REVIEWS

It has been proved remarkably difficult to get compelling evidence for changes in enzymes brought about by selection, not to speak of adaptive changes1.

Although Darwin’s theory of evolution by natural selection is generally accepted by biologists for morphological traits (including behavioural and physiological), the importance of natural selection in molecular evolution has long been a matter of debate. The neutral theory2 maintains that most observed molecular variation – both polymorphism within species and divergence between species – is due to random fixation of selectively neutral mutations. Well established cases of molecular adaptation have been rare3. Several tests of neutrality have been developed and applied to real data, and although they are powerful enough to reject strict neutrality in many genes, they rarely provide unequivocal evidence for positive darwinian selection.

Most convincing cases of adaptive molecular evolution have been identified through comparison of synonymous (silent; $d_S$) and nonsynonymous (amino acid-changing; $d_N$) substitution rates in protein-coding DNA sequences, thus providing fascinating case studies of natural selection in action on the protein molecule. Selected examples are listed in Table 1; see Hughes4 for detailed descriptions of many case studies. Here, we summarize recent methodological developments that improve the power to detect adaptive molecular evolution, and examine their strengths and weaknesses, so that they can be used to detect more cases of molecular adaptation.

Measuring selection using the nonsynonymous/synonymous ($d_N/d_S$) rate ratio

Traditionally, synonymous and nonsynonymous substitution rates (Box 1) are defined in the context of comparing two DNA sequences, with $d_S$ and $d_N$ as the numbers of synonymous and nonsynonymous substitutions per site, respectively5. Thus, the ratio $\omega = d_N/d_S$ measures the difference between the two rates and is most easily understood from a mathematical description of a codon substitution model (Box 2). If an amino acid change is neutral, it will be fixed at the same rate as a synonymous mutation, with $\omega = 1$. If the amino acid change is deleterious, purifying selection (Box 1) will reduce its fixation rate, thus $\omega < 1$. Only when the amino acid change offers a selective advantage is it fixed at a higher rate than a synonymous mutation, with $\omega > 1$. Therefore, an $\omega$ ratio significantly higher than one is convincing evidence for diversifying selection.

The codon-based analysis (Box 2) cannot infer whether synonymous substitutions are driven by mutation or selection, but it does not assume that synonymous substitutions are neutral. For example, highly biased codon usage can be caused by both mutational bias and selection (e.g. for translational efficiency6), and can greatly affect synonymous substitution rates. However, by employing parameters $\pi_j$ for the frequency of codon $j$ in the model (Box 2), estimation of
substitution rates will fully account for codon-usage bias (Box 1), irrespective of its source. Because parameter \( \omega \) is a measure of selective pressure on a protein, it differentiates codon-based analyses from the more general tests of neutrality proposed in population genetics. These general tests often lack the power to determine the sources of the departure from the strict neutral model, such as changes in population size, fluctuating environment or different forms of selection.

**Estimation of \( d_{S} \) and \( d_{N} \) between two sequences**

Two classes of methods have been suggested to estimate \( d_{N} \) and \( d_{S} \) between two protein-coding DNA sequences. The first class includes over a dozen intuitive methods developed since the early 1980s (Refs 5,9–15). These methods involve the following steps: counting synonymous (\( S \)) and nonsynonymous (\( N \)) sites in the two sequences, counting synonymous and nonsynonymous differences between the two sequences, and correcting for multiple substitutions at the same site. \( S \) and \( N \) are defined as the sequence length multiplied by the proportions of synonymous and nonsynonymous changes before selection on the protein. Most of these methods make simplistic assumptions about the nucleotide substitution process and also involve ad hoc treatment of the data that cannot be justified. Therefore, we refer to these methods of estimating \( d_{S} \) and \( d_{N} \) as approximate methods. The methods of Miyata and Yasunaga, and Nei and Gojobori, assume an equal rate for transitions \((T \leftrightarrow C \text{ and } A \leftrightarrow G)\) and transversions \((T,C \leftrightarrow A,G)\), as well as a uniform codon bias. Because transitions at the third ‘wobble’ position are more likely to be synonymous than transversions, ignoring the transition/transversion rate ratio leads to underestimation of \( \omega \) and overestimation of \( \gamma \) (Ref. 10). Efforts have been made to incorporate the transition/transversion rate bias (Box 1) when counting sites and differences. The effect of

---

**Box 1. Glossary**

**Codon-usage bias**: unequal codon frequencies in a gene.

**Nonsynonymous substitution**: a nucleotide substitution that does not change the encoded amino acid.

**Prior probability**: the probability of an event (such as a site belonging to a site class) before the collection of data.

