How probable is common ancestry according to different evolutionary processes?

Elliott Sober [a], Mike Steel [b],*

[a] Philosophy Department, University of Wisconsin, Madison 53706, WI, USA
[b] School of Mathematics and Statistics, Allan Wilson Centre, University of Canterbury, Christchurch 8140, New Zealand

HIGHLIGHTS

- Darwin and contemporary biologists argue that all present-day life traces back to one or a few common ancestors.
- We investigate the relationship of different evolutionary processes to the hypothesis of common ancestry.
- We describe how different evolutionary processes confer different probabilities on the common ancestry thesis.

ARTICLE INFO

ABSTRACT

Darwin and contemporary biologists argue that all present-day life traces back to one or a few common ancestors. Here we investigate the relationship of different evolutionary processes to this hypothesis of common ancestry. We identify the property of an evolutionary process that determines what its probabilistic impact on the common ancestry thesis will be. The point of this exercise is to understand how the parts of Darwin's powerful theory fit together, not to call into question common ancestry or natural selection, since these two pillars of Darwin's theory enjoy strong support.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Darwin's views about natural selection (including his claim that selection is the main but not the exclusive cause of evolution) and his thesis about common ancestry (that everything now alive on earth traces back to one or a few original progenitors) are logically independent of each other (Mayr, 2007), but that did not prevent Darwin from illustrating them both in the single figure that appears in the *Origin of Species*. That diagram (which occurs on a fold-out page immediately following p. 116 of Darwin, 1859) is redrawn below in Fig. 1; there are 11 ancestors (labeled A, B, C, …, L) at the bottom of the page and 15 descendants at the top. Those 15 trace back to only 3 of the original 11. Why did the lineages stemming from 8 of the original 11 go extinct? Darwin's answer is natural selection. The horizontal axis in Darwin's figure represents a quantitative phenotype. Notice that when an ancestor produces several offspring in Darwin's drawing,

the offspring with extreme phenotypes are usually the ones that go on to have offspring of their own; those with intermediate phenotypes usually do less well. This is Darwin's principle of divergence. This selection process not only causes organisms in the present to trace back to a small number of ancient ancestors; it also serves to increase life's diversity. The 15 descendants at the top of the diagram have more horizontal spread than the 11 at the bottom. Darwin says that his principle of divergence describes what "tends" to happen in processes of natural selection, not what happens invariably. This is why he includes in his figure the lineage stemming from F. F and its present-day descendant have the same phenotype.

Darwin (1859, pp. 111–126) gives different characterizations of his principle of divergence (Kohn, 2009). In addition to the idea that selection favors extreme phenotypes, he says that selection favors parents that diversify their offspring and that it favors organisms that diversify their own internal structure. These three formulations are logically independent of each other. Current biology views two of them with reserve. Selection often favors extreme phenotypes, but it also often favors intermediate phenotypes; a classic example of the latter is birth weight in humans (Bell, 1997). Natural selection

* Corresponding author. Fax: +64 33642587.
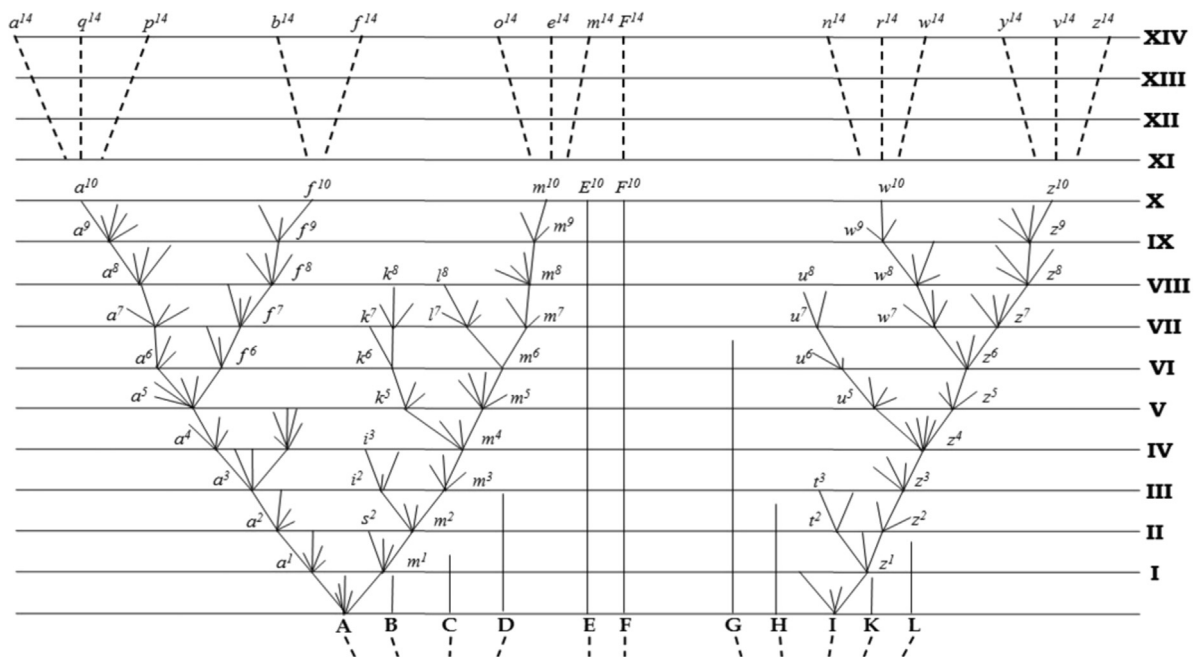E-mail address: mike.steel@canterbury.ac.nz (M. Steel).

