J Mol Evol (2005) 61:691–696
DOI: 10.1007/s00239-005-0039-x

JOURNAL OF MOLECULAR EVOLUTION



# Molecular Evolution of *CXCR1*, a G Protein-Coupled Receptor Involved in Signal Transduction of Neutrophils

**Ying Liu,[1,2,3] Su Yang,[1,2] Alice A. Lin,[4] Luca L. Cavalli-Sforza,[4] Bing Su[1,2]**

[1] Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China
[2] Kunming Primate Research Center, Chinese Academy of Sciences, Kunming, Yunnan, China
[3] Graduate School of the Chinese Academy of Sciences, Beijing, China
[4] Department of Genetics, Stanford University, Stanford, CA 94305, USA

Received: 20 February 2005 / Accepted: 14 June 2005 [*Reviewing Editor:* Dr. Manyuan Long]

**Abstract.** Human neutrophils are a type of white blood cell, which forms an early line of defense against bacterial infections. Neutrophils are highly responsive to the chemokine, interleukin-8 (IL-8) due to the abundant distribution of CXCR1, one of the IL-8 receptors on the neutrophil cell surface. As a member of the GPCR family, CXCR1 plays a crucial role in the IL-8 signal transduction pathway in neutrophils. We sequenced the complete coding region of the *CXCR1* gene in worldwide human populations and five representative nonhuman primate species. Our results indicate accelerated protein evolution in the human lineage, which was likely caused by Darwinian positive selection. The sliding window analysis and the codon-based neutrality test identified signatures of positive selection at the N-terminal ligand/receptor recognition domain of human *CXCR1*.

**Key words:** *CXCR1* — Darwinian positive selection — G protein-coupled receptors — Neutrophils — Primates

## Introduction

The mammalian immune system is constantly under selective pressure in fighting against microbial infections. Genes involved in this process are likely subject to Darwinian positive selection by accumulating amino acid substitutions in a relatively rapid pace. The adaptive evolution of the mammalian major histocompatibility complex (MHC) multigene family is a classical example of positive selection acting on genes involved in the immune system (Piontkivska and Nei 2003; Yeager and Hughes 1999).

Neutrophils are a type of white blood cell (leukocyte), which forms an early line of defense against bacterial infections. They are highly responsive to interleukin-8 (IL-8), one of the critical chemokines in the immune system. IL-8 preferentially activates neutrophils to induce chemotactic and cytotoxic responses such as exocytosis of lysosomal enzymes and production of superoxide anion (Brennan 1993; Smith et al. 1992; Walz et al. 1991). Hence, it plays a fundamental regulatory role in native immunity, inflammation, and host defense against infections. IL-8 carries out its function by binding and activating its receptors, CXCR1 and CXCR2, which are located on the neutrophil cell surface.

CXCR1, which regulates leukocyte trafficking in the immune system, is a member of the GPCR (G protein-coupled receptors) family. The *CXCR1* gene is located on human chromosome 2q35 encoding a protein of 350 amino acids. There are two exons

The GenBank accession numbers of sequences reported herein are AY916760–AY916773.

*Correspondence to:* Bing Su, Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; *email:* sub@ail.kiz.ac.cn

in the *CXCR1* gene, but only exon 2 encodes proteins. Furthermore, the *CXCR1* gene was considered one of the candidate genes for several human disorders, such as rheumatoid arthritis, systemic lupus erythematosus, insulin-dependent diabetes mellitus, and juvenile amyotrophic lateral sclerosis (Copeman et al. 1995; Cornelis et al. 1998; Gaffney et al. 1998; Hadano et al. 1999; Shaw et al. 1996). Therefore, it plays a crucial role within the pathway of signal transduction in the immune system.

In this study, we undertook an evolutionary survey of sequence substitution patterns of *CXCR1* in primates to detect potential adaptive evolution. Our data showed that there were adaptive amino acid changes in the primate lineage leading to humans, which were likely caused by Darwinian positive selection and potentially contributing to functional modification of the immune system during human evolution.

