Letter to the Editor

Positive Selection in the Evolution of Mammalian Interleukin-2 Genes

Jianzhi Zhang[1] *and Masatoshi Nei*

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University

In vertebrates, there are many secreted regulatory proteins that participate in host defense. Interleukin-2 (IL-2) is one of these proteins, and it is secreted primarily by activated T lymphocytes. Interaction between IL-2 and its receptor on the T cell membrane triggers several signal transduction pathways, resulting in clonal expansion of T cells. While this proliferation-promoting activity is believed to be its main function, IL-2 can stimulate the functional differentiation of T cells as well. IL-2 is also known to be a proliferation and differentiation factor for a variety of cell types, such as B cells, natural killer cells, and myeloid cells (reviewed in Goldsmith and Greene 1994; Gaulton and Williamson 1994). Due to its ability to upregulate the immune system, IL-2 has been widely used in immunotherapy for a number of diseases, including cancers and AIDS (e.g., Macey and Johnston 1990; Zou et al. 1999). IL-2 has also been of interest to evolutionists for testing the molecular-clock hypothesis (Gillespie 1989; Ohta 1995). Here, we analyze the sequences available in GenBank and report detection of positive Darwinian selection in ancestral IL-2 genes of artiodactyls and discuss its implications.

The IL-2 gene sequences of 15 mammalian species were obtained from GenBank (see fig. 1*A* for species names) and were aligned using CLUSTAL V (Higgins, Bleasby, and Fuchs 1992) with some visual adjustments. The number of codons in the alignment was 152 after the gaps were removed. (GenBank accession numbers and the alignment are available on request). We assumed that the phylogenetic tree of the species used was as that shown in figure 1*A*, which is generally accepted by molecular evolutionists (Hayasaka, Fujii, and Horai 1996 and references therein; de Jong 1998) and is also supported by the IL-2 gene.

We first estimated the numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions per site between 105 pairs of the 15 IL-2 sequences by the modified Nei-Gojobori method (Nei and Gojobori 1986; Zhang, Rosenberg, and Nei 1998) (table 1). The Jukes-Cantor corrected d_S values were all smaller than 0.7, suggesting that there was no serious saturation of synonymous substitutions. This was particularly so for comparisons among the nonrodent species, for which the d_S values were generally smaller than 0.2. It was observed that d_N was greater than d_S for many comparisons involving species of artiodactyls, carnivores, and primates. To examine the possibility of the occurrence of positive selection and to find the evolutionary lineages in which it would have occurred, we estimated the numbers of synonymous (b_S) and nonsynonymous (b_N) substitutions per site for each tree branch (fig. 1*A*) from the pairwise distances (see Zhang, Rosenberg, and Nei 1998). Several branches showed higher values of b_N than b_S. For example, the b_N/b_S ratio was 3.5 for the exterior branch leading to the deer (branch *w* in fig. 1*A*), and the difference between b_N and b_S for this branch was statistically significant ($Z = 1.87$, $P < 0.05$; one-tail Z test). For branch *z*, which preceded branch *w* (fig. 1*A*), $b_N/b_S = 2.1$, and b_N was again significantly greater than b_S ($Z = 1.79$, $P < 0.05$). These results suggest that positive selection operated in branches *z* and *w*. In no other branches of the tree is b_N significantly greater than b_S. However, it is interesting to note that two branches preceding branch *z*, namely, *x* and *y* (fig. 1*A*), also showed $b_N > b_S$, although their difference was not significant. Because the four branches *x*, *y*, *z*, and *w* are contiguous, it is possible that a wave of positive selection operated for all four branches. To examine this hypothesis, we analyzed the four branches together. This analysis showed that $b_N/b_S = 2.0$ and that the difference between b_N and b_S was highly significant ($Z = 2.12$, $P < 0.01$). The total numbers of synonymous and nonsynonymous substitutions for the four branches were estimated to be 15.1 and 76.7, respectively, from the b_S values and the b_N values. Therefore, the numbers are not small and the result from the Z test is reliable (Zhang, Kumar, and Nei 1997). These results suggest that positive selection operated for the evolutionary lineage between node α and the deer, covering branches *x*, *y*, *z*, and *w* (fig. 1*A*), which corresponds to an evolutionary time of 85–90 Myr (Kumar and Hedges 1998).