Fig. 1. The only illustration in Darwin's *Origin* represents both common ancestry and the principle of divergence.

sometimes is disruptive and sometimes it is stabilizing; there is no a priori reason to think that one is common and the other rare. A similar reservation can be lodged against the idea that selection favors organisms that exhibit greater internal diversity. Selection sometimes promotes increased division of labor (specialization of parts), but often it does not. Darwin (1859, p. 148) notes in his discussion of parasites that selection sometimes favors simplification; parasites often lose structures that were present in their free-living ancestors. The idea in Darwin's trifecta that corresponds most closely to current biology is his suggestion that parents that diversify their offspring will be favored over parents that do not. This idea is alive and well in discussions of the evolution of sexual reproduction (Burt, 2000); organisms that reproduce sexually will do better than organisms that reproduce asexually when the environment is sufficiently unpredictable.

The question we will investigate here—how different evolutionary processes (including different forms of natural selection) confer probabilities on the thesis of common ancestry–is distinct from the question of which observations lend strong support to natural selection and which do so for common ancestry. Darwin addresses this last question in the following passage from the *Origin*:

> … adaptive characters, although of the utmost importance to the welfare of the being, are almost valueless to the systematist. For animals belonging to two most distinct lines of descent, may readily become adapted to similar conditions, and thus assume a close external resemblance; but such resemblances will not reveal– will rather tend to conceal their blood-relationship to their proper lines of descent (Darwin, 1859, p. 427).

Darwin's idea is that adaptive similarities provide scant evidence for common ancestry, whereas similarities that are neutral or deleterious provide evidence that is more weighty. As noted by Sober and Steel (2014), a simple likelihood comparison supports Darwin's claim about adaptive similarities under a rather general model of evolution. Supposing two extant taxa A and B share a trait x, let us consider the likelihood ratio $LR_{CA/SA}$ of the following two hypotheses:

(CA) Taxa A and B have a most recent common ancestor that existed t units of time in the past, trait x was present in this ancestor with some probability p, and the trait's evolution down each of the two lineages leading from that ancestor to A and to B followed continuous-time stochastic processes.

(SA) Taxa A and B do not trace back to a common ancestor, state x was independently present in these two taxa with probability p at t time units in the past, and the trait's subsequent evolution down these two lineages to the present followed continuous-time stochastic processes.

Sober and Steel point out that if t is small, then:

$$LR_{CA/SA} = \frac{\Pr(A, B \text{ share trait } x | CA)}{\Pr(A, B \text{ share trait } x | SA)} \approx p/p^2 = 1/p.$$

If an adaptive trait has a higher value for p than a neutral trait does, and if a neutral trait has a higher value for p than a deleterious trait possesses, then the value of $LR_{CA/SA}$ is higher for neutral and deleterious traits than it is for traits that are adaptive, thus vindicating Darwin's statement about adaptive similarities. This argument has two limitations: it requires t to be small and it considers only two taxa. While the value of $LR_{CA/SA}$ is not the main subject of the present paper, we expand on the $1/p$ argument by providing an exact expression for the likelihood ratio for two taxa when the evolving trait has two states; we also provide a bound on the ratio that applies for any number of taxa when the underlying continuous-time stochastic process is a stationary Markov process. Proofs for both are given in the Appendix.