## Materials and Methods

### *DNA Samples*

We sequenced a total of 31 human individuals from the major continental populations, including 8 Africans, 7 Europeans, and 8 East Asians (6 Chinese and 2 Cambodians), and 8 Melanesians. In addition, five nonhuman primate species were sequenced, which reflect a 25 million-year history of primate evolution (Goodman et al. 1998). The nonhuman primate panel included three great ape species (one chimpanzee, *Pan troglodytes;* one gorilla, *Gorilla gorilla;* and one orangutan, *Pongo pygmaeus*), one lesser ape species (white-browed gibbon, *Hylobates hoolock*), and one Old World monkey species (rhesus monkey, *Macaca mulatta*). All the DNA samples were from collections in Kunming Cell Bank of CAS, Kunming Blood Center and Stanford University.

### *PCR and Sequencing*

All the human and nonhuman primate samples were sequenced for the full-length coding region of *CXCR1* (1053 bp). Primers for all the primates were designed by aligning the published sequences of human and mouse (*Ensemble* genome browser at http://www.ensemble.org). The primer sequences are listed below.

- pcr_f: 5′ AGGAGGGTAAACAAGAATCAGG 3′
- pcr_r: 5′ GTTCAACGGGAATGATGGTGC 3′
- seq_f: 5′ ATAGGCTGAGAGGAGGCAGTTG 3′
- seq_r: 5′ CCTTAAACAGTGTACGCAGGGTGA 3′
- seq_r1: 5′ GCCAGGTAACGGTCCAC 3′
- seq_r2: 5′ GCAGGACCAGGTTGTAGGG 3′
- seq_r3: 5′ GTTTTCAGAGGTTGGAAGAGAC 3′

Sequencing was performed in both directions with forward and reverse primers using the BigDye terminator sequencing kit on an ABI 3100 automated sequencer.

### *Data Analysis*

DNA sequences were aligned using DNASTAR (DNASTAR, INC.) and checked manually. The MEGA3.0 program was used in the phylogenetic analysis and calculation of nonsynonymous and synonymous substitution rates ($K_a$ and $K_s$) following the method of Pamilo–Bianchi–Li (Kumar et al. 2001; Li 1993; Pamilo and Bianchi 1993). We assessed the signatures of positive selection of *CXCR1* in diverse primate lineages by codon-based maximum likelihood analysis using the Codeml program in the PAML package (Yang 1998). We compared the likelihood fit of two evolutionary models: the one-ratio model (M0) and the different ratio model (M2). We calculated the probability that the two models should differ in log likelihood (Yang 1998). For the sliding-window analysis, we used the DnaSP4.0 program (Rozas et al. 2003) and the $K_a/K_s$ ratios were calculated following Nei and Gojobori (1986), which was imbedded in the program. The window length is 300 bp (100 codons) with a sliding step of 3 bp (1 codon). The ancestral sequences (internal nodes) in the phylogenetic tree were inferred using the PAML package based on Yang's (1998) method. The one-tailed *Z*-test and Fisher's exact test were used to detect deviation of the $K_a/K_s$ ratios from neutrality ($K_a/K_s = 1$) (Kumar et al. 2001). The MacDonald–Kreitman test was also used to detect positive selection (Fay and Wu 2000; McDonald and Kreitman 1991). The RbDe program (http://icbtools.med.cornell.edu/viseur/viseur.html) was used to reconstruct the two-dimensional protein structure of *CXCR1*.