To examine the types of amino acid substitutions that occurred in branches *x*, *y*, *z*, and *w*, we inferred the ancestral gene sequences at each interior node of the tree by the distanced-based Bayesian method (Zhang and Nei 1997). The accuracy of the inference (posterior probability) was quite high, except for a few sites. Use of alternative ancestral states inferred from the parsimony method (Fitch 1971) at these sites did not change our conclusion. The numbers of synonymous (*s*) and nonsynonymous (*n*) substitutions were then counted for each branch (fig. 1*A*). If the sequence evolution is neutral, n/s is expected to be equal to N/S, where S and N are the potential numbers of synonymous and nonsynonymous sites of the sequence, respectively. For the present data, $N/S = 329/127 = 2.59$. We found that for each of the branches *x*, *y*, *z*, and *w*, n/s is greater than N/S. When the *n* and *s* values for the four branches were summed, we obtained $n/s = 78.3/14.7 = 5.33$, which is

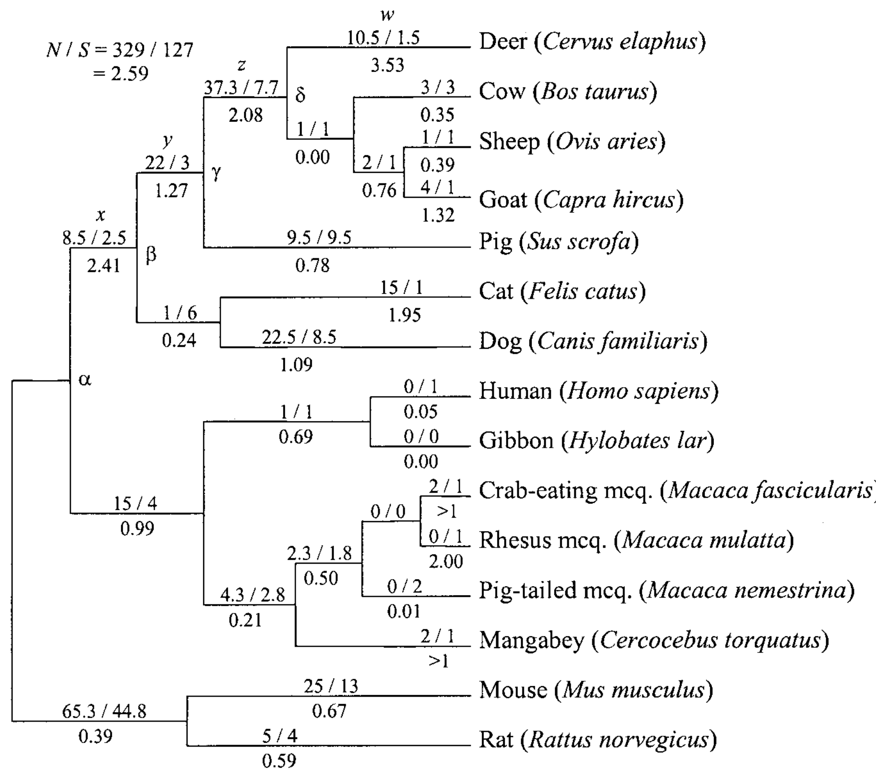
[1] Present address: National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland.

Key words: interleukin-2, positive selection, artiodactyls, substitution pattern.

Address for correspondence and reprints: Jianzhi Zhang, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Building 10, Room 11N104, 9000 Rockville Pike, Bethesda, Maryland 20892. E-mail: jzhang@niaid.nih.gov.