Proposition 1. *For the evolution of a trait under a stationary continuous-time Markov process on two states:*

$$LR_{CA/SA} = 1 + \left(\frac{1}{p} - 1\right) e^{-2r \cdot t},$$

where r is a rate parameter associated with the model. Moreover, if n present taxa have their most recent common ancestor at t time units

in the past, then:

$$\mathrm{LR}_{\mathrm{CA/SA}} \geq 1 + \left(\frac{1}{p^{n-1}} - 1\right) e^{-nr \cdot t}.$$

Notice that the right-hand side of the equality and the right-hand side of the inequality both increase as p is made small.

2. Natural selection and common ancestry

In his discussion of the principle of divergence, Darwin describes how natural selection operates, and argues that it will have the result that all present-day organisms trace back to one or a few original progenitors. We will represent this conclusion about common ancestry by saying that i has a small value in CA_i. CA_1 means that all current life traces back to a single common ancestor, CA_2 says that all life traces back to two common ancestors (but to no fewer), and so on. We do not doubt that selection, as Darwin conceives it, *can* have the result that i is small in CA_i. Our question is whether the operation of natural selection *makes it probable* that i is small in CA_i and whether non-selective processes are able to do the same.

The epistemological problem that we will investigate here is the circumstances in which

For i small, $\Pr(\mathrm{CA}_i | M_1) > \Pr(\mathrm{CA}_i | M_2).$

where M_1 and M_2 are reasonably specific models of the evolutionary process. Here are the models we want to consider:

- Natural selection favoring extreme phenotypes
- Natural selection favoring intermediate phenotypes
- Natural selection that improves the average fitness of the organisms in a population
- Drift without natural selection

In addition to determining whether i being small in CA_i is more probable under some of these models than under others, we also want to see if some of these models make it very probable that i is small in CA_i.

In all of these models, we assume that life on earth began with some unknown number (n) of start-ups. These need not have arisen simultaneously. Present-day life traces back to i of these start-ups and no fewer, where $0 < i \leq n$. For each model, the probability that a start-up has at least one descendant alive today may depend on the characteristics (type) of that start-up; we suppose each start-up is of a type chosen according to fixed (unknown) probability distribution. We let $s(M)$ denote the resulting probability that a start-up in model M will have at least one descendant alive now (technical details are provided in the Appendix). Between the start-ups and the present, there can be branching, as shown in Darwin's diagram, and there also can be reticulation (wherein branches join). Our results do not depend on how frequently branching and reticulation occur in the descendants stemming from a given start-up.

Each of the models we will consider says that start-ups occur independently of each other, and whether a given start-up has a descendant that is alive now is independent of whether any other start-up does. This means that each model's chance that exactly $i = 0, 1, \ldots, n$ of the n start-ups has at least one descendant alive now is a binomial probability with parameters n and $s(M)$. However, the simple binomial formula fails to condition on the event E that at least one of these start-ups has at least one descendant alive now (i.e., that there is life on earth now). To take account of E, we need to divide each binomial probability by $\Pr(E|M) = 1 - (1 - s(M))^n$ and we need to restrict the range of i to values from 1 to n (i.e., excluding 0).

The likelihood ratio we want to consider for a pair of models M_1 and M_2 therefore takes the following form:

$$\mathrm{LR} = \frac{\Pr(\mathrm{CA}_i | M_1 \& E)}{\Pr(\mathrm{CA}_i | M_2 \& E)}.$$

For $j = 1, 2$, the numerator and the denominator of LR each take the form:

$$\Pr(\mathrm{CA}_i | M_j \& E) = \frac{\Pr(\mathrm{CA}_i \& E | M_j)}{\Pr(E | M_j)} = \frac{\Pr(\mathrm{CA}_i | M_j)}{\Pr(E | M_j)},$$

where the first equality is simply a formal identity of probability theory and the second equality holds for $i > 0$ since CA_i entails E. This leads to the following expression for the likelihood ratio:

$$\mathrm{LR} = \left[\frac{s(M_1)}{s(M_2)}\right]^i \left[\frac{1 - s(M_1)}{1 - s(M_2)}\right]^{n-i} \left[\frac{1 - (1 - s(M_2))^n}{1 - (1 - s(M_1))^n}\right].$$

Whether this likelihood ratio is greater than or less that unity depends on the values of $s(M_1)$ and $s(M_2)$ and on whether i is much smaller than n. Notice that $\mathrm{LR} = 1$ when $n = 1$, for any (non-zero) values of $s(M_1)$ and $s(M_2)$, since CA_1 occurs with probability 1 under both models. Thus we consider only the case where $n > 1$ in the following result, the proof of which is given in Appendix A:

Proposition 2. *Suppose $n > 1$. (i) If $i/n \leq s(M_1) < s(M_2)$ then $\mathrm{LR} > 1$. (ii) If $s(M_1) < s(M_2) \leq (i-1)/(n-1)$ then $\mathrm{LR} < 1$. (iii) If $s(M_1) \neq s(M_2)$, and neither probability is 0 or 1, then as i increases, LR goes from being on one side of 1 to the other exactly once.*

Notice that Proposition 2(i) implies that, for strong common ancestry ($i = 1$), likelihood favors the model that has the lower s-value when for each model at least one start-up is expected to survive to the present (i.e., $n \cdot s(M_1) \geq 1$ and $n \cdot s(M_2) \geq 1$). Part (ii) of Proposition 2 implies that when all n start-ups survive (i.e., $i = n$), the model with the higher s-value will have the higher likelihood. All three parts of Proposition 2 are illustrated in Fig. 2. In this example, $n = 100$ start-ups, $s(M_1) = 0.01$, and $s(M_2) = 0.05$.

Darwin says in the 5th edition of the *Origin* that his thesis that i is small in CA_i does not depend on how much bigger n is than i:

> No doubt it is possible … that at the first commencement of life many different forms were evolved; but if so, we may conclude that only a few have left modified descendants (Darwin and Peckham, 1959, p. 753).

This is true as far as the meaning of the thesis of common ancestry is concerned. However, Proposition 2 shows that the relation of i to n *does* affect which of two process models makes a small value for i in CA_i more probable.

Which properties of a model determine its value for $s(M)$? To answer this question, we divide the time between the n start-ups
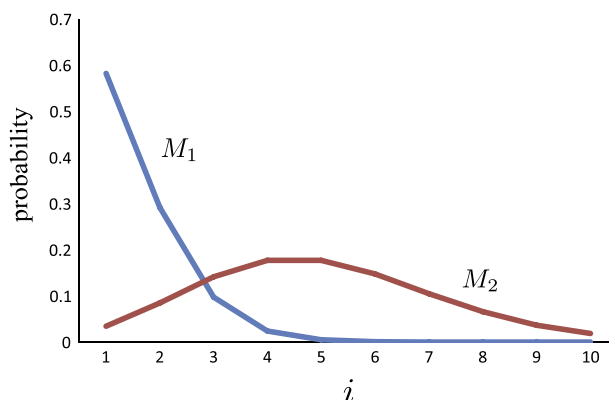

Fig. 2. Graphs of $\Pr(\mathrm{CA}_i | M_1 \& E)$ and $\Pr(\mathrm{CA}_i | M_2 \& E)$ as functions of $i = 1, 2, \ldots$, where $n = 100$, $s(M_1) = 0.01$ and $s(M_2) = 0.05$.

and the present into generations, with the start-ups occurring in generation 0 and the present descendants being in generation T. Let f_g be the average fitness of the individuals that are alive in generation g ($g = 0,...,T$-1). We assume a discrete branching process in which each individual has an independent Poisson-distributed number of surviving offspring in the next generation; it is left open whether the fitnesses of individuals (i.e., their expected numbers of offspring) within each generation vary. Appendix B provides a proof of the following result:

Proposition 3. *A model's value for $s(M)$ is fully determined by the sequence $f_0, f_1, f_2, ..., f_{T-1}$, and $s(M)$ does not further depend on how the individual fitness values vary within any generation. Moreover, increasing any of these f-values (singly or in combination) increases the value of $s(M)$.*

In saying here that "variance is irrelevant", we mean both variation *among* the descendants at any given time and variation *within* the descendants at any given time (if those descendants are themselves populations). The same point holds for variation among descendants at different times. Proposition 3 has the following consequences:

- A selection model M_1 and a drift model M_2 can have the same value for $s(M)$, in which case $\Pr(CA_i | M_1 \& E) = \Pr(CA_i | M_2 \& E)$, for each $i > 0$.
- Two selection models M_3 (where selection favors extreme phenotypes) and M_4 (where selection favors intermediate phenotypes) can have the same value for $s(M)$, in which case $\Pr(CA_i | M_3 \& E) = \Pr(CA_i | M_4 \& E)$, for each $i > 0$.

Notice that we say "can", not "must." There is nothing *intrinsic* to the distinction between selection and drift, nor to the distinction between selection for extreme phenotypes and selection for intermediate phenotypes, that makes a difference to the probability that i is small in CA_i.

Another model that is worth considering is suggested by the following passage from the *Origin*:

> If under a nearly similar climate, the eocene inhabitants of one quarter of the world were put into competition with the existing inhabitants of the same or some other quarter, the eocene fauna or flora would certainly be beaten and exterminated; as would a secondary fauna by an eocene, and a palæozoic fauna by a secondary fauna. I do not doubt that this process of improvement has affected in a marked and sensible manner the organization of the more recent and victorious forms of life, in comparison with the ancient and beaten forms; but I can see no way of testing this sort of progress (Darwin, 1859, p. 337).

