

The Rise and Fall of DNA Hybridization,
ca. 1980-1995,
or How I Got Interested in Science Studies

Jonathan Marks
Department of Anthropology
University of North Carolina at Charlotte
Charlotte, NC 28223

Phone: 704-687-2519
Fax: 704-687-3091
Email: jmarks@uncc.edu

For: Workshop on “Mechanisms of Fraud in Biomedical Research,” organized by Christine Hauskeller and Helga Satzinger. The Wellcome Trust, London, Oct. 17-18, 2008.

Draft: 15 July 2011, typos corrected 2018

[Without unreported data alterations] it is virtually certain that Sibley and Ahlquist would have concluded that *Homo*, *Pan*, and *Gorilla* form a trichotomy.

Sibley, Ahlquist, and Comstock, *Journal of Molecular Evolution* 30:225 (1990).

There are probably better ways to deal with the problems of inter-experimental variation that Drs. Sibley and Ahlquist faced, but their methods were logical, and made very little difference to inferences from the complete data.

Kirsch and Krajewski, *American Scientist* 81:410 (1993).

I will begin this paper with a question posed as an anthropologist: What would motivate an ostensibly honest and reputable scientist to publish an easily demonstrable falsehood – in defense of friends who are accused of fraud? The accused themselves had already publicly acknowledged the significant role of unreported data alterations in determining the conclusions of their study; now, three years later, their friends try to minimize it by saying that those alterations were not really important. Why would they say such a thing, given that it had already been refuted?

The answer must be that they perceived the stakes to be very high, and their own interests to lie with the accused.

The History: Charles Sibley, the Law, and DNA hybridization

On July 13, 1974, the *New York Times* ran an odd science story on its front page. About a month earlier, the director of Yale's Peabody Museum had been ordered to pay a fine of \$3,000 for violating the Lacey Act, which mandates that Americans must respect wildlife laws of other nations. Charles Sibley, Yale ornithologist and director of the Peabody Museum, had been accused of "illegally importing bird parts taken abroad in violation of foreign wildlife laws." Specifically, he had employed agents to steal the eggs of endangered bird species so he could analyze the proteins in their egg whites and produce a molecular phylogeny.¹

Now colleagues rallied around Sibley on the grounds that he was being "persecuted" by "extreme conservationists". Consequently, a few days later, the *Times* followed the story with an editorial:

The case against Dr. Charles G. Sibley, distinguished director of Yale's Peabody Museum, rests on the simple proposition that scientists, like politicians, are not above the law. An outstanding ornithologist, Dr. Sibley has been fined \$3,000 in a civil procedure for having systematically imported birds' eggs and egg whites that he was not licensed to import.

Scientists affirm the importance of Dr. Sibley's work, which involves a new method of classifying bird species by the protein content of the albumen. If his offense had been no more than the occasional and unsolicited receipt of an egg taken without a permit, he would perhaps be justified in complaining of "persecution" for merely technical violation of the Lacey Act – a statute which contains, among other

¹ Ferretti, F. (1974) Fining of bird scholar stirs colleagues. *The New York Times*, 13 July:

provisions, a ban on the importing of any animal or animal part taken contrary to a foreign country's wildlife protection laws.

Unfortunately the case involves more than that. Dr. Sibley appears to have used the services, in England, of an organized ring of illegal operatives, whose raids included the taking of eggs from the nests of such rare birds as the peregrine falcon, the stone curlew and the ringed plover. Dr. Sibley, it was charged, willfully received some of the material as well as the eggs of birds less endangered but nevertheless not stipulated in permits issued to him.

