

Tertiary Windowing to Detect Positive Diversifying Selection

Ann-Charlotte Berglund,^{1,2} Björn Wallner,² Arne Elofsson,² David A. Liberles^{1,2}

¹ Computational Biology Unit, BCCS, University of Bergen, 5020 Bergen, Norway

² Stockholm Bioinformatics Center, SCFAB, Stockholm University, 10691 Stockholm, Sweden

Received: 20 July 2004 / Accepted: 20 October 2004 [Reviewing Editor: Dr. Rasmus Nielsen]

Abstract. As a protein-encoding gene evolves, different selective pressures act on the gene temporally and spatially. An examination of the ratio of non-synonymous-to-synonymous nucleotide substitution rate ratios (K_a/K_s) has proven to be a valuable method to examine selective pressures on protein encoding genes, including detecting positive diversifying selection. To gain power over averaging all sites in a gene together, examination of sites in primary sequence windows has frequently been employed. However, selection acts on folded proteins and sites that are close in tertiary space may not be close in primary sequence. A new method for the examination of K_a/K_s ratios based upon windows in tertiary structure is introduced and applied to the leptin gene family in mammals. Tertiary sequence windowing detects new sites under positive diversifying selection and detects positive diversifying selection with a more significant signal along various branches of the leptin gene family tree.

Key words: Positive diversifying selection — Substitution rate — Protein structure — Leptin

Introduction

Among the most successful methods for detecting adaptive evolution in protein-coding genes are

methods based on comparing the nonsynonymous substitution rate, K_a , to the synonymous substitution rate, K_s . Under the theory of neutrality the nonsynonymous-to-synonymous substitution rate ratio, K_a/K_s , should equal one. A ratio significantly greater than one is evidence of positive diversifying selection, whereas a ratio significantly smaller than one is evidence for purifying selection. A large number of genes have already been identified as being under positive selective pressure at various points in their evolutionary history using K_a/K_s (Liberles et al. 2001).

There exists a number of different methods for estimating the K_a/K_s rate ratio. The earliest methods estimate K_a and K_s by comparing pairs of homologous sequences in extant species (Li et al. 1985; Nei and Gojobori 1986; Li 1993; Pamilo and Bianchi 1993; Goldman and Yang 1994; Comeron 1995; Ina 1995). This approach leads to averaging over time as well as over all codons in the sequences. However, we know that many protein-coding genes are under purifying selection for most of their evolutionary time (see Siltberg and Liberles 2002). Further, since proteins must fold, most of their sites are under purifying selection, while the behavior of a protein can change even if only a few amino acid positions have been substituted (Golding and Dean 1998). Consequently, for many proteins these methods will indicate negative selection, even though positive diversifying selection may have acted on the protein during a short evolutionary time or at a few sites. The first of these problems is addressed by considering the phylogeny and the reconstructed ancestral sequences of the protein-coding gene family of

interest (Messier and Stewart 1997; Yang 1998; Liberles 2001). These methods make it possible to detect adaptive evolution along branches in a phylogeny. To take care of the second problem, several approaches for detecting sites under positive diversifying selection have been developed. Hughes and Nei (1989) divided the sites into different categories based on *a priori* knowledge about functional domains of the proteins under investigation. However, this type of *a priori* knowledge might not always be available, and also we might conduct a study with the aim to detect potentially functionally evolving sites. Approaches that divide the sequences into different categories without taking into account the functional domains of the proteins include the covarion method, where the sequences in a phylogeny are divided into variable and invariable sites (Siltberg and Liberles 2002), and the maximum-likelihood methods where the K_a/K_s rate ratio is allowed to vary over sites in the sequences (Nielsen and Yang 1998; Yang et al. 2000).

In a large-scale study, Endo et al. (1996) investigated 3595 protein-coding genes by sliding a window of 20 codons along multiple sequence alignments. Their method was later improved by Fares et al. (2002) using the method by Suzuki and Gojobori (1999) and estimating an optimal window size based on sequence divergence. Here we propose another windowing method based on the tertiary structure of the protein. The advantage of this approach to the primary windowing approaches is that selection is frequently dictated by specific binding pockets as well as interacting residues, and some of these residues are distantly located in primary sequence, but close in the tertiary structure of the protein. The tertiary windowing method is compared to the primary windowing method, the most commonly used method for subdividing functional sites during the analysis of K_a/K_s .