Here Darwin is saying that selection improves fitness. This has the following consequence:

- If M_2 is a drift model in which $f_0 = f_1 = f_2 = \cdots = f_{T-1}$ and M_5 says that selection improves fitness with the result that $f_0 < f_1 < f_2 < \cdots < f_{T-1}$, then $s(M_2) < s(M_5)$, provided that both models start with the same value for f_0.

Selection when it improves fitness can do a worse job of getting i to be small in CA_i than drift does.

Two more alternatives to natural selection are worth mentioning, since Darwin thought about them both. The first is the doctrine of special creation. If a benevolent deity separately created numerous kinds of organism, and if this God would not destroy what he had created, i would not be small in CA_i. The second is Lamarck's theory of evolution, according to which the major kinds of organism now

alive trace back to separate start-ups.[1] Natural selection (as Darwin understood it) would be evidence that i is small in CA_i if these alternatives were the only games in town.

In summary, the models that confer the highest probability on CA_i when i is small are ones in which the probability of the survival of a start-up to the present (i.e. $s(M)$) lies close to the proportion of start-ups that survive; thus if there were a large number of start-ups, this favors models in which $s(M)$ is small. The impact of selection on $s(M)$ depends on its type – selection that tends to increase fitness between generations can lead to larger values of $s(M)$ than drift; however other types of selection (favoring extreme or intermediate phenotypes) can lead to the same $s(M)$ value, and this value could even be shared by a drift model.

In our analysis, we have assumed that whether a given start-up has descendants alive now is independent of whether other start-ups have extant descendants. This assumption is, at best, only approximately true. On the one hand, the success of one start-up may make other start-ups less viable, due to predation or competition. On the other hand, interactions between the lineages stemming from separate start-ups can be symbiotic, as in the emergence of hybrid lineages via events such as endosymbiosis in early life (Margulis, 1970). The assumption of independence ignores the intricacies of these events and interactions (the details of which would be difficult to estimate) by providing a simple null model that allows tractable calculations and avoids introducing numerous additional parameters. The conclusions drawn here generalize only to the extent that the survival of distinct start-ups can be viewed as "approximately" independent.

The relationship of Darwin's hypothesis that i is small in CA_i to his hypothesis that natural selection has been an important cause of trait evolution is multi-faceted. First, there is Mayr's point that the two are logically independent. Second there is Darwin's point that there can be strong evidence for common ancestry even if selection never caused a trait to evolve. Third, there is the point that common ancestry entails evolution, and Darwin's theory would be woefully incomplete unless it provided a mechanism that can cause traits to evolve (Darwin, 1859, p. 3). Fourth, there is the use that Darwin makes of the thesis of common ancestry to test hypotheses about natural selection. To figure out which traits natural selection will cause a species to have, one needs to know the traits that were present in the ancestors of that species. Without a time machine, the way to get a grip on the ancestral state of the lineage is to look at the target species's collateral relatives (Darwin, 1859, p. 187). It is the fact of common ancestry that allows Darwin to use the observed characteristics of extant species to make inferences about ancestors (Sober, 2011).

In this paper we have added to these points by describing two connections between common ancestry and natural selection in Darwin's theory. First, we have shown that there are forms of natural selection that make it highly probable that i is small in CA_i. When natural selection is conceived of in this way, we have an important harmony between these two parts of Darwin's theory: natural selection reinforces the thesis of common ancestry. However, Darwin describes other types of natural selection and these fail to make it highly probable that i is small in CA_i. In addition, nonselective processes can make it highly probable that i is small in CA_i. Thus, the case for common ancestry does not depend on natural selection's being the main cause of trait evolution, but certain forms of natural selection make common ancestry highly probable.

---

[1] For discussion of the nontheistic theories that preceded Darwin's, see Rupke (2005).