**Positive selection**: Darwinian selection favoring advantageous mutations with positive selective coefficients. The term is used interchangeably with molecular adaptation and adaptive molecular evolution.

**Posterior probability**: the probability of an event conditional on the observed data, which reflects both the prior assumption and information in the data.

**Purifying selection**: natural selection against deleterious mutations with negative selective coefficients. The term is used interchangeably with negative selection or selective constraints.

**Synonymous substitution**: a nucleotide substitution that does not change the encoded amino acid.

**Transition/transversion ratio**: unequal substitution rates between nucleotides, with a higher rate for transitions (changes between T and C) and between A and G) than transversions (all other changes).
Box 2. A model of codon substitution

The codon is considered the unit of evolution. The substitution rate from codon i to j (\( \omega \)) is given as:

\[
q_{ij} = \begin{cases} 
0, & \text{if } i \text{ and } j \text{ differ at more than one position,} \\
\pi_j, & \text{for synonymous transversion,} \\
\kappa \pi_j, & \text{for synonymous transition,} \\
\omega \pi_j, & \text{for nonsynonymous transversion,} \\
\omega k \pi_j, & \text{for nonsynonymous transition.}
\end{cases}
\]

Parameter \( \kappa \) is the transition/transversion rate ratio, \( \pi_j \) is the equilibrium frequency of codon \( j \) and \( \omega = \frac{d_S}{d_N} \) measures the selective pressure on the protein. The \( q_{ij} \) are relative rates because time and rate are confounded in such an analysis. Given the rate matrix \( Q = (q_{ij}) \), the transition probability matrix over time \( t \) is calculated as:

\[
P(t) = (p(t)) = e^{Qt}
\]

where \( p(t) \) is the probability that codon \( i \) becomes codon \( j \) after time \( t \). Likelihood calculation on a phylogeny involves summing over all possible codons in extinct ancestors (internal nodes of the tree). After Refs 16,18,27,79.

<table>
<thead>
<tr>
<th>Method and/or model</th>
<th>( \kappa )</th>
<th>( S )</th>
<th>( N )</th>
<th>( d_S )</th>
<th>( \omega )</th>
<th>( \ell^2 )</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nei and Gojobori</td>
<td>1.0</td>
<td>109.4</td>
<td>316.6</td>
<td>0.0095</td>
<td>0.0569</td>
<td>0.168</td>
<td>9</td>
</tr>
<tr>
<td>Li</td>
<td></td>
<td>NA</td>
<td>NA</td>
<td>0.0104</td>
<td>0.0517</td>
<td>0.201</td>
<td>11</td>
</tr>
<tr>
<td>Ina</td>
<td>2.1</td>
<td>119.3</td>
<td>299.9</td>
<td>0.0101</td>
<td>0.0523</td>
<td>0.193</td>
<td>14</td>
</tr>
<tr>
<td>Yang and Nielsen</td>
<td>6.1</td>
<td>61.7</td>
<td>367.3</td>
<td>0.0063</td>
<td>0.1065</td>
<td>0.078</td>
<td>15</td>
</tr>
<tr>
<td>ML methods&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Fequal, ( \kappa = 1 )</td>
<td>1.0</td>
<td>109.5</td>
<td>317.5</td>
<td>0.0093</td>
<td>0.0557</td>
<td>0.167</td>
<td>633.67</td>
</tr>
<tr>
<td>(2) Fequal, ( \kappa ) estimated</td>
<td>3.0</td>
<td>124.6</td>
<td>301.4</td>
<td>0.0099</td>
<td>0.0480</td>
<td>0.206</td>
<td>632.47</td>
</tr>
<tr>
<td>(3) F1×4, ( \kappa = 1 ) fixed</td>
<td>1.0</td>
<td>129.1</td>
<td>296.9</td>
<td>0.0092</td>
<td>0.0671</td>
<td>0.137</td>
<td>612.40</td>
</tr>
<tr>
<td>(4) F1×4, ( \kappa ) estimated</td>
<td>3.9</td>
<td>137.1</td>
<td>288.9</td>
<td>0.0093</td>
<td>0.0624</td>
<td>0.149</td>
<td>610.48</td>
</tr>
<tr>
<td>(5) F3×4, ( \kappa = 1 ) fixed</td>
<td>1.0</td>
<td>63.2</td>
<td>362.8</td>
<td>0.0084</td>
<td>0.0973</td>
<td>0.087</td>
<td>560.76</td>
</tr>
<tr>
<td>(6) F3×4, ( \kappa ) estimated</td>
<td>5.4</td>
<td>60.6</td>
<td>365.4</td>
<td>0.0084</td>
<td>0.1061</td>
<td>0.079</td>
<td>557.85</td>
</tr>
<tr>
<td>(7) F61, ( \kappa = 1 ) fixed</td>
<td>1.0</td>
<td>58.3</td>
<td>367.7</td>
<td>0.0082</td>
<td>0.1145</td>
<td>0.072</td>
<td>501.39</td>
</tr>
<tr>
<td>(8) F61, ( \kappa ) estimated</td>
<td>5.3</td>
<td>55.3</td>
<td>370.7</td>
<td>0.0082</td>
<td>0.1237</td>
<td>0.066</td>
<td>498.61</td>
</tr>
</tbody>
</table>