Materials and Methods

Tertiary Windowing Method

After the sequences for the protein-coding genes were retrieved and aligned, the phylogeny was constructed using MrBayes (Huelsenbeck and Ronquist 2001). The reconstruction of the ancestral sequences was done using the maximum-parsimony method (Fitch 1971). In this study we were only considering closely related sequences, where the results from maximum-parsimony and maximum-likelihood methods give similar results (Zhang and Nei 1997). The protein structures were retrieved from the Protein Data Bank (PDB) (Berman et al. 2000), and since the positions for all residues might not be resolved, the sequence from PDB had to be aligned with the sequences in the original multiple alignment.

The tertiary windows were constructed by letting each codon act as the center of a window, and all codons within a prespecified distance in Ångströms from the center were added to the window.

Consequently, the number of codons in each window will depend on the number of close neighbors to the center codon. For calculating K_a and K_s we used the PBL method (Li et al. 1985; Li 1993; Pamilo and Bianchi 1993). The average number of synonymous and nonsynonymous changes, and the average number of synonymous and nonsynonymous sites were then added together for each of the codons in the same tertiary window, and the K_a/K_s rate ratio and the difference $K_a - K_s$ were calculated for each window.

Hypothesis Testing

The null hypothesis for testing our estimated K_a/K_s values is near-neutral evolution, that is, only negative or no selection has acted on the protein as it has evolved. To test if the windows that have a K_a/K_s ratio greater than one are significant, the probability distribution of the difference, $K_a - K_s$, is estimated. The reason why the difference is used instead of the ratio is that the difference is always a real number, even when $K_s = 0$. The significance level we have chosen is 0.05. To correct for multiple testing, Bonferroni correction and the False Discovery Rate are utilized (Benjamini and Hochberg 1995).

A problem in hypothesis testing when using a windowing method where the windows overlap is that many of the tests are redundant, as we are testing the same region more than once. To overcome this problem, we only consider windows that overlap less than a determined threshold. The set of windows to test is deduced by the following procedure. First, the windows with $K_a/K_s > 0$ are listed in ascending order. The top window is included to test. Then eliminate windows that overlap with the selected window above the threshold and repeat until a decision has been made on each window.

To estimate the probability distribution 400 sets of sequences were simulated from the original data set under the model of neutral evolution using the program EVOLVER (Yang 2002).

Leptin

The leptin protein is central to the regulation of energy metabolism in mammals and has been used as a test case for positive diversifying selection in previous studies (see, e.g., Benner et al. 1998). We used the same phylogeny (Fig. 2) for the leptin-coding gene (146 residues plus a signal peptide that consists of 21 amino acid residues), as used in several previous studies (Benner et al. 2002; Siltberg and Liberles 2002; Gaucher et al. 2003). The crystal structure of mature leptin has been resolved by Zhang et al. (1997) and is depicted in Fig. 1. In previous analyses, between two and four lineages have been detected with positive diversifying selection. Three lineages were detected in Liberles (2001), the lineages leading to Hominidae, orangutan, and rhesus monkey. In Benner et al. (2002), the two lineages leading to Hominidae and orangutan showed evidence of positive diversifying selection and in Siltberg and Liberles (2002), the four lineages leading to Hominidae, orangutan, gorilla, and rhesus monkey showed evidence of positive diversifying selection. Therefore, we decided to investigate if we could detect sites evolving under positive diversifying selection along the lineages present in this subtree. This resulted in testing six lineages for sites evolving under positive diversifying selection: the lineages leading to Hominidae, rhesus monkey, orangutan, the last common ancestor of human, gorilla, and chimpanzee, gorilla, and the last common ancestor of chimpanzee and human. The corresponding tree showing these relationships is shown in Fig. 2.

Windows with $K_a/K_s > 0$ that did not overlap beyond thresholds of 10, 33, and 80% were considered and results were tested for consistency.

