Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 223 (2003) 377–385
Journal of Theoretical Biology
www.elsevier.com/locate/jtbi

Testing fundamental evolutionary hypotheses

David Penny*, Michael D. Hendy, Anthony M. Poole[1]

Allan Wilson Center for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand

Received 3 July 2002; received in revised form 22 February 2003; accepted 24 February 2003

Abstract

Sober and Steel (J. Theor. Biol. 218, 395–408) give important limits on the use of current models with sequence data for studying ancient aspects of evolution; but they go too far in suggesting that several fundamental aspects of evolutionary theory cannot be tested in a normal scientific manner. To the contrary, we show examples of how some alternatives to the theory of descent can be formulated in such a way that they lead to predictions that can be evaluated (and rejected). The critical factor is a logical formulation of the alternatives, even though not all possible alternatives can be tested simultaneously. Similarly, some of the limits using DNA sequence data can be overcome by other types of sequence derived characters. The uniqueness (or not) of the origin of life, though still difficult, is similarly amenable to the testing of alternative hypotheses.

Keywords: Evolution; Genetic code; Intelligent Design; Phylogeny limits; Origin of life

1. Introduction

Sober and Steel, in a recent contribution to this journal (2002) critically examine the "Hypothesis of Common Ancestry"—that all life on earth traces back to a single common ancestor. They rightly point out that this is accepted within biology without rigorous testing, and they present theoretical results as to why the theory may be difficult to test. Our paper is a response to their claims, and illustrates how the Hypothesis of Common Ancestry has been tested in the past, and how difficult aspects (such as whether more than one ancestral lineage contributed to modern life) can be tested further. Our conclusion is that the Hypothesis of Common Ancestry is testable in principle, and it is not intrinsically different from other scientific theories. Nevertheless, Sober and Steel's paper is an important challenge; the reliability of current methods for building evolutionary trees from DNA sequence data is rightly criticized, as is the notion of a single (lineage-based) origin of all modern life. Thus questions analysed by Sober and Steel are fundamental in evolution—issues too often neglected. What are the expected limits of reconstructing evolutionary trees from sequences? How many ancestors are there for life on earth? Are there tests that distinguish between single- and multiple-origin hypotheses? We accept fundamental points in their article but secondary problems distract from the key issues, and the problems identified are solvable by good science. We focus on four main themes:

(a) the theory of descent leads to testable predictions,
(b) science does not claim to have absolute tests of hypotheses,
(c) the limits to phylogeny reconstruction depend on the model, and
(d) are there one, or more than one, common ancestors of life?

Sober and Steel suggest that the hypothesis of common ancestry is so ingrained in the minds of biologists that, when attempting to reconstruct the relationships that link a set of species, "the typical question is *which* tree is the best one, not *whether* there is a tree in the first place" (Sober and Steel, 2002). Historically, this is certainly not the case; many forms of relationship between species are possible (Fig. 1) and there is no a priori reason to assume a Steiner tree (Fig. 2). The concept of species having a continuity

*Corresponding author. Institute of Molecular BioSciences, Massey University, Private Bag, Palmerston North 11222, New Zealand. Tel.: +64-6-3505033; fax: +54-6-350-5688.
E-mail address: d.penny@massey.ac.nz (D. Penny).
[1] Current address: Department of Molecular Biology and Functional Genomics, Stockholm University, SE 106 91, Stockholm, Sweden.