<sup>a</sup>GenBank accession numbers are V00516 (human) and M12158 (orangutan).
<sup>b</sup>Fequal, equal codon frequencies (\( \omega = 1/61 \)) are assumed; F1×4, four nucleotide frequencies are used to calculate codon frequencies (3 free parameters); F3×4, nucleotide frequencies at three codon positions are used to calculate codon frequencies (9 free parameters); F61, all codon frequencies are used as free parameters (60 free parameters).
<sup>c</sup>\( \ell^2 \) is the log-likelihood value.

Statistical tests can be used to test whether \( d_S \) is significantly higher than \( d_S \). For approximate methods, a normal approximation is applied to \( d_S - d_N \). For ML, a likelihood-ratio test can be used. In this case, the null model has \( \omega \) fixed at 1, whereas the alternative model estimates \( \omega \) as a free parameter. Twice the log-likelihood difference between the two models is compared with a \( \chi^2 \) distribution with one degree of freedom to test whether \( \omega \) is different from one.

Computer simulation has been used to examine the performance of different estimation methods; the findings are consistent with observations made in real data analyses<sup>a,b</sup>. We demonstrate the effects of different estimation procedures using human and orangutan \( \alpha_2 \)-globin genes (Table 2). For comparison, different assumptions are made in ML concerning the transition/transversion rate bias and the codon-usage bias. The simpler models are each rejected when compared with more complex models by likelihood-ratio tests, confirming biased transition rates and codon usage. Thus, estimates from ML accounting for both biases (Model 8, Table 2) are expected to be the most reliable. We make the following observations:

- Assumptions appear to matter more than methods.
- The approximate methods and ML produce similar results under similar assumptions. The method of Nei and Gojobori is similar to ML under a model that ignores both transition/transversion bias and codon-usage bias (Model 1, Table 2), whereas the methods of Ina and Li are similar to ML under a model accounting for the transition/transversion bias but ignoring codon-usage bias (Model 2, Table 2). The method of Yang and Nielsen<sup>15</sup> is similar to ML under a model accounting for both biases (Model 6, Table 2). However, for distantly related sequences, \( \alpha_2 \) treatment in approximate methods can lead to serious biases even under the correct assumptions<sup>19</sup>.
- Ignoring the transition/transversion rate bias leads to underestimation of \( S \), overestimation of \( d_S \) and underestimation of the \( \omega \) ratio<sup>10</sup>.
- Codon-usage bias in these data has the opposite...
effect to the transition/transversion bias; ignoring codon-usage bias leads to overestimation of $\omega_i$, underestimation of $d_k$ and overestimation of $\omega$. This gene is extremely GC-rich at the third codon position, with base frequencies at 9% (T), 52% (C), 1% (A) and 37% (G). Most changes at the third position (before selection at the amino acid level) are transversions between C and G. Thus, the number of synonymous sites is less than half that expected under equal base and codon frequencies. Although, in theory, the bias caused by unequal codon frequencies can be in the opposite direction\cite{25}; we have not encountered a real gene showing that pattern. Such codon-usage bias appears to have missed previous analyses examining the relationship between the GC content at silent sites and $d_k$, because those studies ignored the codon-usage bias when estimating $d_k$ (Ref. 21).