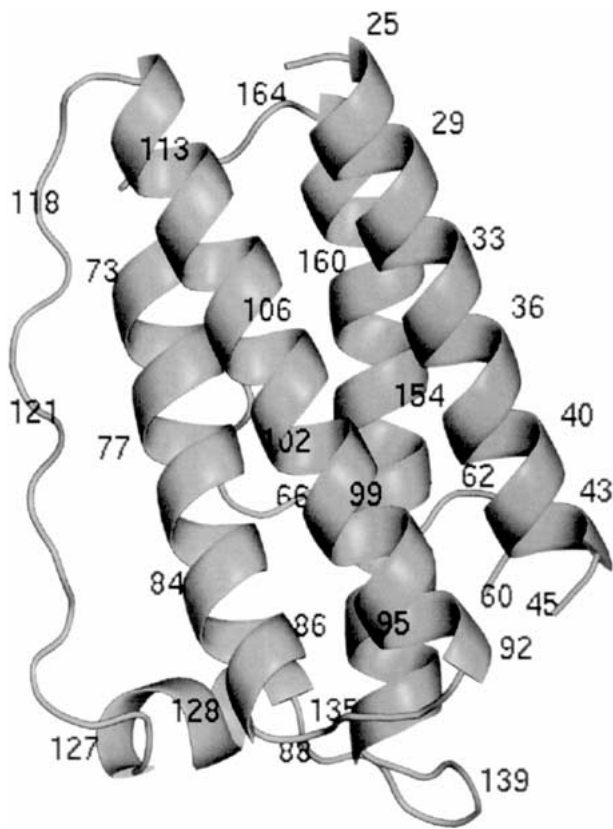


Fig. 1. The three-dimensional structure of leptin (Zhang et al. 1997) is shown with corresponding residue numbers indicated.

Results

In order to estimate the distribution of $K_a - K_s$ in different windows under the null hypothesis (neutral evolution), 400 data sets were simulated from the leptin gene family using the program EVOLVER (Yang 2002). The sequences evolved down the given phylogeny assuming neutral evolution with 52% of sites evolving neutrally ($\omega = 1$) and 48% of sites under conservative selective pressure ($\omega = 0.2$). This is a conservative estimate of the distribution, which is much more neutral than has been observed in most real gene families (see, e.g., Liberles et al. 2001). Both the primary and the tertiary windowing analyses were then used to analyze the simulated data sets.

For the primary windowing two different window sizes were used—three codons and five codons—and two different thresholds—0.10 and 0.33. While larger windows have been traditionally used with this method, the smaller windows showed greater power in this analysis. The results are summarized in Tables 1 and 2. For a window size of three codons, only one window showed a significantly higher K_a value compared to K_s , and that was the window consisting of residues 87–89 along the lineage leading to rhesus (*Macaca mulatta*). The FDR approach was used for a threshold of overlaps of 30%, and for $\alpha = 0.33$ six

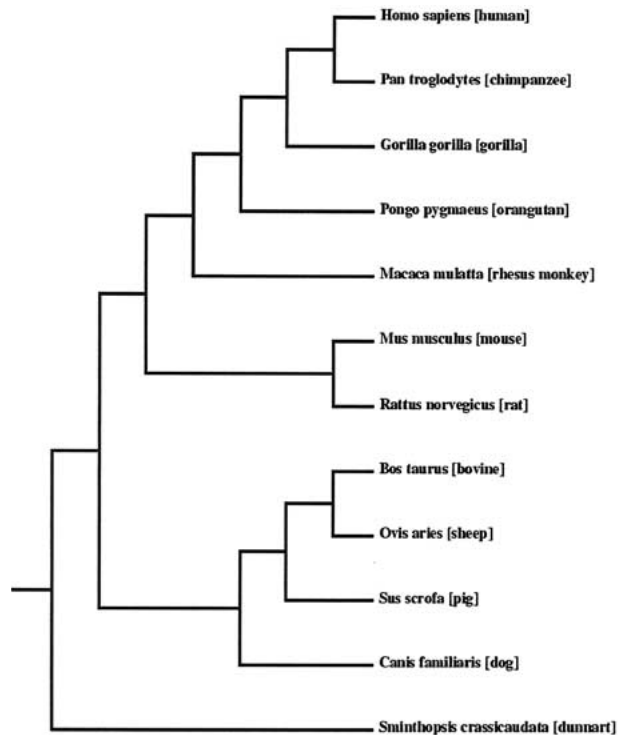


Fig. 2. A phylogenetic tree of the leptin gene family is depicted. This tree was calculated using MrBayes (Huelsenbeck and Ronquist 2001) using a previously calculated gene family with multiple sequence alignment (Siltberg and Liberles 2002). This tree was used for all analyses of leptin. Hominidae (from the text) refers to the clade containing human, chimpanzee, gorilla, and orangutan.

