloan Foundation and by grants from the National Institutes of Health (GM-45344) and the National Science Foundation (BSR-910718). W.M.F. was supported by the National Science Foundation (DEB9096152).

- Atchley, W. R. & Hall, B. K. (1991) *Biol. Rev.* **66**, 101–157.
- Cole, M. D. (1986) *Annu. Rev. Genet.* **20**, 361–384.
- Luscher, G. & Eisenman, R. N. (1990) *Genes Dev.* **4**, 2025–2035.
- Marcu, K. B., Bossone, S. A. & Patel, A. J. (1992) *Annu. Rev. Biochem.* **61**, 809–861.
- Meichle, A., Philipp, A. & Eilers, M. (1992) *Biochim. Biophys. Acta* **1114**, 129–146.
- Cole, M. D. (1991) *Cell* **65**, 715–716.
- Blackwood, E. M., Luscher, B. & Eisenman, R. N. (1992) *Genes Dev.* **6**, 71–80.

- Kato, G. J., Lee, W. M. F., Chen, L. & Dang, C. V. (1992) *Genes Dev.* **6**, 81–92.
- Ayer, D. E., Kretzner, L. & Eisenman, R. E. (1993) *Cell* **72**, 211–222.
- Zervos, A. S., Gyuris, J. & Brent, R. (1993) *Cell* **72**, 223–232.
- Walker, C. W., Boom, J. D. & Marsh, A. G. (1992) *Oncogene* **7**, 2007–2012.
- Blackwood, E. M., Luscher, B., Kretzner, L. & Eisenman, R. N. (1991) *Cold Spring Harbor Symp. Quant. Biol.* **56**, 109–117.
- Eisenman, R. N. (1989) in *Oncogenes and the Molecular Origins of Cancer*, ed. Weinberg, R. A. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 175–221.
- Bishop, J., Eilers, M., Katzen, A. L., Kornberg, T., Ramsay, G. & Schirm, S. (1991) *Cold Spring Harbor Symp. Quant. Biol.* **56**, 99–107.
- DePinho, R. A., Schreiber-Agus, N. & Alt, F. W. (1991) *Adv. Cancer Res.* **57**, 1–45.
- Sheiness, D. & Bishop, J. M. (1979) *J. Virol.* **31**, 514–521.
- Schwab, M., Alitalo, K., Klemmner, K.-H., Varmus, H. E., Bishop, J. M., Gilbert, F., Brodeur, G., Goldstein, M. & Trent, J. (1983) *Nature (London)* **305**, 245–248.
- Kohl, N. E., Kamda, N., Schreck, R. R., Bruns, G., Latt, S. A., Gilbert, F. & Alt, F. W. (1983) *Cell* **35**, 359–367.
- Nau, M. M., Brooks, B. J., Battey, J., Suasville, E., Gazdar, A. F., Kirsh, I., McBride, O. W., Bertness, V., Hollis, G. F. & Minna, J. D. (1985) *Nature (London)* **318**, 69–73.
- Sugiyama, A., Kume, A., Nemoto, K., Lee, S. Y., Asami, Y., Nemoto, Y., Nishimura, S. & Kuchino, Y. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9144–9148.
- Xu, L., Wallen, R., Patel, V. & DePinho, R. A. (1993) *Oncogene* **8**, 2547–2553.
- Schreiber-Agus, N., Horner, J., Torres, R., Chiu, F. C. & DePinho, R. A. (1993) *Mol. Cell. Biol.* **13**, 2765–2775.
- Amati, B., Brooks, M. W., Levy, N., Littlewood, T. D., Evan, G. I. & Land, H. (1993) *Cell* **72**, 233–245.
- Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992) *Comput. Appl. Biosci.* **8**, 189–191.
- Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
- Benton, M. J. (1990) *J. Mol. Evol.* **30**, 409–424.
- Hayasaka, K., Gojobori, T. & Horai, S. (1988) *Mol. Biol. Evol.* **5**, 626–644.
- Rabbits, T. H., Hamlyn, P. H. & Baer, R. (1983) *Nature (London)* **306**, 760–765.
- Vriz, S., Taylor, M. & Mechali, M. (1989) *EMBO J.* **8**, 4091–4097.
- King, M. W. (1991) *Nucleic Acids Res.* **19**, 5777–5783.
- Bisbee, C. A., Baker, M. A. & Wilson, A. C. (1977) *Science* **195**, 785–787.
- Tikhonenko, A. T. & Linial, M. L. (1992) *J. Virol.* **66**, 946–955.
- Smith, D. R., Vennstrom, B., Hayman, M. J. & Enrietto, P. J. (1985) *J. Virol.* **56**, 969–977.
- Watson, D. K., Reddy, E. P., Duesberg, P. H. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2146–2150.
- Reddy, E. P., Reynolds, R. K., Watson, D. K., Schultz, R. A., Lautenberger, J. & Papas, T. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2500–2504.
- Stewart, M. A., Forrest, D., McFarlane, R., Onions, D., Wilkie, N. & Neil, J. C. (1986) *J. Virol.* **154**, 121–134.
- Schreiber-Agus, N., Torres, R., Horner, J., Lau, A., Jamrich, M. & DePinho, R. A. (1993) *Mol. Cell. Biol.* **13**, 2456–2468.
- Morton, C. C., Nussenzweig, M. C., Sousa, R., Sorenson, C. D., Pettengill, O. S. & Shows, T. B. (1989) *Genomics* **4**, 367–375.
- Kato, G. J., Barrett, J., Villa-Garcia, M. & Dang, C. V. (1990) *Mol. Cell. Biol.* **10**, 5914–5920.
- Prendergast, G. C. & Ziff, E. B. (1992) *Trends Genet.* **8**, 91–96.
- Sarid, J., Halazonetis, T. D., Murphy, W. & Leder, P. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 170–173.
- Min, S. & Taparowsky, E. J. (1992) *Oncogene* **7**, 1531–1540.
- Roy, A. L., Carruthers, C., Gutjahr, R. & Roeder, R. G. (1993) *Nature (London)* **365**, 359–361.
- Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
- Sokal, R. R. & Rohlf, F. J. (1981) *Biometry* (Freeman, New York).