Acknowledgments

Appendix A. Proofs of Propositions 1–3

A.1. Proof of Proposition 1

By Proposition 8.5.2 of Semple and Steel (2003),

$$\Pr(A, B \text{ share trait } x|CA) = p[p + (1-p)e^{-\mu}]^2 + (1-p)[p(1-e^{-\mu})]^2, \tag{1}$$

where $\mu = r \cdot t$, and where r is the substitution rate divided by $1 - p^2 - (1-p)^2 = 2p(1-p)$. The first term on the right of (1) handles the case where the state x was present in the ancestor of A and B, while the second term accounts for the possibility that x was absent in this ancestor; the squaring present in both these terms recognizes that the two lineages leading from this ancestor A and B are conditionally independent, given the ancestral state. Straightforward algebra applied to (1) now shows that

$$\Pr(A, B \text{ share trait } x|CA) = p^2 + (1-p)pe^{-2\mu}. \tag{2}$$

Since the model is stationary, $\Pr(A, B \text{ share trait } x|SA) = p^2$, so by dividing (2) by p^2 and substituting $\mu = r \cdot t$, we arrive at the equation $LR_{CA/SA} = 1 + ((1-p)/p)e^{-2r \cdot t}$, as claimed.

For the second part of Proposition 1, not only do we allow more taxa (and an arbitrary tree topology) but we also provide a result that generalizes the 2-state non-symmetric model to the 'equal input model' on any number of states (for details, see Semple and Steel, 2003). For our calculations we exploit an equivalent description of this model sometimes referred to as 'Fortuin–Kasteleyn' type of random cluster model (see Section 2.1 of Matsen et al., 2008). Let $C = 1, 2, \ldots n$ be the number of components of the partition of the set of leaves of T induced by an independent Poisson process that acts with intensity r along the edges of the tree (the partition regards two leaves as being in the same component if the path between them does not cross an edge on which the Poisson event has occurred). Here r is the substitution rate, divided by 1 minus the sum of the squares of the stationary probabilities of the states. Then if ψ_x denote the probability that, under the equal input model, all n leaves of T are all in state x, and if p denotes the stationary probability of x, the random cluster description allows us to write ψ_x as follows:

$$\psi_x = \mathbb{E}[p^C] = \sum_{i=1}^{n} \Pr(C = i)p^i. \tag{3}$$

Notice that from Eq. (3), we have

$$\psi_x \geq p \cdot \Pr(C = 1) + p^n \cdot \Pr(C > 1).$$

Now, $\Pr(C = 1) = e^{-rL}$, where L is the sum of the branch lengths of T, Consequently, we have

$$\psi_x \geq pe^{-rL} + p^n(1 - e^{-rL}) = p^n + p(1 - p^{n-1})e^{-rL}. \tag{4}$$

Moreover, if the most recent common ancestor of the leaves of T is at time t in the past then $L \leq nt$ (and with equality precisely if T were a star tree). It thus follows from (4) that

$$\psi_x \geq p^n + p(1 - p^{n-1})e^{-nr \cdot t}.$$

Dividing this by p^n (the probability of the same event under (SA)), we obtain the lower bound on $LR_{CA/SA}$ given in Proposition 1. Notice that this bound is exact when $n = 2$.

A.2. Proof of Proposition 2

In the models we compare, the n start-up organisms may be of different types, with type A_j having probability a_j of occurrence as a start-up, and a probability s_j of having at least one surviving descendant at the present. The values a_j and s_j may depend on the model of evolution, so we let $s(M) = \sum_{j=1}^{n} a_j s_j$. Note that we have $\sum_{j=1}^{n} a_j = 1$ and that $s(M)$ is the probability that a randomly selected start-up (sampled according to the a_j frequencies) has survivors at the present. Suppose that each of the n start-ups arose independently (with the probabilities described), and the lineages generated by different start-ups survive to the present (or not) independently of each other. In this case, $\Pr(CA_i|M)$ is the binomial probability of obtaining exactly i successes in n trials, where the probability of success on each trial is $s(M)$.

By the independence assumptions above,

$$\Pr(CA_i|M\&E) = \binom{n}{i} s(M)^i (1 - s(M))^{n-i}/(1 - (1 - s(M)^n)).$$

Therefore, we obtain the stated expression for the likelihood namely:

$$LR = \frac{x^i(1-x)^{n-i}(1-(1-y)^n)}{y^i(1-y)^{n-i}(1-(1-x)^n)}, \tag{5}$$

where $x = s(M_1), y = s(M_2)$. Routine differential calculus shows that the (real-valued) function $\phi(t) = t^j(1-t)^{m-j}$ satisfies

$$\frac{d\phi}{dt} \leq 0 \text{ for } t \geq j/m \quad \text{and} \quad \frac{d\phi}{dt} \geq 0 \text{ for } t \leq j/m. \tag{6}$$

For Part (i), since $0 \leq x < y$ then $1 - (1-y)^n > 1 - (1-x)^n$, and so,

$$x^i(1-x)^{n-i}(1-(1-y)^n) > x^i(1-x)^{n-i}(1-(1-x)^n). \tag{7}$$

Now, applying (6) with $j = i$ and $m = n$ gives

$$x^i(1-x)^{n-i} \geq y^i(1-y)^{n-i}, \tag{8}$$

since $i/n \leq x < y$. Combining (7) and (8) gives

$$x^i(1-x)^{n-i}(1-(1-y)^n) > y^i(1-y)^{n-i}(1-(1-x)^n)$$

and so, by (5), $LR > 1$, as claimed.