0022-5193/03/$ - see front matter
doi:10.1016/S0022-5193(03)00099-7

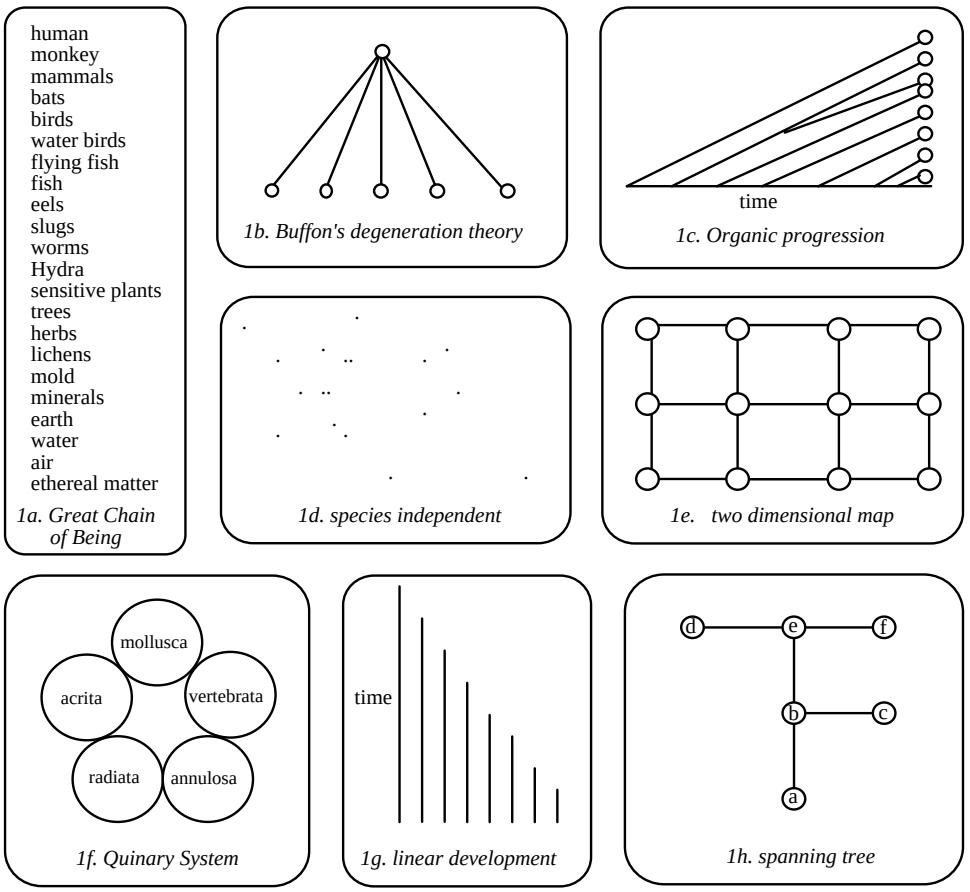

Fig. 1. Possible patterns for classifying species—examples of relationships between species that were considered by early biologists: (a) the Great Chain of Being, favored particularly by zoologists; most authors considered it static, others imagined species ascending the chain with time, some thought species degenerated with time; (b) a small number of species "degenerated" from one original species (for example, giant cats from one original "perfect" form) but the figure could represent several species created from a perfect archetype (idea) with each species modified for local conditions; (c) a representation of Lamarck's ideas of continued spontaneous generation of new "monads" which then ascended a form of the Great Chain of Being (shown with limited speciation); (d) a form of special creation where each species was designed for its environment without any overall pattern; (e) a two-dimensional surface (map) with species, or groups of species, occupying defined regions; (f) the quinary system with five "osculating" circles that were repeated at each level of classification, intermediate forms occurred at each intersection (osculation); (g) a linear development with species arising at the same time but ascending to "higher" forms at different rates; (h) a spanning tree that links existing species, species "b" could be derived from "a" (which remains unaltered), species "c" derived from "b" and so forth.

through time was only developed in the late 17th century (and only after continuous spontaneous generation of complex organisms was invalidated, see Farley, 1977). Higher life forms were no longer thought to "transmute" into different kinds during the lifetime of an individual. Many proposals relating these new entities (species) are shown in Fig. 1 and/or discussed in Bowler (1984). It took over 2000 years (from the time of the ancient Greeks), and over 150 years from the concept of permanent species, before a rooted Steiner tree was proposed by Charles Darwin. Some of the earlier ideas had common ancestors for subgroups, others did not. By providing a *mechanism* (natural selection, and descent with modification), Darwin could suggest a scientific model (pattern and mechanism) for species relationships. Darwin's evolutionary tree was neither obvious, nor easy to find. We claim that any alternative (as in Fig. 1) is testable individually, but that it is logically impossible to compare any one hypothesis against "all possible alternatives" (including those not yet specified). Rejecting all possible alternatives is logically equivalent to "proving" the theory.

2. The theory of descent leads to testable predictions

Sober and Steel consider three previous arguments that have been used to argue in favour of the hypothesis of Common Ancestry. Two, related to the origin of life and the genetic codes, are dealt with in Section 4. The third is an analysis of Penny et al. (1982) in which the theory of descent was tested by examining evolutionary relationships of mammals using five independent data-sets. We argue here that the specific criticisms of our analysis by Sober and Steel do not affect the validity of the test, and point out further tests that corroborate it.

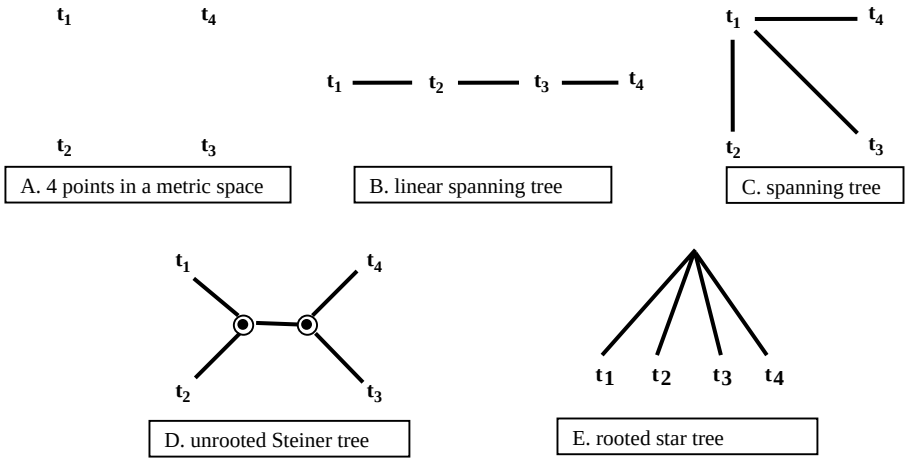

Fig. 2. Terminology for trees as used in the text. (A) Shows any four points in a metric space, (B,C) are spanning trees that link these four points, two of 16 possible spanning trees are shown. At least one of the 16 will be a minimal spanning tree for the metric used. In contrast, a Steiner tree ((D), shown as an unrooted tree) allows new internal points to be introduced. In general, a minimal Steiner tree will be shorter than a minimal spanning tree. In evolution we aim eventually for a rooted Steiner tree, and interpret the new internal points as ancestral to some current species. But there is nothing in the Steiner tree per se that requires it to be rooted. Steiner trees are extensively studied in mathematics (see Cieslik, 1998) and were well studied long before evolutionary trees were formalized as Steiner trees (Hendy et al., 1978). It is also a well-studied problem how much shorter a Steiner tree can be than a spanning tree (Cieslik, 2001). (F) is a rooted star tree used for comparing with Steiner trees.