- Different methods can produce different estimates, even when the sequences are highly similar. The sequences used in Table 2 are only about 10% different at silent sites and <1% different at nonsynonymous sites; however, estimates of $\omega$ are three times different. Because all estimation procedures partition the total numbers of sites and differences into synonymous and nonsynonymous categories, underestimation of one means overestimation of the other, thus resulting in large errors in the $\omega$ ratio.

**Detecting lineage-specific episodes of darwinian selection**

If, for most of the time, a gene evolves under purifying selection but is occasionally subject to episodes of adaptive change\cite{26}, a comparison between two distantly related sequences is unlikely to yield a $d_k/d_s$ ratio significantly greater than one. Methods have been developed to detect positive selection (Box 1) along specific lineages on a phylogeny. If the gene sequences of the extinct ancestors were compared with another model assuming an independent lineage (the free-ratio model), to test the neutral hypothesis for branches in a phylogeny\cite{18,27}. Using such models, two lineages in a primate phylogeny with highly elevated $d_k$ on some branches and $d_k/d_s$ > 1 in that lineage. Similarly, a model assuming the same $d_k/d_s$ ratio is identical among lineages\cite{18,27}.

It should be noted that variation in the $\omega$ ratio among lineages is a violation of the strictly neutral model\cite{21,28,29}, but it is not sufficient evidence for adaptive evolution. In particular, if nonsynonymous mutations are slightly deleterious, they will have a higher probability of fixation in a small population than in a large one\cite{30}, and thus lineages of different population sizes will have different $\omega$ ratios. Besides positive selection, relaxed selective constraint can also elevate the $\omega$ ratio – it might be difficult to distinguish the two if the estimated $\omega$ is not larger than one. Furthermore, it is incorrect to use the free-ratio model to identify lineages of interest and then to perform further tests on the $\omega$ ratios for those lineages using the same data without any correction\cite{31}.

**Box 3. Likelihood and Bayes**

The statistical-estimation theory used in the methods discussed in this review can be explained with the following simple hypothetical example. Suppose that a population is an admixture of two groups of people in the proportions 60% and 40%, and a certain disease occurs at a rate of 1% in Group I and 0.1% in Group II. Suppose a random sample of 100 individuals is taken from the population, what is the probability that three of them carry the disease? The probability that a random individual carries the disease ($D$) is an average over the two groups ($G_1$ and $G_2$):

$$p = P(D) = P(G_1) \times P(D|G_1) + P(G_2) \times P(D|G_2) = 0.6 \times 0.01 + 0.4 \times 0.001 = 0.0064$$

(1)

Similarly, the probability that an individual does not carry the disease is:

$$P(D) = P(G_1) \times P(D|G_1) + P(G_2) \times P(D|G_2) = 0.6 \times 0.99 + 0.4 \times 0.999 = 0.9936 = 1 - p$$

(2)

The probability that three out of 100 individuals carry the disease is given by the binomial probability:

$$P = \binom{100}{3} \times p^3 (1-p)^{97} = 0.0227$$

(3)

If Eqn 3 involves an unknown parameter [such as the rate $P(D|G_i)$ in Group I], that parameter can be estimated by maximizing Eqn 3. In that case, Eqn 3 gives the probability of observing the data (sample) and is called the likelihood function.

The second question is to calculate the probability that an individual in the sample who carries the disease is from Group I. The Bayes theorem gives this probability as:

$$P(G_1 \mid D) = P(G_1) \times P(D|G_1) / P(D) = 0.6 \times 0.01/0.0064 = 0.94$$

(4)

Note that this is just the proportion of the contribution from Group I to $P(D)$ in Eqn 1. Thus, this individual is most likely to be from Group I. Similarly, a healthy individual in the sample is more likely to be from Group I than from Group II because:

$$P(G_1 \mid D) = P(G_1) \times P(D|G_1) / P(D) = 0.6 \times 0.990.9936 = 0.5978$$

and

$$P(G_2 \mid D) = 1 - P(G_1 \mid D) = 0.4022$$

(5)

In methods for inferring sites under positive selection\cite{26,27}, we let $D$ be the data at a site and $G_i$ be the $i$th site class with the $d_i/d_s$ ratio $\omega_i$. The probability of observing data at a site is then an average over the site classes (Eqn 1). The product of such probabilities over sites constitutes the likelihood (Eqn 3), from which we estimate any unknown parameters, such as the branch lengths and parameters in the $\omega$ distribution over sites. After the parameters are estimated, we use the Bayes theorem to calculate the probability that any site, given data at that site, is from each site class (Eqns 4 and 5).