## Results and Discussion

In the 1053-bp coding region of *CXCR1*, we observed 118 nucleotide variant sites when the human sequences were aligned with the five nonhuman primate species including Old World monkeys, lesser apes, and great apes. A total of 60 amino acid variant sites (17.1%; 60/350) were observed (Fig. 1). To compare the patterns of sequence substitutions among different primate lineages, we calculated the $K_a/K_s$ ratios for each branch on the primate phylogenetic tree following the Pamilo–Bianchi–Li method (Li 1993; Pamilo and Bianchi 1993) (Fig. 2). We noted that, for all lineages except human and chimpanzee, the $K_a/K_s$ values are $< 1$, suggesting functional constraint of *CXCR1* in these primate lineages. However, for the human and chimpanzee lineages, the $K_a/K_s$ ratios are 2.67 ($p = 0.09$) and 2.2 ($p = 0.13$), respectively (one-tailed $Z$ test), which is suggestive of accelerated protein evolution in these two species, although the likelihood ratio test did not detect $K_a/K_s$ ratio differences between these two lineages and other primate lineages. The likelihood ratio under model M2 is –2156.11, and the likelihood ratio under model M0 is –2156.86 ($p > 0.05$) (Yang 1998).

The observed high $K_a/K_s$ ratios in the human and chimpanzee lineages could be either due to Darwinian positive selection as a result of adaptive evolution or due to relaxation of functional constraint of *CXCR1*. To further test if positive selection was acting on the human lineage, we conducted the MacDonald–Kreitman neutrality test, which was developed to compare between-species nonsynonymous/synonymous substitutions with within-species nonsynonymous /synonymous polymorphisms and detect whether positive selection contribute to the observed

```
                                                                      50
Human        MSNITDPQMW DF-DDLNFTG MPPADEDYSP CMLETETLNK YVVIIAYALV
Chimpanzee   ....I..... .Y........ ...T..G... .R........ .....T...A
Gorilla      ....I..... .......... ...I...... .R........ .....T...A
Orangutan    .......... .YDG.P.... ...I....R. .R........ ....VT....
Gibbon       .W......G. .YDG..Y... ...I...F.. .K........ .....T....
Macaca       ...A.....G .DDY...... ...T...... .R...QS... ....VT....

                                                                     100
Human        FLLSLLGNSL VMLVILYSRV GRSVTDVYLL NLALADLLFA LTLPIWAASK
Chimpanzee   .......... .......... ...I...... .......... ..........
Gorilla      .......... .........G .......... .......... ..........
Orangutan    .......... .......... .......... .......... .......V..
Gibbon       .......... .......... ......I... .......... ..........
Macaca       .......... .......R.. .......... ...M...... ..........

                                                                     150
Human        VNGWIFGTFL CKVVSLLKEV NFYSGILLLA CISVDRYLAI VHATRTLTQK
Chimpanzee   .T......L. .......... .......... .......... ..........
Gorilla      .......... .......... .......... .......... ..........
Orangutan    ........L. .......... .......... .......... ..........
Gibbon       .......... .......... .......... .......... ..........
Macaca       .......... .......... .......... .......... .......I..

                                                                     200
Human        RHLVKFVCLG CWGLSMNLSL PFFLFRQAYH PNNSSPVCYE VLGNDTAKWR
Chimpanzee   .......... ......I... .......... .......... ....N.....
Gorilla      .......... ......I... .......... .......... ..........
Orangutan    .........S ......I... .......... .K........ ..........
Gibbon       ........V. ......I... .......... .......... ..........
Macaca       ..S......S .....VI... .......... ....T..... ..........

                                                                     250
Human        MVLRILPHTF GFIVPLFVML FCYGFTLRTL FKAHMGQKHR AMRVIFAVVL
Chimpanzee   .......... .......... .......H.. .......... ...I......
Gorilla      .......... .......... .......... .......... ..........
Orangutan    .......... .......... .....A.C.. .......... ..........
Gibbon       .......... ......L... .......H.. .......... ....V.....
Macaca       .......... ..TL..LI.. .......H.. ....I..... ..........

                                                                     300
Human        IFLLCWLPYN LVLLADTLMR TQVIQESCER RNNIGRALDA TEILGFLHSC
Chimpanzee   .......... .......... .......... .......... ..........
Gorilla      .......... .......... .......... .....L.... ..........
Orangutan    .......... .......... ..L.K..... ..D..W.... ..........
Gibbon       .......... ...FT..... ..L.K..... .KD.SK..E. ...F......
Macaca       .......... .......... .HL.K..... ..D....... ..........

                                                                     350
Human        LNPIIYAFIG QNFRHGFLKI LAMHGLVSKE FLARHRVTSY TSSSVNVSSN
Chimpanzee   .......... .......... .......... .......... ..........
Gorilla      .......... .......... .......... .......... S.........
Orangutan    .......... .......... .......... .....H.... ..........
Gibbon       .......... .......... .......... .....H.... ..........
Macaca       .......... .......... ..T....... .....H.... ..........
```