In Penny et al. (1982) we compared minimal-length trees from five datasets of protein sequences, each with the same 11 species. Our conclusion was that the theory of evolution leads to quantitative predictions that are testable and is thus, in principle, falsifiable. Sober and Steel (2002) suggest that,

1. in some way our test depended on the parsimony optimality criterion,
2. parsimony assumes that the taxa are genealogically related,
3. our method relied on something called "character congruence",
4. a tree can generate noncongruent characters, and
5. unrelated taxa, by some unknown rules, can generate data that appears tree-like.

Our prediction from the theory of descent was that orthologous genes in mammals should lead to similar trees—they are expected to share the same evolutionary history. We found minimal-length trees from five protein datasets, and showed that the trees were much more similar than expected by chance. To do this, we

(a) developed a branch and bound search algorithm (guaranteed to find all minimal-length trees),
(b) implemented a tree-comparison metric to measure closeness objectively, and
(c) calculated the expected distribution of this metric.

We responded to a controversy in Nature (Anon, 1981a, b) as to whether evolution was a falsifiable theory. This involved two issues (see Halstead, 1980). There were comments by Karl Popper that evolution did not appear a normal scientific theory—rather it was a "metaphysical research programme" that could generate normal scientific theories. Then there were new exhibits at the British Museum of Natural History that appeared to question whether evolution had indeed occurred.

2.1. Test 1. The theory of descent

Parsimony was selected as our optimality criterion because it was the only one developed at that time, and we were able to implement a branch and bound search algorithm to guarantee optimality. Although logically other criteria such as maximum likelihood could be used, even today optimality cannot be guaranteed even for a single tree (Chor et al., 2000), and other search methods only find local optima. But nothing in the logic of the test depends on the optimality criterion used. Indeed, it is an excellent evaluation of the power of optimality criteria to compare their effectiveness in selecting highly similar trees from independent datasets. For example, if the trees selected by an ML program were more similar than the parsimony trees from the five datasets, then that is evidence that ML is more effective. The test must be done with real data, not data simulated on a tree in the first place. The particular optimality criterion used in our paper is not a central issue as to whether the theory of descent leads to falsifiable predictions.

The above discussion also answers the second criticism (an optimality criterion in some sense assumes a tree). A minimal-length Steiner tree can be calculated for any data, just like an average or a correlation coefficient. The calculation (average, correlation coefficient, or length of a Steiner tree) is independent of the

interpretation of the data. Our paper does not aim to "prove" the theory of descent for mammals, it allows a comparison against a null alternative (that there was no tree-like information in the data). Before continuing with the queries, it is necessary to show other examples in which alternative theories are tested.

2.2. Test 2. Influenza viruses from space

Another example of testing the theory of descent against alternatives is reported by Henderson et al. (1987, 1989) who test Hoyle and Wickramasinghe's (1984, 1986) claim that influenza viruses continue to arrive from outer space via comets. We examined two data sets (with 9 and 12 viral sequences) from epidemics between 1933 and 1980. Under the theory of descent, sequences should be close to a linear tree (Fig. 2B) with the sequences in the same order as the epidemics (1933 at one end to 1980 at the other). In contrast, if each epidemic was carried on different comets (which had formed millions of years earlier) then their order of arrival on earth should not correlate with their phylogeny. Indeed, there is no reason to expect any tree-like structure because comets could arise in different parts of the galaxy.

The first test calculated the probability of the sequences occurring on a linear tree in the same order as the years of the epidemics. The probability was less than one in 10^{-6} that the observed order occurs by chance—the theory of descent model survived this strong test, and that version of the comet model was falsified. The second test had an even stronger result. We compared a Steiner tree against a star-tree model (Fig. 1E). If the sequences were genuinely represented by a star tree, then there was only about one chance in 10^{64} of obtaining the observed pattern of only 11 nucleotide changes occurring twice. A fundamental point is that a general model (such as influenza arriving on comets) may not be testable "as a whole". Instead it is necessary to formalize versions of the model and test them (Riddiford and Penny, 1984). The theory of descent, combined with transmission between hosts, passed both tests but each version of the comet model that we could formalize (including a third one) was strongly rejected. We do not accept the Sober and Steel (2002) view that all possible alternatives to a model must be rejected simultaneously.

2.3. Test 3. Intelligent design

We can test the theory of descent versus a theory of individual creation of species—with each species being intelligently designed for its environment. Consider photosynthetic enzymes from plants living in a hot, dry desert (a cactus and a desert grass) with those from a moist-temperate grass. A wise creator might design similar photosynthetic enzymes for leaves functioning under hot dry conditions (the cactus and a desert grass). This version of intelligent design would predict the following rooted tree for these enzymes:

((cactus, desert grass), temperate grass)—see Fig. 3A.