Another straightforward application of the theory is ancestral sequence reconstruction; in this case, we replace $G_i$ with a reconstruction (characters at interior nodes of the phylogeny) at a site. When we calculate the likelihood function, the probability of data at a site $P(D)$ is a sum over all possible ancestral reconstructions ($G_{si}$) (Eqns 1 and 2). After parameters are estimated, the reconstruction that makes the greatest contribution to $P(D)$ is the most likely (Eqns 4 and 5).

The Bayes method discussed here is known as the empirical Bayes, because it uses estimates of parameters and does not account for their sampling errors. This might be a concern if parameters are estimated from small samples or if the posterior probabilities are sensitive to parameter estimates. An alternative approach is the hierarchical Bayes method, which accounts for the uncertainty in unknown parameters by averaging over their prior distribution.

Note that the reconstructed ancestral sequences\cite{24}, as well as the inferred site classes in the site-class models\cite{28,29}, are pseudo data and involve systematic biases. To appreciate such biases, note that in the previous example, the Bayes calculations (Eqns 4 and 5) predict that each of the 100 individuals in the sample, healthy or sick, are from Group I. Although this is the best prediction, the accuracy is low. If such inferred group identities are used for further statistical analysis, misleading results might follow.
Methods based on ancestral reconstruction might not provide reliable statistical tests because they ignore errors and biases in reconstructed ancestral sequences (Box 3). The ML method has the advantage of not relying on reconstructed ancestral sequences. It can also easily incorporate features of DNA sequence evolution, such as the transition/transversion rate bias and codon-usage bias, and is thus based on a more realistic evolutionary model. When likelihood-ratio tests suggest adaptive evolution along certain lineages, ancestral reconstruction might be useful to pinpoint the involved amino acids and to infer ancestral proteins, which can be synthesized and examined in the laboratory.

### Detecting amino acid sites under darwinian selection

The methods discussed so far assume that all amino acid sites are under the same selective pressure, with the same \( \omega \) ratio. The analysis effectively averages the \( \omega \) ratio across all sites and positive selection is detected only if that average is \( >1 \). This appears to be a conservative test of positive selection because many sites might be under strong purifying selection owing to functional constraint, with the \( \omega \) ratio close to zero.

A few recent studies addressed this problem. Fitch and colleagues \(^{33,34}\) used parsimony to reconstruct ancestral DNA sequences, and counted changes at each codon site along branches of the tree. They tested whether the proportion of nonsynonymous substitutions at each site is greater than the average over all sites in the sequence. Suzuki and Gojobori\(^{15}\) took a more systematic approach. For each site in the sequence, they estimated the numbers of synonymous and nonsynonymous sites and differences along the tree using reconstructed ancestral sequences, and then tested whether the proportion of nonsynonymous substitutions differed from the neutral expectation (\( \omega = 1 \)). Suzuki and Gojobori's criterion is more stringent than Fitch et al.'s, because the \( \omega \) ratio averaged over sites is almost always \( <1 \). These methods are expected to require many sequences in the data set so that there are enough changes at individual sites. Furthermore, the reliability of significance values produced by these methods might be affected by the use of ancestral reconstruction, which is most unreliable at the positively selected or variable sites\(^{34}\), and by codon composition bias, which is most extreme at a single site.

In a likelihood model, it is impractical to use one \( \omega \) parameter for each site. The standard approach is to use a statistical distribution to describe the variation of \( \omega \) among sites; for example, we might assume several classes of sites in the protein with different \( \omega \) ratios\(^{36,37}\). The test of positive selection then involves two major steps: first, to test whether sites exist where \( \omega > 1 \), which is achieved by a likelihood-ratio test comparing a model that does not allow for such sites with a more general model that does; and second, to use the Bayes theorem to identify positively selected sites when they exist. Sites having high posterior probabilities (Box 1) for site classes with \( \omega > 1 \) are potential targets of diversifying selection. The theory is explained in Box 3 (Refs 20,36,37).