Table 1. Primary windowing was performed with a window size of three codons

Branch	Adj p -value	Residues	No. codons	$K_a - K_s$
A				
To <i>Macaca mulatta</i>	0.017	87–89	3	0.8047
B				
To <i>Macaca mulatta</i>	0.043	87–89	3	0.8047
To <i>Macaca mulatta</i>	0.153	86–88	3	0.5493
To <i>Macaca mulatta</i>	0.153	29–31	3	0.5493
To <i>Macaca mulatta</i>	0.153	27–29	3	0.5493
To <i>Pongo pygmaeus</i>	0.165	38–40	3	0.5493
To <i>Pongo pygmaeus</i>	0.165	37–39	3	0.5493

Note. Windows with $K_a - K_s > 0$. The adjusted p -values are the Bonferroni-corrected p -values. (A) Threshold of 10%. (B) Threshold of 33%. With a threshold of 33% the number of tests has increased from 21 to 53. Also, the number of redundant tests has increased, and thus for a threshold of 33% we are considering the false discovery rate that for $\alpha = 0.33$ gives six windows.

windows were returned. Of these six windows, three regions could be detected: the region from amino acid 86 to 89, the region from amino acid 27 to 31, and the region from amino acid 37 to 40. For a window size of five codons, the Bonferroni-adjusted p -values were all above the chosen significance value 0.05. With a

Table 2. Primary windowing was performed with a window size of five codons

Branch	Adj <i>p</i> -value	Residues	No. codons	$K_a - K_s$
A				
To <i>Pongo pygmaeus</i>	0.085	38–42	5	0.2939
To <i>Macaca mulatta</i>	0.097	87–91	5	0.2939
To <i>Macaca mulatta</i>	0.162	27–31	5	0.2554
B				
To <i>Pongo pygmaeus</i>	0.121	38–42	5	0.2939
To <i>Macaca mulatta</i>	0.138	87–91	5	0.2939
To <i>Macaca mulatta</i>	0.232	28–32	5	0.2554
To <i>Macaca mulatta</i>	0.232	25–29	5	0.2554

Note. Windows with $K_a - K_s > 0$. The adjusted *p*-values are the Bonferroni-corrected *p*-values. (A) Threshold of 10%. The false discovery rate for $\alpha = 0.20$ gives three windows. (B) Threshold of 33%. With a threshold of 33% the number of tests has increased from 21 to 53. The false discovery rate for $\alpha = 0.25$ gives four windows.

Table 3. Tertiary windowing was performed with a window size of 4.5 Å

Branch	Adj <i>p</i> -value	Residues	No. codons	$K_a - K_s$
A				
To <i>Macaca mulatta</i>	0.009	86–88, 98	4	0.3466
To <i>Macaca mulatta</i>	0.009	25–27, 29	4	0.3466
To hominoid	0.019	138–140	3	0.3989
B				
To <i>Macaca mulatta</i>	0.013	86–88, 98	4	0.3466
To <i>Macaca mulatta</i>	0.013	25–27, 29	4	0.3466
To hominoid	0.029	138–140	3	0.3989
C				
To <i>Macaca mulatta</i>	0.045	86–88, 98	4	0.3466
To <i>Macaca mulatta</i>	0.045	25–27, 29	4	0.3466
To <i>Macaca mulatta</i>	0.090	85, 87–90, 128	6	0.2554
To hominoid	0.101	138–140	3	0.3989

Note. Windows with $K_a - K_s > 0$. The adjusted *p*-values are the Bonferroni-corrected *p*-values. (A) Threshold of 10%. (B) Threshold of 33%. (C) Threshold of 80%. With a threshold of 80% the number of tests has increased from 19 to 99. Also, the number of redundant tests has increased, and thus for a threshold of 80% we are considering the false discovery rate that for $\alpha = 0.25$ gives four windows.

threshold of 10%, using FDR with $\alpha = 0.20$ gave three windows along two lineages: the lineage leading to orangutan (*Pongo pygmaeus*) and the lineage leading to *Macaca mulatta*. For a threshold of 33%, FDR with $\alpha = 0.25$ gave four windows along the same lineages as the 10% threshold.