This brings together enzymes from similar physical environments; under stress from high temperatures and strong water deficits. In contrast, the theory of descent predicts that the grass enzymes would be more similar:

(cactus, (desert grass, temperate grass))—see Fig. 3B.

This unites sequences sharing a more recent common ancestor, irrespective of their current physical environment. In practice, common ancestry gives the correct prediction for photosynthetic enzymes.

Many similar tests can be designed. The logic is identical for comparing protein sequences in the hairs of polar bears and snow rabbits with, say, those of a rabbit in a warm environment. Under intelligent design, the proteins in the two species living under Arctic conditions could be created to give maximum insulation under freezing conditions. Thus, hair proteins from species living in the Arctic would be similar for functional reasons. This test may not have been done, but the point is that the theory of descent leads to testable predictions. It is possible for Intelligent Design to fudge predictions to make them identical to the theory of descent, but this is unsatisfactory. It provides no mechanism that leads to the observed data, and it leads to a creator appearing to be the "Great Deceiver" who deliberately misleads rational humans.

3. Science does not claim to have absolute tests of hypotheses

Sober and Steel appear to assume that there must be a "definitive test" of any major scientific hypothesis; this is a fundamental question on the nature of science. However, science does not have absolute tests that "prove" a theory, we can never even think of all possible hypotheses. Could a better hypothesis for the structure of water be developed in 100 years? Similarly, there is no simple test that will prove the general theory of relativity once and for all. In reality, we may be able to reject a class of models, but we mainly test those that have been explicitly formulated. Once a new hypothesis has been proposed, it will be subject to testing against the old. It is sometimes claimed that "core" aspects of theories are "protected" from testing (see Riddiford and Penny, 1984). However, we argued that although some hypotheses are hard to test, there is more personal reward for scientists to find new innovative ways of testing any aspect of a difficult theory.

Scientific tests are comparative rather than universal—is X a better explanation of the data than Y? There are exceptions in that a hypothesis may test a much

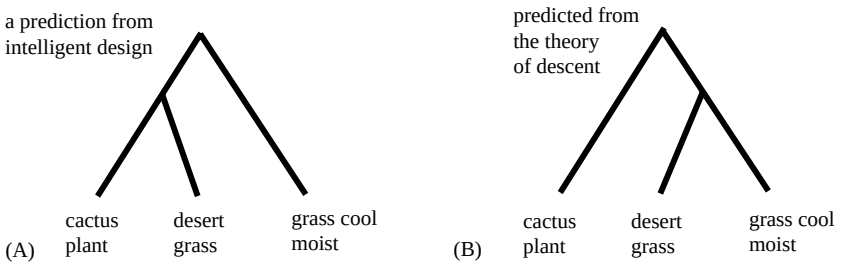

Fig. 3. A test of intelligent design versus the theory of descent. (A) is a possible prediction from Intelligent Design where the most similar protein sequences are found in the most similar physical environments. (B) is the prediction from the theory of descent, the most similar proteins are those found in the two grasses (because they share the most recent common ancestor).

more general hypothesis and our 1982 paper was (partially) such a one. Our claim, that we currently still stand by, was that we demonstrated that tests of the theory of descent are possible in principle. Certainly it was limited to mammals, but that is sufficient to show that tests are possible for other sets of organisms. We demonstrated that a testing mechanism exists, not that the theory was correct (which cannot be done in a Popperian way). In other words, we can do a test which can reject evolution, or the model of origin of viruses from comets, or versions of an intelligent design model.

Sober and Steel also raise the important issue of too many parameters in a model, eventually allowing any data to be generated on any tree. We have not found this in practice, though we do suspect the problem will be more acute at the limits imposed by Theorem 1 of Sober and Steel (2002). Artificial cases are possible where two trees give the same data, but adding another taxon destroys this (Waddell, 1995, pp. 426–435). The second point is biochemical, the rate of sequence evolution is proportional to the mutation rate during DNA synthesis and repair and this depends on up to 70 enzymes (see Lin et al., 2002). The basic error rate of DNA repair is independent of where the nucleotide fits in the gene, there is no mechanism that allows each site to have its own rate over all of evolution. This restricts the number of parameters.

A tree will not always be the correct model if a network (that allows cycles in the graph) is required (hybrids between plants, and the endosymbiotic origins of chloroplasts and mitochondria). Plant genes come from at least three sources and a network is an appropriate model—though, in practice, genes are considered separately and a tree drawn from each. Gene conversion is more complex because only a portion of the gene may be converted to another sequence. Lateral transfer of genes between bacterial "species" occurs by plasmids and other mechanisms to the extent that some authors (for example Doolittle, 2000) consider it the dominant process—others limit it (Jain et al., 1999). None of these cases overrides the use of the tree relationship between eukaryotic species as being the most useful model. Evolution, like other aspects of science, leads to testable predictions.