Mol. Biol. Evol. 17(9):1413–1416. 2000

A



B

```
                                      +                +-                              +
                                   -+ - +- +           -+ +   -       +   -     -        -  +80
Node α MYKMQLLSCI ALTLALVTNS APTSSSKETQ QQLEQLLLDL QLLLNGINNY KNPKLSRMLT FKFYMPKKAT ELKHLQCLEE
Node β .......... .......A.. .......... .......... ......V... E......... .......... ........V.
Node γ .......... .......A.G .......N.K K...P..... ....KEVK.. E.A....... .......Q.. ........V.
Node δ ...I...... .......A.G ......GN.M KEVKS..... ....EKVK.P E.L.....H. .N.....N.. .....K..L.
Deer   ...I...... .......A.G ......GN.M KEVKS..... ....EKVK.P E.L...K.H. .N.F...N.. .....N..L.

                         -
                         +        -  ++    - -       +    ++   +              +     152
Node α ELKPLEEVLN LAQSKNFHLR IKELISNINV TVLKLKGSET TFTCEYDDET ATIVEFLNRW ITFCQSIIST LT
Node β .......... .......... ....M..... .......... S......... ........K. .......F.. ..
Node γ ...A...... ......SDS. ...SMN.... ...E...... S......... V.A.....K. .......Y.. ..
Node δ ...L...... ..P...LNP. ..DSMD..KR I..E.Q.... S.......A. VKA.....K. .......Y.. M.
Deer   ...L..D..S .SP...LNPK ..DSMD..KR I..E.Q.... S.K....AA. VKA.....K. .....R.Y.. M.
```

FIG. 1.—Positive selection in the evolution of IL-2 genes. *A*, Test of positive selection. The *n* and *s* values are shown above the branches, whereas b_N/b_S ratios are given below the branches. When both b_S and b_N are estimated to be 0, the ratio is not shown, and when $b_S = 0$ and $b_N > 0$, a ratio of >1 is indicated. The transition/transversion ratio was estimated to be 1.5 by Kimura's (1980) method. Mcq = macaque. *B*, Charge changes of IL-2 for the branches *x, y, z,* and *w*. An amino acid change that increases the positive charge is denoted by a "+," whereas a change that decreases the positive charge is denoted by a "−." The sequences presented are inferred ancestral sequences, except that of the deer.

significantly greater than *N/S* ($P < 0.005$; Fisher's exact test). This *n/s* test suggests that positive selection operated in branches *x, y, z,* and *w* and strengthens the conclusion obtained by the b_N–b_S test. In addition, the *n/s* test also shows that the exterior branch leading to the cat has an *n/s* ratio significantly greater than *N/S* ($P < 0.05$). This significant difference was not found in the previous b_N–b_S test, although the b_N/b_S ratio was high (1.95). The difference between the two test results is probably due to the fact that *s* was estimated to be 1 from the ancestral sequences, but it was estimated to be 2.7 from b_S. At this moment, it is difficult to tell which test is more reliable for this particular branch, and more sequences from carnivores are needed to resolve this discrepancy.

In the b_N–b_S test, we examined positive selection for the branches in which high b_N/b_S ratios were observed. Strictly speaking, this statistical test is inappropriate because it is equivalent to conducting multiple tests. In such a situation, the test is likely to be too liberal. Nevertheless, this will not be a serious problem if additional evidence of positive selection is found, as is shown below.

Comparing the ancestral sequences at the nodes α, β, γ, and δ (fig. 1*A*) and the extant sequence from the deer, we noticed that there were many amino acid substitutions that involved charge changes (fig. 1*B*). Let us call the nonsynonymous nucleotide substitutions resulting in charge changes "radical substitutions" and the rest of the nonsynonymous substitutions "conservative

Table 1
Synonymous ($d_S \times 100$, below diagonal) and Nonsynonymous ($d_N \times 100$, above diagonal) Distances Between IL-2 Sequences of Placental Mammals