**Fig. 1.** The amino acid sequence alignment of CXCR1 in humans and nonhuman primates.

substitution pattern (McDonald and Kreitman 1991). We observed four polymorphic sites (one synonymous and three nonsynonymous) in the coding region of *CXCR1* in the human populations (31 worldwide samples in total). The between-species (human *vs.* chimpanzee) nonsynonymous/synonymous substitution ratio (14/3) is slightly higher than the nonysynonymous/synonymous polymorphism within human populations (3/1) according to the MacDonald–Kreitman test, but not statistically significant ($p > 0.05$, two-tailed Fisher's exact test). Table 1 lists the allele frequencies of the four polymorphic sites

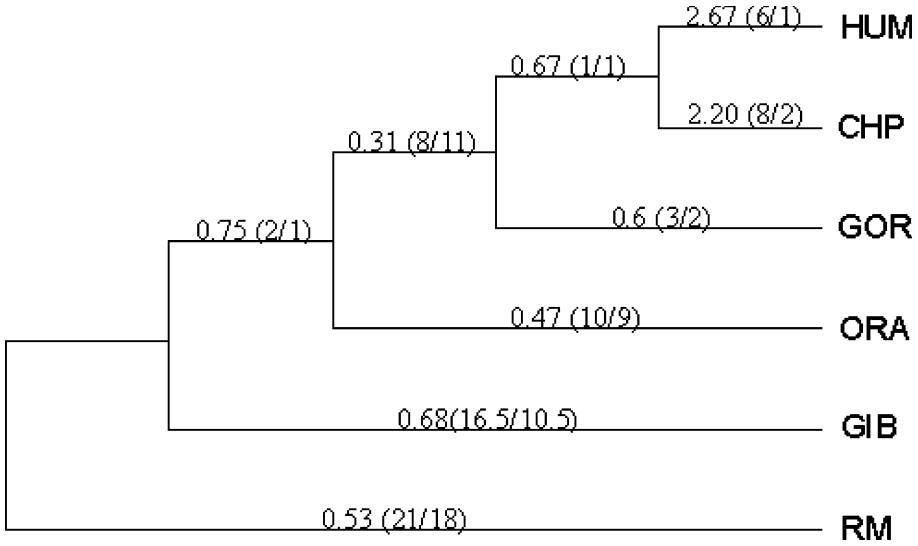


**Fig. 2.** The $K_a/K_s$ ratios of different evolutionary lineages in primates. The complete coding sequences were used for $K_a/K_s$ calculation with the Pamilo–Bianchi–Li method (Pamilo and Bianchi 1993; Li 1993). The consensus sequence of humans was used in the analysis. The numbers of nonsynonymous and synonymous substitutions are shown in parentheses. HUM, human; CHP, chimpanzee; GOR, gorilla; ORA, orangutan; GIB, white-browed gibbon; RM, rhesus monkey.