4. The limits to phylogeny reconstruction depend on the model

The conclusion of Sober and Steel (their Theorem 1) that current models of sequence evolution eventually limit phylogeny reconstruction is both important and fundamental; it has major consequences for studies of ancient divergences. Indeed, this subject has already moved away from confidence in the accuracy of ancient divergences inferred from a single gene, towards cautious phylogenetic interpretation. Examples such as Microsporidia have been recognized—these are a group of simple eukaryotic organisms originally thought to have branched off early from the main eukaryote trunk, but in fact are simplified fungi (Williams et al., 2002). Similarly, Lockhart et al. (2000) argued on empirical grounds that much of the information left for ancient bacterial divergences is artefactual (deviations from the model). In Penny et al. (2001), using that poor-cousin simulation, we reached a similar conclusion to Sober and Steel. We took estimated rates of molecular evolution for sites free to vary, and then simulated datasets with 1000 sites for periods ranging from 2 million to 2.5 billion years. The ability of current programs to recover the correct trees from the sequences was evaluated on these datasets; the results are fully congruent with Theorem 1 of Sober and Steel—information is eventually lost. In our simulations, by 500 million years, there was little information about deep phylogeny left in sequences under the standard model of molecular evolution.

However, it is necessary to qualify their conclusion in at least two ways—to sequence data and to the mechanisms of evolution they describe. Their current model is restricted to two-state characters and with sites staying in the same rate class over the whole tree. From our simulation results (Penny et al., 2001), we expect that four-state characters will lead to only marginally

longer retention of information under the model (sites always in the same rate class). It would be excellent to see their theorem extended to four states, a mathematical proof is always more convincing. However, other models may do better; a covarion model may allow primary sequences to retain information for longer times (Penny et al., 2001). This model allows sites to interchange between being fixed and being variable, thus freezing some phylogenetic information. This interchange occurs as the 3D structure of protein changes during evolution. It would be especially useful to extend the Sober/Steel theorem to include the Tuffley and Steel (1997) implementation of the covarion model. However, again we agree on the main issue; under our *current* models, sequences run out of useful information as the time to the common ancestor becomes large. There is currently far too much overconfidence in the ability of standard methods to recover ancient divergences.

While we are in agreement with Sober and Steel on the difficulties of reconstructing ancient divergences, using gene sequences directly is not the only way to reconstruct past evolutionary events. Paralogous genes (sequences that arose by gene duplication) increase the amount of information available, and thereby increase the chance of recovering information about the root of a tree. The emerging picture from genome-scale analyses is that both gene and genome duplication is much more frequent than previously supposed. For example, Lynch and Conery (2000) carried out a genome-level search in nine eukaryote genomes, and proposed that the extent of gene duplication is high. While this work has been subject to criticism (Long and Thornton, 2001; Zhang et al., 2001), the relevant point is that their results paint an optimistic picture for the use of such data for phylogenetic analyses. While there are substantial technical difficulties with such analyses (the problem of information loss over time being one example), these data permit testing of hypotheses not amenable to testing using only sequence data (Wolfe, 2001).

Similarly, the validity of the conclusion of information loss is thus far limited to primary sequence data; secondary and tertiary structures appear to retain information longer. We can recover trees from RNase P secondary structure even when we are unable to align the RNA itself (Collins et al., 2000). Similarly, Bujnicki (2000) used tertiary structures of proteins to infer evolutionary relationships. With the three classes of ribonucleotide reductase, support for their common ancestry from sequence data was very weak, their common origin has only been definitively demonstrated with the solution of their 3D structures (see Logan et al., 1999). This case is important for estimating the number of origins of DNA synthesis (see later). It is uncertain how useful 2D and 3D structures will be in general, but the Sober/Steel theorem does not address this issue. It was precisely because we expected primary sequences to run out of evolutionary information that we investigated secondary structure (Collins et al., 2000). But given the power of the Sober/Steel theorem, the onus is on anyone using only sequence data for ancient events to demonstrate that there is information left.

5. Are there one, or more than one, common ancestors of life?

Sober and Steel claim "It is a central tenet of modern evolutionary theory that all living things now on earth trace back to a single common ancestor", and suggest that it is impossible to establish whether more than one start-up contributed to modern life (Fig. 4d of Sober and Steel, 2002). We suggest it is possible to investigate the question scientifically, as follows. On biochemical grounds, it is argued that genetically encoded protein synthesis preceded DNA synthesis (and therefore DNA replication—see Poole et al., 1999, 2000, Fig. 4). However, such qualitative analyses of biochemical data do not distinguish between independent origins (as per Fig. 4d in Sober and Steel, 2002) and our sequential model (Fig. 4). Consider the following hypotheses:

Hypothesis 1. DNA synthesis arose in a descendant of the organism in which protein synthesis arose.

Hypothesis 2. DNA synthesis arose in a descendant of an independent (and now extinct) lineage with an unrelated protein synthetic machinery, and that there was transfer (mechanism unspecified) of either trait (ribosome or DNA synthesis) such that they ended up in the same lineage.

We assume that genetically encoded protein synthesis (and therefore a genetic code) was a prerequisite[2] for DNA synthesis (see Poole et al., 2000). We also assume, for the interim, that where multiple genetic codes are possible they all are equally fit. Multiple start-ups leading to a complete genetic code are permitted under hypothesis 1. Where the number of start-ups is much lower than the possible number of genetic codes, genetically encoded traits arising in a start-up would be unlikely to contribute to any other start-up (where start-ups $\ll$ genetic codes, the codes will tend to be incompatible). For non-coding traits (for example, RNA genes), this argument cannot be used. Hypothesis 1 is unlikely if start-ups $\gg$ genetic codes.