Nielsen and Yang\(^{36}\) implemented a likelihood-ratio test based on two simple models. The null model, M1 (neutral), assumes a class of conserved sites with \( \omega = 0 \) and another class of neutral sites with \( \omega = 1 \). The alternative model, M2 (selection), adds a third class of sites with \( \omega \) estimated from the data. (The model codes are those used in the codeml program in the PAML package.) If M2 fits the data significantly better than M1 and the estimated \( \omega \) ratio for the third class in M2 is \( >1 \), then some sites are under
poorly understood. It is also not entirely clear how to relate with their chemical properties, the relationship is
Although amino acid substitution rates are known to correlate and estimate different \( \omega \) parameters for sites from different structural and functional domains\(^{30}\).

**Limitations of current methods and future directions**
All the methods for detecting positive selection reviewed here appear to be conservative. They detect selection only if \( d_S > d_K \) is higher than \( d_S \) – selection that does not cause excessive nonsynonymous substitutions, such as balancing selection, might not be detected. The pairwise comparison has little power because it averages the \( \omega \) ratio over sites and over time. Methods for detecting selection along lineages work only if the \( \omega \) ratio averaged over all sites is \( >1 \). Similarly, the test of positive selection at sites works only if the \( \omega \) ratio averaged over all branches is \( >1 \). If adaptive evolution occurs only in a short time interval and affects only a few crucial amino acids, none of the methods is likely to succeed. Constancy of selective pressure at sites appears to be a much more serious assumption than constancy among lineages, especially for genes likely to be under continuous selective pressure, such as the HIV \( env \) gene. Indeed, models of variable selective pressures among sites\(^{36,37}\) have been successful in detecting positive selection, even in a background of overwhelming purifying selection indicated by an average \( \omega \) ratio much smaller than one\(^{37}\). Models that allow \( \omega \) to vary among both lineages and sites should have increased power.

The methods discussed here also assume the same \( \omega \) ratio for all possible amino acid changes; for example, at a positively selected site, all amino acid changes are assumed to be advantageous, which is unrealistic. Although amino acid substitution rates are known to correlate with their chemical properties, the relationship is poorly understood\(^{40,44}\). It is also not entirely clear how to define positive selection in a model accounting for chemical properties.

It will be interesting to perform computer simulations to examine the power of various detection methods and to investigate how this is affected by important factors, such as the size of the gene, sampling of species (sequences) and the level of sequence divergence. Including more sequences in the data should improve the power of site-based analyses. Sequence divergence is also important because neither very similar nor very divergent sequences contain much information. Very divergent sequences might also be associated with problems with alignment and unequal nucleotide compositions in different species. Analyses discussed here, which require information from both synonymous and non-synonymous substitutions, are expected to have a narrower window of suitable sequence divergences than phylogeny reconstruction. The large-sample \( \chi^2 \) approximation to the likelihood-ratio test statistic might also be examined, but limited simulations suggest that typical sequence data (with more than 100 codons) are large enough for it to be reliable. For very short genes or gene regions and especially at low sequence divergences, Monte Carlo simulation might be needed to derive the null distribution.

The likelihood analysis assumes no recombination within a gene. If recombination occurs, different regions will have different phylogenies. Empirical data analysis suggests that the phylogeny does not have much impact on tests of positive selection and identification of sites, and one might suspect that recombination will not cause false positives by the likelihood-ratio test. However, simulation studies are necessary to understand whether this is the case.

**Acknowledgements**
We thank D. Haydon, J. Mallet, T. Ohta, A. Pomiankowski, V. Vacquier, W. Swanson and three anonymous referees for comments. We also thank several users of the PAML package (http://abacus.gene.ucl.ac.uk/software/paml.html), in particular C. Woelk, for comments and suggestions concerning the implementation. This work is supported by grant #31/G10434 from the Biotechnology and Biological Sciences Research Council (UK).