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Deer		4.5	4.5	5.5	18.4	24.6	27.1	28.1	27.9	28.3	27.4	27.4	27.9	47.8	43.6
2. Cow	4.5		1.8	2.8	16.7	20.6	26.1	24.3	24.1	24.5	23.7	23.7	24.1	46.0	41.9
3. Sheep	3.7	4.1		1.5	16.8	21.3	25.6	24.6	24.4	24.8	24.0	24.0	24.4	46.1	41.8
4. Goat	3.7	4.1	1.6		18.0	22.6	26.9	25.7	25.4	25.9	25.0	25.0	25.5	46.3	43.2
5. Pig	13.2	14.5	13.2	13.1		15.5	16.4	17.3	17.3	17.7	17.0	17.0	17.4	37.0	30.5
6. Cat	14.7	14.6	15.7	15.6	17.6		10.8	12.4	12.4	12.4	12.1	12.1	12.4	32.4	27.1
7. Dog	16.9	20.3	21.4	21.4	21.4	8.3		14.4	14.4	14.5	14.1	14.1	14.5	34.2	28.8
8. Human	16.6	18.4	17.7	18.2	16.8	12.3	14.8		0.0	1.5	0.9	0.9	1.2	28.8	24.0
9. Gibbon	16.1	17.9	17.2	17.7	15.8	11.4	13.8	0.8		1.5	0.9	0.9	1.2	28.8	23.8
10. C-Mcq.	17.0	18.9	18.2	18.6	16.7	12.2	14.5	3.2	2.4		0.6	0.6	0.9	28.9	24.7
11. R-Mcq.	17.0	18.9	18.2	18.6	16.7	12.2	14.5	3.2	2.4	0.0		0.0	0.3	28.9	24.7
12. P-Mcq.	18.0	19.9	19.2	19.7	17.7	13.2	15.5	4.0	3.2	0.8	0.8		0.3	28.9	24.7
13. Mangabey ..	17.0	18.9	18.1	18.6	16.7	12.2	14.5	3.2	2.4	0.0	0.0	0.8		28.9	24.3
14. Mouse	66.4	70.5	68.1	69.6	61.4	56.7	64.5	63.9	63.6	63.4	63.5	63.5	63.3		9.7
15. Rat	65.4	69.4	68.1	68.5	57.0	52.6	59.9	53.3	52.5	52.3	52.4	52.4	52.2	15.0	

NOTE.—d_N is underlined when it is greater than d_S. C-Mcq. = crab-eating macaque; R-Mcq. = Rhesus macaque; P-Mcq. = pig-tailed macaque.

substitutions'' (Hughes, Ota, and Nei 1990). The total numbers of conservative (*c*) and radical (*r*) nonsynonymous substitutions for branches *x, y, z,* and *w* are 42.3 and 36.0, respectively. The potential numbers of conservative (*C*) and radical (*R*) sites of the sequences are computed to be 209 and 120, respectively. These numbers were computed by the method of Zhang (2000), which is an extension of the method of Hughes, Ota, and Nei (1990) and takes into account the transition/transversion bias. When there is no difference in selection intensity on conservative and radical substitutions, r/c is expected to be equal to R/C. In most genes, $r/c < R/C$, apparently because of a strong purifying selection against radical substitutions in comparison with conservative substitutions (Zhang 2000). In the present case, however, r/c (0.85) is higher than R/C (0.57), and their difference is statistically significant ($P < 0.05$; Fisher's exact test). This result, together with the observation of a significantly higher rate of nonsynonymous than synonymous substitution, strongly suggests that positive selection has promoted amino acid substitutions that result in charge changes in the evolution of IL-2 in branches *x, y, z,* and *w*. In no other branches of the tree was r/c significantly greater than R/C. In fact, the overall r/c ratio (0.58) for all of the branches excluding *x, y, z,* and *w* was almost identical to R/C (0.57), suggesting that for these evolutionary lineages, the selective pressure was nearly the same for conservative and radical amino acid substitutions.

It is worth pointing out that although the rate of charge change substitution was accelerated by positive selection in the four branches examined, the net charge of the IL-2 protein did not change much. This occurred apparently because there were substitutions involving both positively charged and negatively charged amino acids. For example, in branch *z,* there were 16 amino acid substitutions involving charge changes, and 9 of them increased the net charge of the protein, whereas the remaining 7 decreased it. Interestingly, at some sites, amino acid charges changed back and forth. For instance, at site 30, a substitution from the noncharged glutamine to the positively charged lysine occurred in branch *y,* but a subsequent substitution in branch *z* changed it back to a noncharged amino acid (methionine). Similar patterns of charge changes were observed at sites 45, 46, 76, and 98. These observations suggest that positive selection in branches *x, y, z,* and *w* did not increase or decrease the net charge of IL-2, but, rather, changed the amino acid charges at individual sites. This pattern is similar to that of the antigen-binding cleft of major histocompatibility complex molecules, in which positive selection promotes charge profile diversity (Hughes, Ota, and Nei 1990). It is, however, in striking contrast to that of primate eosinophil-cationic protein (ECP), in which the net charge dramatically increased under directional selection during a short period of evolutionary time after a gene duplication event and a novel function subsequently evolved (Zhang, Rosenberg, and Nei 1998).