For hypothesis 2 to be correct, the genetic code in the two unrelated lineages must be identical. Thus, either there is only one possible genetic code, or

[2]The only known mechanism of deoxyribonucleotide synthesis is ribonucleotide reduction that requires protein radical chemistry. Control of protein radicals apparently requires complex proteins; so the emergence of DNA requires not only genetically encoded protein synthesis, but also must have post-dated a complete genetic code.

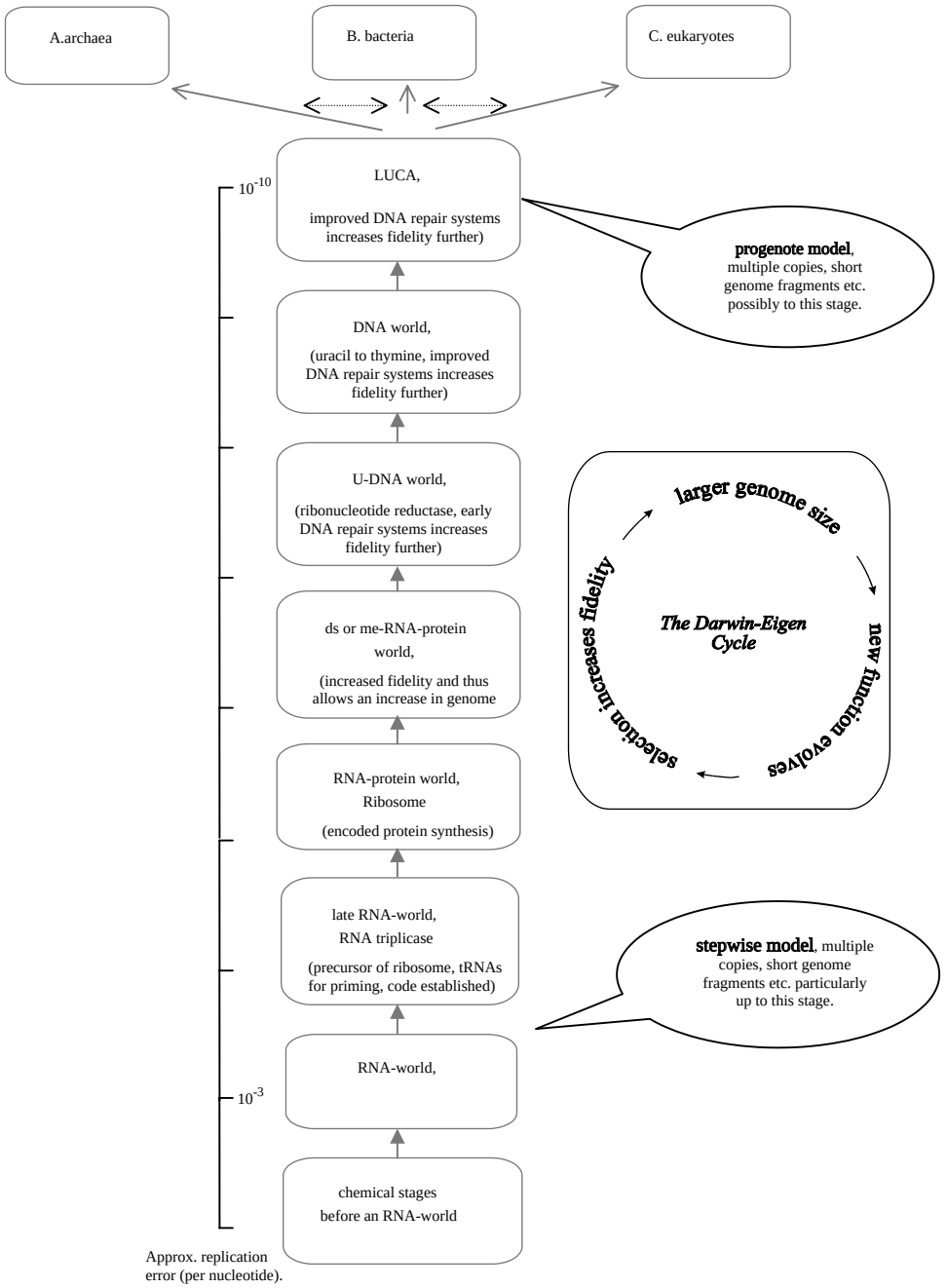

Fig. 4. A stepwise theory, showing parallel increases in replication fidelity and genome size in a positive feedback Darwin–Eigen cycle. The insert is a summary of this Darwin–Eigen cycle, a positive feedback loop between increased replication fidelity and the maximum possible genome size. The figure is extended from the work of Poole et al. (1999, 2000). In contrast, the progenote model of Woese (2000) has extensive horizontal gene transfer up to the LUCA stage, essentially without a lineage-based ancestor.

start-ups $\gg$ possible genetic codes. A further factor is the absolute frequency of transfer of traits, and the relative frequency of transfer between lineages with a common ancestor, relative to unrelated lineages with a common genetic code. The above model is simple: few start-ups and many possible codes favor hypothesis 1, while many start-ups and few possible codes (assuming reasonably frequent trait transfer or lineage fusion) are consistent with either. The problem is how to establish the number of possible codes, and the number of start-ups. The origin of the genetic code is an active area of research, and comparative testing of different models for its origin and evolution is amenable to standard scientific enquiry (e.g. Ronneberg et al., 2001). The number of start-ups is trickier, but both the emergence of multiple codes plus multiple start-ups can potentially be dealt with by considering the problem over time. As noted by Sober and Steel (2002), subsequent start-ups may not have the

same probability of emergence as the initial start-up; how could such a qualitative statement be turned into a quantitative one?