**References**


Bielański, J. et al. Rates of nucleotide substitution and mammalian nuclear gene evolution: approx. and maximum-likelihood methods lead to different conclusions. Genetics (in press)


Ward, T.J. et al. (1997) Nucleotide sequence evolution at the k-casen gene of Apis mellifera: evidence for positive selection within the family Bovidae. Genetics 147, 1863–1872


The evolutionary origin of snakes (or Serpentes) has been discussed for over 130 years and their phylogenetic position within squamates is still debated. Around 2700 snake species are alive today and these are divided into three main groups1–3 (Box 1): tiny fossorial (burrowing) scolecophidians (blindsnakes); aniloids (pipesnakes), which are mostly semi-fossorial; and macrostomatans, which include more familiar taxa, such as boas, pythons, vipers and cobras. In addition to the more obvious diagnostic features of body elongation, limblessness and jaws that can engulf surprisingly large prey, other key features of snakes include absence of eyelids and external ears, and the presence of deep forked tongues (linked to their highly attuned and sophisticated chemosensory systems4).

Ancestral diggers or swimmers?

Hypotheses concerning snake interrelationships fall into two main groups. For some researchers, snakes descend from terrestrial squamates that developed fossorial (burrowing) habits. Two groups of lizards exhibiting such habitats, amphisbaenians and dibamids, have often been regarded as snakes’ closest living relatives4. Amphisbaenians, in particular, resemble scaly, loose-skinned earthworms, whose shovel-shaped or wedge-like heads function as soil-shunting devices. Specializations shared by snakes (Fig. 1a), amphisbaenians (Fig. 1b) and dibamids include loss, reduction and consolidation of skull bones; braincase enclosure; dorsal displacement of jaw-closing muscles; loss or reduction of limbs and girdles; and increased uniformity along the vertebral column. Furthermore, differences between the eyes of lizards and snakes are consistent with a model in which structures that were barely useful in a burrower underwent progressive reduction. Thus, whereas lizards, like humans, distort eye lens curvature to focus on objects, snakes lack ciliary muscles and are compelled to move the entire lens back and forth relative to the retina. Moreover, unlike lizards, snakes lack both a fovea and coloured oil droplets in retinal cells5.

Alternative hypotheses postulate that snakes are related to mosasauroids (Fig. 1c); spectacular marine reptiles from the upper half of the Cretaceous period, some 65–100 Mya6. Mosasauroids and snakes share reduced ossification of the pelvic and hindlimbs as well as specialized features of the jaw suspension and intramandibular joint kinetics (presence of a hinge allowing a degree of lateral movement within the lower jaw; Fig. 1a,c,d; Fig. 2, red circle). Phylogenetically, mosasauroids would be the nearest monophyletic sister group of snakes, with varanoid lizards (monitors) as the immediate sister group to this pair. Given this theory of relationships, the latest common ancestor of mosasaurs and snakes has been argued to have been a limbed, aquatic or semi-aquatic squamate5,7–11.

Snakes are one of the most extraordinary groups of terrestrial vertebrates, with numerous specializations distinguishing them from other squamates (lizards and their allies). Their musculoskeletal system allows creeping, burrowing, swimming and even gliding, and their predatory habits are aided by chemoreceptors, an extraordinary degree of cranial kinesis and, sometimes, of cranial kinesis and, sometimes, powerful venoms. Recent discoveries of indisputable early fossil snakes with posterior legs are generating intense debate about the evolutionary origin of these reptiles. New cladistic analyses dispute the precise significance and phylogenetic placement of these fossils. These conflicting hypotheses imply radically different scenarios of snake origins and relationships with wide biological implications.

Michael Coates and Marcello Ruta are at the Dept of Biology, Darwin Building, University College London, Gower Street, London, UK WC1E 6BT (m.coates@ucl.ac.uk; m.ruta@ucl.ac.uk).

Note that the implied ecological shift from an aquatic to a terrestrial environment in snake ancestry suggests that mosasaurs’ (implied) aquatic habits were also primitive for Serpentes. Subsequently, snakes reduced and lost their limbs, although rudiments of the posterior pair remain in some forms, such as pythons.

Fossils: perfect missing links...

Renewed interest in the origin of snakes has been triggered by the recognition and discovery of three remarkable fossil forms with hind legs. Each of these ancient snakes is around 97 My old and originates from lowermost Upper Cretaceous sediments in the Middle East.

Pachyrhachis problematicus, from Israel (Fig. 1d–f), rapidly assumed a central position in debates about snake phylogeny6,7,12. It has miniature hindlimbs articulated with a rudimentary pelvic girdle (Fig. 1e,f), but sadly, its feet are missing. Currently described from only two specimens, it