**Table 1.** Frequency distribution of sequences polymorphism of *CXCR1* in human populations

| | Allele frequency | | | | |
|---|---|---|---|---|---|
| Polymorphism site | Caucasian ($n=14$) | African ($n=16$) | Asian ($n=16$) | Melanesian ($n=16$) | Total ($n=62$) |
| **G/T** ($^{Arg}31^{Met}$) | 0.93 | 0.75 | 0.94 | 1 | 0.91 |
| **C/T** ($96^{synonymous}$) | 0 | 0.125 | 0 | 0 | 0.032 |
| **G/C** ($^{Ser}276^{Thr}$) | 0 | 0.187 | 0.062 | 0 | 0.064 |
| **C/T** ($^{Arg}335^{Cys}$) | 0 | 0.062 | 0 | 0.062 | 0.032 |

*Note*. The bold alleles refer to the derived variants and the sample size in parentheses is the number of chromosomes.

observed in the human populations, in which no between-population allele frequency difference was observed. Hence, the result of the MacDonald–Kreitman test could not reject the null hypothesis of neutral evolution in the human lineage.

Because natural selection could have different selective pressures on different domains of a gene, we performed a sliding window analysis by comparing the $K_a/K_s$ ratios between human and chimpanzee sequences across the coding region of *CXCR1*. Figure 3 shows the result of the sliding window analysis, in which high $K_a/K_s$ ratios were observed in the N-terminal region of *CXCR1*. We also conducted a codon-based neutrality test (Yang 1998) to detect individual amino acid sites under positive selection. A total of five sites showed the signature of positive selection ($p < 0.05$ for sites 12 and 14; $p < 0.01$ for sites 24, 228, and 286) when tested using the one rate model M0 in *PAML* (Yang 1998). Three of the five sites (site 12, 14, 24) are located in the N-terminal region of *CXCR1*, with two of them (sites 12, 24) having nonsynonymous substitutions in the human lineage. A previous study indicated that the protein sequence of *CXCR1* is conserved in the transmembrane and intracellular domains but less conserved in the extracellular domains (Rajagopalan and Rajarathnam 2004). The binding of *IL-8* involves close interactions between the *CXCR1* N-domain and *IL-8* N-loop residues and between *CXCR1* exoloops and the transmembrane and *IL-8* N-terminal residues (Fernando et al. 2004; Rajagopalan and Rajarathnam 2004). Both the sliding window analysis and the codon-based neutrality test tend to favor positive selection acting on the N-terminal domain of human *CXCR1*, the critical region for ligand/receptor signaling for neutrophils. This explains the insignificance of the MacDonald–Kreitman test because the signature of positive selection was possibly averaged out when the complete coding region was used. In addition, as one of the critical players in the immune system, *CXCR1* is the receptor that specifically binds IL-8 (Fernando et al. 2004). Therefore the relatively rapid protein evolution of human *CXCR1* is not likely the result of relaxation of functional constraint. Consequently, adaptive evolution is more likely the cause of the observed high $K_a/K_s$ ratio in the human lineage although relaxation of functional constraint cannot be completely ruled out. Whether the high $K_a/K_s$ ratio in the chimpanzee lineage was also caused by positive selection is yet to be tested.

To see whether the adaptive evolution of *CXCR1* in humans had an effect on its ligand, IL-8, we compared the sequences of *IL-8* among human, chimpanzee, and rhesus macaque (GenBank