We have recently described a model describing the effect of intra- and interspecific competition over evolutionary time; the model, which we call evolutionarily stable niche discontinuity (ESND), accounts for the emergence of evolutionarily stable strategies for resource access (Poole et al., 2003), and can in principle be applied to any system with the basic properties of intra- and interspecific competition. In brief, competition both within species and between species occupying two different fitness peaks on a fitness landscape prevents members of either species from successfully moving away from their current niche towards that of the other species. Over time, the peaks become further separated as multiple traits contribute to the success of each species occupying each peak.

Applying this to the origin of the genetic code, there is no inherent reason why more than one code does not persist (other than the argument for "extinction of family names" noted by Sober and Steel, and assuming more than one code is possible, and that these are of equal fitness). However, such a selection-based model emphasises that the genetic code is not a single trait; it can be broken down into multiple traits (64 triplet codons that code for 20 amino acids and 3 stop signals, 61 tRNAs corresponding to the 61 coding codons, and 20 aminoacyl-tRNA synthetases for charging the tRNAs with their cognate amino acids). We would predict that as the number of adaptive changes increases over time, the probability of fixation of an identical start-up that emerged in the same location but at a later time would reduce, because additional traits now contribute to the fitness of the first start-up, meaning it will always outperform later start-ups (local optima at early stages notwithstanding). This would also be the case for different start-ups which have the same initial fitness; where one has a head-start, it will outcompete subsequent start-ups.

Indeed, such "ancestor-descendant" competitions have been carried out with *E. coli* (Lenski et al., 1998) and similar experiments could be designed using in vitro selection protocols. For instance, RNA-based aminoacyl-tRNA synthetases have been "evolved" through in vitro selection (Saito et al., 2001; Lee et al., 2000), and it would be possible to establish (through a competition experiment) how the presence of the "incumbent" influences the *de novo* emergence of additional aminoacyl-tRNA synthetases. Such experiments are technically demanding, but not impossible, and would provide a starting point from which to estimate the effect of the "incumbent" start-up on additional start-ups.

A final point concerns competition between start-ups that have never been in contact. We predict (Poole et al., 2003) that ESNDs would break down where non-coevolved competitors come into contact (e.g. introduction of exotic species into a new habitat), where horizontal transfer of a trait or traits results in the recipient being able to compete with incumbents (an example is the multiple independent emergence of pathogenic Shigella strains of *E. coli*, Pupo et al., 2000). With horizontal gene transfer, it is likely that the number of coevolved components will limit the success of fixation of a transfer (Jain et al., 1999), such that it would be unlikely that part of the coevolved machinery contributing to the genetic code would be easily transferred to another variant. Sober and Steel rightly point out difficulties with examination of events predating the LUCA. Nevertheless, we are convinced that these can be investigated with standard scientific reasoning. Advances in molecular experimentation are enabling the testing of theories once considered too complex to be reliably investigated—a viral system for investigating the prisoner's dilemma is one such example (Turner and Chao, 1999).

6. Conclusions

Sober and Steel's (2002) considerations on the difficulties in studying ancient evolutionary events are both timely and welcome. We disagree on details and think that clarifying some unfocused aspects in the paper helps get to key issues. Their Theorem 1 equally well supports the idea that there is strong evolutionary information in sequences for testing the theory of descent—as long as it is well within the limits imposed by the theorem. Importantly, it is time, not evolution, that is information destroying. Processes such as gene duplication and species divergence can increase the amount of information in the sense that these increase the chance of recovering information about the root.

The issues raised by Sober and Steel (2002) are basic and must be considered by a much wider range of researchers, but we do not see them as unique to evolution. They are common problems at the forefront of science. It is fundamental that researchers know the limits of their measuring instruments (in this case, recovering evolutionary trees from sequence data). The issues surrounding the testability of evolutionary theory are solvable by better science. There will seldom be one definitive test that will settle any major scientific theory "once and for all". Rather, we make specific tests of aspects of general theories. In the case of evolution we see that all aspects are able to lead to testable predictions, evolution is typical in this respect. However, at the level of the fundamental questions about early evolution, the Sober/Steel paper is a major contribution as it stands; all researchers in the subject should take it seriously.

Acknowledgements

We thank the New Zealand Marsden Fund for support.

References

Anon, 1981a. How true is the theory of evolution. Nature 290, 75–76.

Anon, 1981b. Darwin's death in South Kensington. Nature 290, 135.

Bowler, P.J., 1984. Evolution: The History of an Idea. University of California Press, Berkeley, CA.

Bujnicki, J.M., 2000. Phylogeny of the restriction endonuclease-like superfamily inferred from comparison of protein structures. J. Mol. Evol. 50, 39–44.

Chor, B., Holland, B.R., Penny, D., Hendy, M.D., 2000. Multiple maxima of likelihood in phylogenetic trees: an analytic approach. Mol. Biol. Evol. 17, 1529–1541.