For Part (ii), notice that the identity $1 - a^n = (1-a)(1 + a + a^2 + \cdots + a^{n-1})$ applied to $a = (1-x)$ and $a = (1-y)$ gives

$$1 - (1-x)^n = x(1 + (1-x) + (1-x)^2 + \cdots + (1-x)^{n-1}),$$

and

$$1 - (1-y)^n = y(1 + (1-y) + (1-y)^2 + \cdots + (1-y)^{n-1}).$$

Thus, since $0 \leq x < y$, and so $(1-x) > (1-y) \geq 0$, we obtain

$$\frac{x}{1-(1-x)^n} = \frac{1}{\sum_{j=0}^{n-1}(1-x)^j} < \frac{1}{\sum_{j=0}^{n-1}(1-y)^j} = \frac{y}{1-(1-y)^n}. \tag{9}$$

Moreover, applying (6) with $j = i - 1, m = n - 1$ gives

$$x^{i-1}(1-x)^{n-i} \leq y^{i-1}(1-y)^{n-i}, \tag{10}$$

since $x < y \leq (i-1)/(n-1)$. Multiplying the left-hand sides of (9) and (10), and applying these two inequalities gives

$$\frac{x^i(1-x)^{n-i}}{1-(1-x)^n} < \frac{y^i(1-y)^{n-i}}{1-(1-y)^n},$$

and so, by (5), $LR < 1$, as claimed.

For Part (iii), notice that the likelihood ratio can be written as

$$LR = Ae^{ai}, \tag{11}$$

where

$$A = \left(\frac{1-s(M_1)}{1-s(M_2)}\right)^n \cdot \frac{1-(1-s(M_2))^n}{1-(1-s(M_1))^n}$$

and

$$a = \ln\left(\left(\frac{s(M_1)}{1-s(M_1)} \right) \Big/ \left(\frac{s(M_2)}{1-s(M_2)} \right) \right).$$

The assumption that $s(M_1), s(M_2) \neq 0, 1$ implies that A and a are well-defined, and the assumption that $s(M_1) \neq s(M_2)$ implies that $a \neq 0$. Now for $a \neq 0$, the equation $Ae^{as} = 1$ has unique real solution for s, namely $s = \ln(1/A)/a$. It follows from (11) that as i increases, LR crosses the value 1 for at most one value of i. Furthermore, since $n > 1$, LR must cross the value 1 *exactly* once, since no probability distribution can (strictly) dominate another, and when $n > 1$, there are at least two values that i can take with positive probability.

A.3. Proof of Proposition 3

For each start-up, we model the population of its descendants as a discrete branching process, starting with a single individual at time 0 and ending with (zero or more) extant individuals at the present, which we will take as generation T. For any given one of the n start-ups, let s_T be the probability that there is at least one extant individual existing at the present that is a descendant of this start-up (this will be the s_j value referred to at the start of the previous proof, if the start-up is individual j). In this branching process, each individual is assumed to have a Poisson-distributed number of surviving offspring in the next generation; however, the expected number of surviving offspring from a single start-up can vary both within a generation and between generations. Thus, for individual x in generation j, we will let $\lambda(x,j)$ be the expected number of surviving offspring x has in generation $j+1$; we may regard this as a measure of the "fitness" of this individual.[2] We will let $\overline{\lambda_j}$ be the average $\lambda(x,j)$ value $\lambda(x,j)$ across all individuals x in generation j who are descendants of a given start-up (provided there is at least one individual present in generation j; otherwise, $\overline{\lambda_j}$ is undefined). Thus if $\mathcal{G}(j)$ is the (random) set of individuals in generation j then, provided this set is non-empty, we have

$$\overline{\lambda_j} = \frac{1}{|\mathcal{G}(j)|} \sum_{x \in \mathcal{G}(j)} \lambda(x,j),$$

where $|\mathcal{G}(j)|$ denotes the number of individuals in $\mathcal{G}(j)$.