A number of host defense genes have been shown to evolve rapidly by positive selection in response to ever-changing pathogens (e.g., Tanaka and Nei 1989; Hughes, Ota, and Nei 1990). These genes are all known to interact directly with foreign antigens. In the present case, however, there is no evidence that IL-2 or its receptor directly contacts pathogens. Nevertheless, there is a possibility that they do contact pathogens, as a number of T cell membrane receptors that function in physiological processes are used by pathogens to enter host cells (Cairns and D'Souza 1998). For instance, the chemokine receptor CCR5, which normally binds to macrophage inflammatory proteins 1α, 1β, and RANTES to mediate chemotaxis, is used by the human immunodeficiency virus (HIV) as a coreceptor to enter human T cells (reviewed in Cairns and D'Souza 1998). To examine whether pathogens have been driving the rapid evolution of IL-2, an evolutionary analysis of the IL-2 receptor will be necessary. Interestingly, our preliminary sequence analysis of the alpha chain of the receptor did indicate an elevated rate of evolution in an ancestral lineage of artiodactyls (data not shown) similar to the lineage where IL-2 was found to evolve rapidly. How-

ever, because the number of sequences is limited, that result remains inconclusive. Thus, the selective agent on IL-2 is unclear, but the intriguing possibility of direct contact of IL-2 or its receptor with pathogens is worth pursuing in the future, particularly in the context of frequent use of IL-2 clinically.

Acknowledgments

We thank two anonymous reviewers for their comments on an earlier version of the manuscript. This work was supported by NIH and NSF research grants to M.N.

LITERATURE CITED

CAIRNS, J. S., and M. P. D'SOUZA. 1998. Chemokines and HIV-1 second receptors: the therapeutic connection. Nat. Med. **4**:563–568.

DE JONG, W. W. 1998. Molecules remodel the mammalian tree. Trends Ecol. Evol. **13**:270–275.

FITCH, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20**:406–416.

GAULTON, G. N., and P. WILLIAMSON. 1994. Interleukin-2 and the interleukin-2 receptor complex. Chem. Immunol. **59**: 91–114.

GILLESPIE, J. H. 1989. Lineage effects and the index of dispersion of molecular evolution. Mol. Biol. Evol. **6**:636–647.

GOLDSMITH, M. A., and W. C. GREENE. 1994. Interleukin-2 (IL-2). Pp. 27–30 *in* N. A. NICOLA, ed. Guildbook to cytokines and their receptors. Oxford University Press, Oxford, England.

HAYASAKA, K., K. FUJII, and S. HORAI. 1996. Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. Mol. Biol. Evol. **13**:1044–1053.

HIGGINS, D. G., A. J. BLEASBY, and R. FUCHS. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8**:189–191.

HUGHES, A. L., T. OTA, and M. NEI. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major histocompatibility complex genes in mammals. Mol. Biol. Evol. **7**:491–514.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KUMAR, S., and S. B. HEDGES. 1998. A molecular timescale for vertebrate evolution. Nature **392**:917–920.

MACEY, M.-G., and D. H. JOHNSTON. 1990. Interleukin-2 (IL-2) immunotherapy. Med. Lab. Sci. **47**:220–225.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

OHTA, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40**:56–63.

TANAKA, T., and M. NEI. 1989. Positive Darwinian selection observed at the variable-region genes of immunoglobulins. Mol. Biol. Evol. **6**:447–459.

ZHANG, J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J. Mol. Evol. **50**:56–68.

ZHANG, J., S. KUMAR, and M. NEI. 1997. Small-sample tests of episodic adaptive evolution: A case study of primate lysozyme genes. Mol. Biol. Evol. **14**:1335–1338.

ZHANG, J., and M. NEI. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J. Mol. Evol. **44**(Suppl. 1):S139–S146.

ZHANG, J., H. F. ROSENBERG, and M. NEI. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

ZOU, W., A. FOUSSAT, C. CAPITANT, I. DURAND-GASSELIN, L. BOUCHET, P. GALANAUD, Y. LEVY, and D. EMILIE. 1999. Acute activation of CD8+ T lymphocytes in interleukin-2 treated HIV-infected patients. J. Acquir. Immune Defic. Syndr. **22**:31–38.

NARUYA SAITOU, reviewing editor

Accepted May 29, 2000