Collins, L.J., Moulton, V., Penny, D., 2000. Use of RNA secondary structure for evolutionary relationships: the case of RNase P and RNase MRP. J. Mol. Evol. 51, 194–204.

Cieslik, D., 1998. Steiner Minimal Trees. Kluwer, Dordrecht.

Cieslik, D., 2001. The Steiner Ratio. Kluwer, Dordrecht.

Doolittle, W.F., 2000. The nature of the universal ancestor and the evolution of the proteome. Curr. Opin. Struct. Biol. 10, 355–358.

Farley, J., 1977. The Spontaneous Generation Controversy from Descartes to Oparin. The Johns Hopkins University Press, Baltimore, MD.

Halstead, B., 1980. Popper: good philosophy, bad science? New Scientist 87, 215–217.

Henderson, I.M., Penny, D., Hendy, M.D., 1987. Models for the evolution of influenza viruses. Nature 326, 22.

Henderson, I.M., Hendy, M.D., Penny, D., 1989. Influenza viruses, comets, and the science of evolutionary trees. J. Theor. Biol. 140, 289–303.

Hendy, M.D., Penny, D., Foulds, L.R., 1978. Identification of phylogenetic trees of minimal length. J. Theor. Biol. 71, 441–452.

Hoyle, F., Wickramasinghe, N.C., 1984. Living Comets. Cardiff University Press, Cardiff.

Hoyle, F., Wickramasinghe, N.C., 1986. The case for life as a cosmic phenomenon. Nature 322, 509–511.

Jain, R., Rivera, M.C., Lake, J.A., 1999. Horizontal transfer among genomes: the complexity hypothesis. Proc. Natl Acad. Sci. USA 96, 3801–3806.

Lee, N., Bessho, Y., Wei, K., Szostak, J.W., Suga, H., 2000. Ribozyme-catalyzed tRNA aminoacylation. Nat. Struct. Biol. 7, 28–33.

Lenski, R.E., Mongold, J.A., Sniegowski, P.D., Travisano, M., Vasi, F., Gerrish, P.J., Schmidt, T.M., 1998. Evolution of competitive fitness in experimental populations of *E. coli*: what makes one genotype a better competitor than another? Antonie Van Leeuwenhoek 73, 35–47.

Lin, Y.-H., Waddell, P.J., Penny, D., 2002. Pika and Vole mitochondrial genomes add support to both rodent monophyly and glires. Gene 294, 119–129.

Lockhart, P.J., Huson, D., et al., 2000. How molecules evolve in bacteria. Mol. Biol. Evol. 17, 835–838.

Logan, D.T., Andersson, J., Sjöberg, B.-M., Nordlund, P., 1999. A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. Science 283, 1499–1504.

Long, M., Thornton, K., 2001. Gene duplication and evolution. Science 293, 1551a.

Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.

Penny, D., Foulds, L.R., Hendy, M.D., 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature 297, 197–200.

Penny, D., McComish, B.J., Charleston, M.A., Hendy, M.D., 2001. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. J. Mol. Evol. 53, 711–723.

Poole, A.M., Jeffares, D.C., Penny, D., 1999. Prokaryotes, the new kids on the block. Bioessays 21, 880–889.

Poole, A., Penny, D., Sjöberg, B.-M., 2000. Methyl-RNA: an evolutionary bridge between RNA and DNA? Chem. Biol. 7, R207–R216.

Poole, A.M., Phillips, M.J., Penny, D., 2003. Prokaryote and eukaryote evolvability. Biosystems 69, 163–185.

Pupo, G.M., Lan, R., Reeves, P.R., 2000. Multiple independent origins of Shigella clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc. Natl Acad. Sci. USA 97, 10567–10572.

Riddiford, A., Penny, D., 1984. The scientific status of evolutionary theory. In: Pollard, J.W. (Ed.), Evolutionary Theory: Paths into the Future. Wiley, London and New York, pp. 1–38.

Ronneberg, T.A., Freeland, S.J., Landweber, L.F., 2001. Genview and Gencode: a pair of programs to test theories of genetic code evolution. Bioinformatics 17, 280–281.

Saito, H., Watanabe, K., Suga, H., 2001. Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. RNA 7, 1867–1878.

Sober, E., Steel, M., 2002. Testing the hypothesis of common ancestry. J. Theor. Biol. 218, 395–408.

Tuffley, C., Steel, M.A., 1997. Modeling the covarion hypothesis of nucleotide substitution. Math. Biosci. 147, 63–91.

Turner, P.E., Chao, L., 1999. Prisoner's dilemma in an RNA virus. Nature 398, 441–443.

Waddell, P.J., 1995. Statistical methods of phylogenetic analysis: including Hadamard conjugations, LogDet transformations and maximum likelihood. Ph.D. Thesis, Massey University, Palmerston North.

Williams, B.A.P., Hirt, R.P., Lucocq, J.M., Embley, T.M., 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. Nature 418, 865–869.

Woese, C.R., 2000. Interpreting the universal phylogenetic tree. Proc. Natl Acad. Sci. USA 97, 8392–8396.

Wolfe, K.H., 2001. Yesterday's polyploids and the mystery of diploidization. Nat. Rev. Genet. 2, 333–341.

Zhang, L., Gaut, B.S., Vision, T.J., 2001. Gene duplication and evolution. Science 293, 1551a.