For the tertiary windowing method, three different window sizes were considered, with spheres of radius 4.5, 6.5, and 8.0 Å. The results are summarized in Tables 3–5. For these three choices of window sizes we find significant windows along two lineages, the

Table 4. Tertiary windowing was performed with a window size of 6.5 Å

Branch	Adj <i>p</i> -value	Residues	No. codons	$K_a - K_s$
A				
To hominoid	0.013	137–140	4	0.2786
To <i>Macaca mulatta</i>	0.022	85–90, 128	7	0.2027
To <i>Macaca mulatta</i>	0.022	24–29	6	0.2027
B				
To hominoid	0.021	137–140	4	0.2786
To <i>Macaca mulatta</i>	0.035	85–90, 128	7	0.2027
To <i>Macaca mulatta</i>	0.035	24–29	6	0.2027
C				
To hominoid	0.0620	137–140	4	0.2786
To <i>Macaca mulatta</i>	0.1034	85–90, 128	7	0.2027
To <i>Macaca mulatta</i>	0.1034	24–29	6	0.2027

Note. Windows with $K_a - K_s > 0$. The adjusted *p*-values are the Bonferroni-corrected *p*-values. (A) Threshold of 10%. (B) Threshold of 33%. (C) Threshold of 80%. With a threshold of 80% the number of tests has increased from 19 to 91. Also, the number of redundant tests has increased, and thus for a threshold of 80% we are considering the false discovery rate that for $\alpha = 0.33$ gives three windows.

one leading to *Macaca mulatta* and the one leading to hominoid. The significant lineages are consistent over the three window sizes and also over the threshold levels. For a window size of 8.0 Å only one window is detected. Increasing the window size further might capture new regions not captured by the smaller window sizes chosen here.

Along the lineage leading to *Macaca mulatta* there were overlaps of residues present in the primary and tertiary windows. The tertiary windowing method also showed a significant window along the lineage leading to hominoid that consisted of three codons, which are adjacent in the sequence, but this window is not significant using the primary windowing technique. The lineage leading to orangutan detected with primary windowing was only detected as significantly positive with tertiary windowing using a window size of 8.0 Å and an overlap threshold of 33% or less.

It should be noted that depending on the role and position of the sites that have undergone positive diversifying selection, different window sizes will be optimal for detecting different regions. This is even more true when comparing different protein folds with different compactnesses, surface area to volume ratios, topologies, shapes, and even functions. This holds for both the primary and the tertiary windowing methods.

Discussion

On the basis of predicted biological cleavage sites of human leptin, three peptide fragments are obtained,

Table 5. Tertiary windowing was performed with a window size of 8.0 Å

Branch	Adj <i>p</i> -value	Residues	No. codons	$K_a - K_s$
A				
To <i>Macaca mulatta</i>	0.017	24–31	8	0.1438
B				
To <i>Macaca mulatta</i>	0.032	24–31	8	0.1438
To <i>Macaca mulatta</i>	0.109	137–141, 144	6	0.1769
To <i>Macaca mulatta</i>	0.148	81, 84–89, 125–132, 135	16	0.1024
To <i>Macaca mulatta</i>	0.162	86–95, 98, 143	12	0.1003
To <i>Macaca mulatta</i>	0.264	29, 32–40, 154, 155, 158	13	0.0912
To <i>Pongo pygmaeus</i>	0.303	36–44, 151	10	0.1182
C				
To <i>Macaca mulatta</i>	0.117	24–31	8	0.1438
To <i>Macaca mulatta</i>	0.337	24–29, 32, 111, 162, 164	10	0.1182
To hominoid	0.402	137–141, 144	6	0.1769
To <i>Macaca mulatta</i>	0.440	29–37, 107, 158	11	0.1057
To <i>Macaca mulatta</i>	0.544	81, 84–89, 125–132, 135	16	0.1024
To <i>Macaca mulatta</i>	0.596	25, 28, 29, 32, 158–165	12	0.1003
To <i>Macaca mulatta</i>	0.596	86–95, 98, 143	12	0.1003
To <i>Macaca mulatta</i>	0.596	83–90, 94, 95, 98, 128	12	0.1003
To <i>Macaca mulatta</i>	0.596	26–34, 104, 107	11	0.1003