Now suppose we are given (average) fitness values for each generation $f_j, j \geq 0$. We will write "$\overline{\lambda_j} \sim f_j$" provided that for each j between 0 and $T-1$, $G(j) \neq \varnothing$ and $\overline{\lambda_j} = f_j$, or $G(j) = \varnothing$ (thus $\sim$ is weaker than equality, since it leaves open the possibility that $\overline{\lambda_j}$ is undefined if the population is already extinct by generation j).

Let random variable N_j denote the size of the population in generation j for $j = 0, 1, \ldots N$. Thus $N_0 = 1$ and $s_T = Pr(N_T > 0)$. Now, for $j \geq 0$, we have

$$N_{j+1} = \sum_{x \in \mathcal{G}(j)} Y_{jx},$$

where $(Y_{jx} : x \in \mathcal{G}(j))$ is a sequence of independent Poisson random variables, with Y_{jx} having the expected value $\lambda(x,j)$. By the reproductive property of the Poisson distribution (a sum of independent Poisson variables has a Poisson distribution), it follows that, conditional on N_j, N_{j+1} has a Poisson distribution, with mean $\sum_{x \in \mathcal{G}(j)} \lambda(x,j) = f_j \cdot N_j$ (note this equality is also valid when $\mathcal{G}(j) = \varnothing$, since $f_j \cdot N_j = 0$ for any value f_j when $N_j = 0$). Thus, conditional on the $\overline{\lambda_j} \sim f_j$ for $j = 0, \ldots, T-1$, it follows by induction on j (starting with the base case $j = 0$) that the distribution of N_j is fully determined by the t_j values for each j and, in particular, for $j = T$. From this distribution on N_T, we have $s_T = Pr(N_T > 0)$. Notice that, although N_{j+1} has a Poisson distribution conditional on N_j, the (absolute) distribution of N_{j+1} will not, in general, be a Poisson distribution.

Finally, consider what happens when the f_j values are altered. Given two fitness profiles, $\mathbf{f} = (f_0, f_1, \ldots, f_{T-1})$ and $\mathbf{f}' = (f'_0, f'_1, \ldots, f_{T-}\ 1')$, we write $\mathbf{f} < \mathbf{f}'$ if $f_j \leq f'_j$ for all $0 \leq j < T$ with at least one of the inequalities being strict. Given two models of the type described, conditional on the fitness profiles $\mathbf{f}$ and $\mathbf{f}'$, the respective non-extinction probabilities at time T, s_T and s'_T, satisfy

$$\mathbf{f} < \mathbf{f}' \Rightarrow s_T < s'_T. \tag{12}$$

This follows by applying a coupling argument to establish (12) in the special case where $\mathbf{f}' = \mathbf{f} + \epsilon \cdot \mathbf{e}_i$ for any $\epsilon > 0$ (where $\mathbf{e}_i$ is the vector that has 0's except for position $i \in \{0, 1, \ldots, T-1\}$ where the entry is 1), from which the general case follows directly. Thus, any combination of one or more increases in the fitness values will necessarily increase s_T (i.e. larger values for $s(M)$ for the corresponding model M).

References

Bell, G., 1997. Selection – the Mechanism of Evolution. Springer, New York.

Burt, A., 2000. Sex, recombination, and the efficacy of selection – was Weismann right? Evolution 54, 337–351.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, 1st edition John Murray, London.

Darwin, C., 1959. On the origin of species – a variorum edition. In: Peckham, M. (Ed.), University of Pennsylvania Press, Philadelphia.

Kohn, D., 2009. Darwin's keystone – the principle of divergence. In: Ruse, M., Richards, R. (Eds.), The Cambridge Companion to the 'Origin of Species'. Cambridge University, Cambridge, pp. 87–108.

Margulis, L., 1970. Origin of Eukaryotic Cells. Yale University Press, New Haven.

Matsen, F.A., Mossel, E., Steel, M., 2008. Mixed-up trees: the structure of phylogenetic mixtures. Bull. Math. Biol. 70, 1115–1139.

Mayr, E., 2007. Darwin's five theories of evolution. In: What Makes Biology Unique? Cambridge University Press, New York, pp. 97–115.

Rupke, N., 2005. Neither creation nor evolution: the third way in mid-nineteenth century thinking about the origin of species. Ann. Hist. Philos. Biol. 10, 143–172.

Semple, C., Steel, M., 2003. Phylogenetics. Oxford University Press, Oxford, UK.

Sober, E., 2011. Did Darwin Write the Origin Backwards?. Prometheus Books, Amherst, NY.

Sober, E., Steel, M., 2014. Time and knowability in evolutionary processes. Philos. Sci. 81, 537–557.

---

[2] Since the offspring distribution is Poisson, the variance of that distribution is determined by (and is indeed equal to) this λ-value.