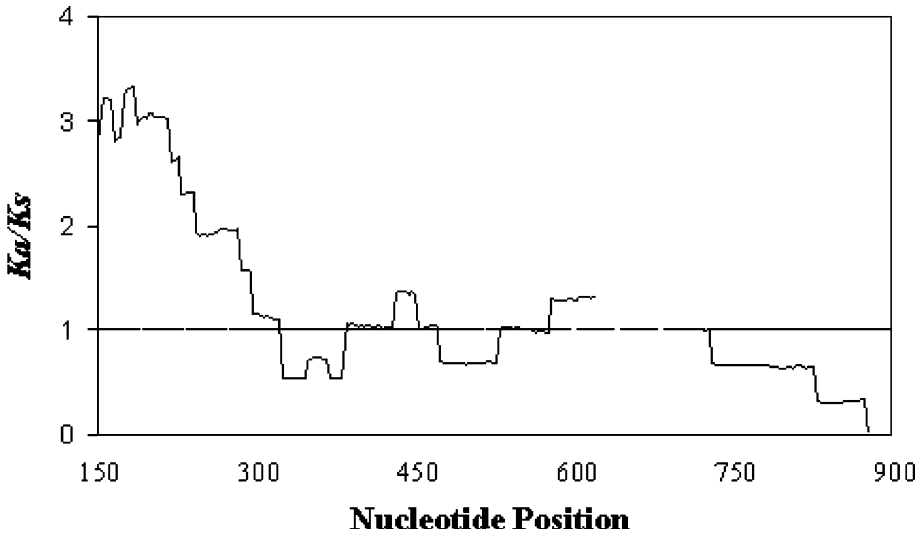


**Fig. 3.** The sliding window analysis of $K_a/K_s$ ratios between human and chimpanzee. The DnaSP4.0 program was used in generating the $K_a/K_s$ ratio plot (Rozas et al. 2003). The window length is 100 codons and the step size is one codon. The nucleotide positions shown are the midpoints of the windows starting from site 150.