Note. Windows with $K_a - K_s$. The adjusted *p*-values are the Bonferroni-corrected *p*-values. (A) Threshold of 10%. (B) Threshold of 33%. Using FDR with $\alpha = 0.30$ gives nine windows (31 tests), of which one is significant at the 0.05 level. (C) Threshold of 80%. With a threshold of 80% the number of tests has increased from 27 to 114. Also, the number of redundant tests has increased, and thus for a threshold of 80% we are considering the false discovery rate that for $\alpha = 0.50$ gives nine windows.

consisting of amino acid residues 22–56, amino acid residues 57–92, and amino acid residues 116–167. Amino acids 1–21 consist of the signal peptide (Samson et al. 2002). Interestingly, feeding mice peptides containing amino acid residues 106–120, 116–130, and 126–140 has been shown to significantly reduce food intake and body weight gain in mice (Grasso et al. 1999). Structurally, several of these amino acid residues have been modeled as playing a major role in the direct physical interaction between leptin and the extracellular domain of its receptor (Hiroike et al. 2000). Both the primary windowing method and the tertiary windowing method were able

to detect residues from the study of Hiroike et al. (2000) along the lineage leading to *Macaca mulatta*. Considering the residues in the three peptides used in the study of Grasso et al. (1999), only the tertiary windowing method detected evidence for positive selection in any of those regions. Along the lineage leading to hominoid the tertiary windowing method with window sizes of 4.5 and 6.5 Å, detected windows at the C-terminal region of the protein, corresponding to the peptide covering amino acids 126 to 140.

Gaucher et al. (2003) picked out the region from residue 85 to residue 119 along the lineage leading to hominoid as having been under positive diversifying selection. This section was only detected under such positive diversifying selection along the lineage leading to *Macaca mulatta* in our analysis. Further definition of functional change and verification of roles of individual residues awaits experimental biochemical analysis.

Along the lineages leading to *Macaca mulatta* and *Pongo pygmaeus*, both windowing methods detected windows in the N-terminal region corresponding to the fragment between amino acids 22–56. The regions detected along the two lineages did not overlap, except for 8.0-Å tertiary windows where a few residues overlapped between the two lineages. The fragment between amino acids 22–56 in human leptin has been implicated in hematopoiesis (Stamatiadis et al. 2003).

Often the tertiary windows consist of amino acid residues that are spread out in the sequence but close in 3D. This can result in an increased ability to detect functionally important evolving regions of the protein in much more detail (and with more significance) than with the primary windowing approach. The extra interacting residues also evolving under positive selective pressure enable better detection of regions in protein structure that have undergone positive diversifying selection. This is seen clearly in Table 3 along the rhesus lineage.

Tertiary windowing is described as a new complementary tool in the molecular evolution toolbox for detecting positive diversifying selection. It has been used in this study coupled with methods that explicitly reconstruct ancestral sequences and count differences along branches. This method performs well for recent events involving closely related sequences. However, tertiary windowing can be applied in the same manner in conjunction with maximum likelihood methods that treat K_a/K_s as a lineage specific parameter.

In windowing methods, K_a and K_s are treated the same way. However, if K_s is a baseline for the amount of selection available to sites in the gene, it may reduce sampling errors from small sample sizes in windows to normalize K_a calculated in windows by a global K_s value. The potential problem with this is if

any genomic level selection has resulted in site-specific differences in the amount of mutation available for protein-level selection. A global K_s normalizer has not been implemented in this study but may have some utility in windowing studies where there is no significant difference in K_s calculated across windows.

While some uncertainty remains about the precise timing of events driving positive diversifying selection during the divergence of primates, tertiary sequence windowing is presented as a valuable tool for diagnosing the evolution of a protein-encoding gene family when a solved three-dimensional structure of the protein exists. While this approach has been applied to the examination of regions and times of positive diversifying selection in the leptin gene family, it can be utilized to examine different modes and rates of protein evolution in general.

Acknowledgments. We thank Erik Sonnhammer for helpful discussions on methodology. Funding and support for this project were provided by the Swedish Foundation for Strategic Research, Vetenskapsrådet, and FUGE, the Norwegian National Functional Genomics Platform.

References

- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* 57:289–300
- Benner SA, Trabesinger-Rueff N, Schreiber DR (1998) Exobiology and post-genomic science. Converting primary structure into physiological function. *Adv Enzyme Reg* 38:155–180
- Benner SA, Caraco MD, Thomson JM, Gaucher EA (2002) Planetary biology: paleontological, geological, and molecular histories of life. *Science* 296:864–868
- Berman HW, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242
- Comeron JM (1995) A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. *J Mol Evol* 41:1152–1159
- Endo T, Ikeo K, Gojobori T (1996) Large-scale search for genes on which positive selection may operate. *Mol Biol Evol* 13:685–690
- Fares MA, Elena SF, Ortiz J, Moya A, Barrio E (2002) A sliding window-based method to detect selective constraints in protein-coding genes and its application to RNA viruses. *J Mol Evol* 55:509–521
- Gaucher EA, Miyamoto MM, Benner SA (2003) Evolutionary, structural and biochemical evidence for a new interaction site of the leptin obesity protein. *Genetics* 163:1549–1553
- Golding GB, Dean AM (1998) The structural basis of molecular adaptation. *Mol Biol Evol* 15:355–369
- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11:725–736
- Grasso P, Leinung MC, Lee DW (1999) Epitope mapping of secreted mouse leptin utilizing peripherally administered synthetic peptides. *Regul Pept* 85:93–100
- Hiroike T, Higo J, Jingami H, Toh H (2000) Homology modeling of human leptin/leptin receptor complex. *Biochem Biophys Res Commun* 275:154–158
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17:754–755
- Hughes AL, Nei M (1989) Nucleotide substitutions major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA* 86:958–962
- Ina Y (1995) New methods for estimating the numbers of synonymous and nonsynonymous substitutions. *J Mol Evol* 40:190–226
- Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 36:96–99
- Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* 2:150–174
- Liberles DA (2001) Evaluation of methods for extermination of a reconstructed history of gene sequence evolution. *Mol Biol Evol* 18:2040–2047
- Liberles DA, Schreiber DR, Govindarajan S, Chamberlin SG, Benner SA (2001) The Adaptive Evolution Database (TAED). *Genome Biol* 2(8):research0028.1–6
- Messier W, Stewart CB (1997) Episodic adaptive evolution of primate lysozymes. *Nature* 385:151–154
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936
- Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol Biol Evol* 19:271–281
- Samson WK, Murphy TC, Robison D, Vargas T, Tau E, Chang JK (1996) A 35 amino acid fragment of leptin inhibits feeding in the rat. *Endocrinology* 137:5182–5185
- Siltberg J, Liberles DA (2002) A simple covarion-based approach to analyse nucleotide substitution rates. *J Evol Biol* 15:588–594
- Stamatiadis DN, Chan JL, Cogswell R, Stefanopoulou HC, Bullen J, Katsilambros N, Stathakis CP, Mantzoros CS (2003) Elevated leptin fragments in renal failure correlate with BMI and haematopoiesis and are normalized by haemodialysis. *Clin Endocrinol* 60:434–441
- Suzuki Y, Gojobori T (1999) A method for detecting positive selection at single amino acid sites. *Mol Biol Evol* 16:1315–1328
- Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573
- Yang Z (2002) Phylogenetic analysis by maximum likelihood (PAML). Version 3.13 University College London, London
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917
- Zhang J, Nei M (1997) Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J Mol Evol* 44:S139–S146
- Zhang F, Basinski MB, Beals JM, Briggs SL, Churgay LM, Clawson DK, DiMarchi RD, Furman TC, Hale JE, Hsiung HM, Schoner BE, Smith DP, Zhang XY, Wery JP, Schevitz RW (1997) Crystal structure of the obese protein leptin-E100. *Nature* 387:206–209