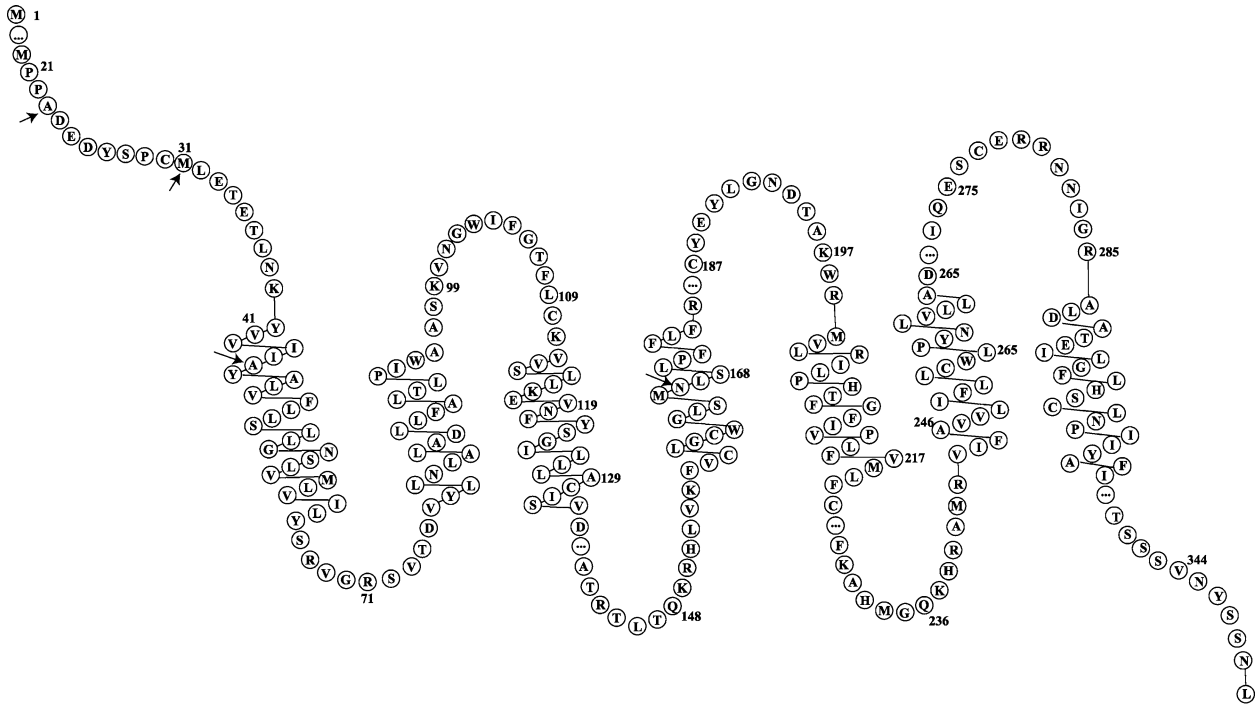

**Fig. 4.** The secondary structure of the human CXCR1. The snake-like diagram was reconstructed by the RbDe program (http://icbtools.med.cornell.edu/viseur/viseur.html). The arrows indicate the four human-specific mutations. The numbers indicate the positions of the amino acid sequence. The open circles are the omitted amino acid sites.

accession numbers NM_000584, XM_526587, and S78555). We found that IL-8 is extremely conserved. Human and chimpanzee share the same protein sequence (data not shown). Consequently, the adaptive evolution of *CXCR1* in humans is not likely driven by the coevolution of ligand and receptor but, rather, by unilateral adaptive changes of *CXCR1*. As the N-terminal and transmembrane of *CXCR1* are involved in ligand/receptor binding and recognition, the four human-specific amino acid substitutions could potentially modify the ligand/receptor interaction, e.g., the specificity and/or affinity (Fig. 4). Because neutrophils are a type of leukocyte, which forms an early line of defense against bacterial infections, the adaptive change of human *CXCR1* was likely caused by the selective pressure of improving the innate immunity.

*Acknowledgments.* We are thankful for the technical help of Xiao-na Fan and Yi-chuan Yu. This study was supported by grants from the Chinese Academy of Sciences (KSCX2-SW-121), the National Natural Science Foundation of China (30370755, 30440018), the National 973 Project (2006CB701506), and the Natural Science Foundation of Yunnan Province of China.

## References


Brennan FM (1993) Cytokines and cytokine receptors: From cloning to the clinic. Keystone symposium on molecular and cellular biology. Int J Exp Pathol 74:519–524

Copeman JB, Cucca F, Hearne CM, et al. (1995) Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (IDDM7) to chromosome 2q31-q33. Nat Genet 9:80–85

Cornelis F, Faure S, Martinez M, et al. (1998) New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. Proc Natl Acad Sci USA 95:10746–10750

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Fernando H, Chin C, Rosgen J, Rajarathnam K (2004) Dimer dissociation is essential for interleukin-8 (IL-8) binding to CXCR1 receptor. J Biol Chem 279:36175–36178

Gaffney PM, Kearns GM, Shark KB, et al. (1998) A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. Proc Natl Acad Sci USA 95:14875–14879

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9:585–598

Hadano S, Nichol K, Brinkman RR, Nasir J, Martindale D, Koop BF, Nicholson DW, Scherer SW, Ikeda JE, Hayden MR (1999) A yeast artificial chromosome–based physical map of the juvenile amyotrophic lateral sclerosis (ALS2) critical region on human chromosome 2q33-q34. Genomics 55:106–112

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17:1244–1245

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: Rates and interdependence between the genes. Mol Biol Evol 10:271–281

Piontkivska H, Nei M (2003) Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol Biol Evol 20:601–609

Rajagopalan L, Rajarathnam K (2004) Ligand selectivity and affinity of chemokine receptor CXCR1. Role of N-terminal domain. J Biol Chem 279:30000–30008

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Shaw MA, Clayton D, Atkinson SE, Williams H, Miller N, Sibthorpe D, Blackwell JM (1996) Linkage of rheumatoid arthritis to the candidate gene NRAMP1 on 2q35. J Med Genet 33:672–677

Smith RJ, Sam LM, Leach KL, Justen JM (1992) Postreceptor events associated with human neutrophil activation by interleukin-8. J Leukoc Biol 52:17–26

Walz A, Zwahlen R, Baggiolini M (1991) Formation and biological properties of neutrophil activating peptide 2 (NAP-2). Adv Exp Med Biol 305:39–46

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

Yeager M, Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol Rev 167:45–58