No doubt the temptation to circumvent bureaucratic red tape was strong, and Dr. Sibley's activities, unlike those of most violators of the Lacey Act, involved no personal profit. Nevertheless, as clear and deliberate evasions of the law, they cannot be justified by scientific purpose. The arrogance of science is no more appealing than the arrogance of commerce – or of government.²

The “arrogance of government” was a reference to Watergate, the infamous “third-rate burglary” of recent memory. The fine itself was the back end of a plea bargain in which criminal charges against Sibley were dropped. Although the leading science journals had news bureaus, the only journal that wrote it up was *Sports Illustrated* – since one of the birds involved was the peregrine falcon, and the issue seemed to be of greater interest to falconers than to molecular evolutionists. A few months later, Sibley was made Vice-President of the American Ornithologists' Union, which dismissed the matter as follows:

The Council considered the charges brought by the Department of the Interior under the Lacey Act against Charles G. Sibley in connection with importing egg-white specimens and adopted the following position statement: Council has examined and discussed the “Notice of Violation” of Section 43 (a) (2) of Title 18, U.S. Code (the “Lacey Act”), served on Dr. Charles G. Sibley, a member and officer of the American Ornithologists' Union, on 20 May 1974, by the Director of the Bureau of Sports Fisheries and Wildlife. It has noted that Dr. Sibley on the same day accepted the proposed civil penalty of \$3000 and paid that amount forthwith by check. The Council does note, however, that the Notice of Violation contains some errors and unwarranted charges. The Council is also familiar with accounts of these charges that appeared in the popular press and notes that, in some cases, these contain serious exaggerations and distortions. The Council of the American Ornithologists' Union does not condone, in any sense whatsoever, such violation by any member of the Union. It is the further view of the Council that the law has now taken its due course in this particular case.³

And life went on.

Over the next few years, Sibley abandoned the study of proteins in the construction of phylogeny for a technique that would isolate and analyze DNA itself. The technique came to be known as DNA hybridization, and was being applied in narrow

² Anonymous (1974) For the birds. *The New York Times*, 17 July.

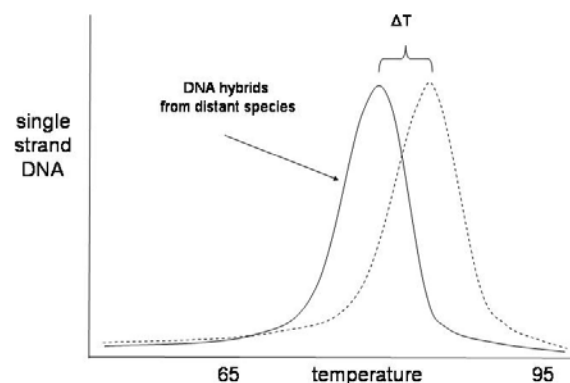
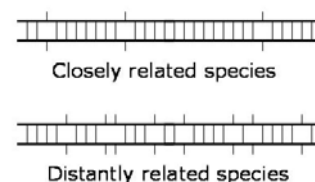
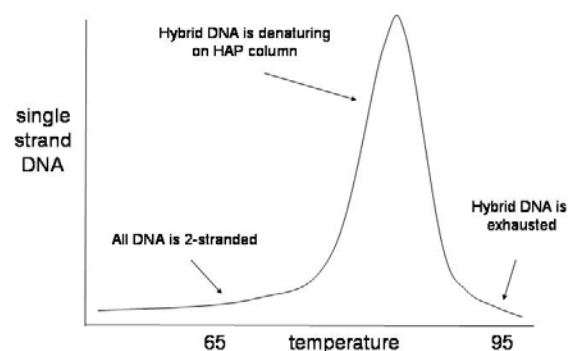
³ Watson, G. E. (1975) Proceedings of the ninety-second stated meeting of the American Ornithologists' Union. *The Auk*, 92:347-368. Quotation from pp. 352-353.

ways to questions of vertebrate phylogeny. Sibley wished to apply it to the birds, and with a protégé, Jon Ahlquist, developed a machine to mass-produce the relevant data, 25 experiments at a time.

The technique involved four steps, predicated on the double-helical structure of DNA, and the experimental ability to separate the two DNA strands from one another reversibly by heating and cooling. First, the DNA was biochemically isolated from blood, discarding the repetitive part of the genome (by boiling the DNA, which dissociates it into single strands, then allowing the DNA strands to rejoin or reanneal for a short period of time, and then discarding the fraction that reanneals rapidly, which is presumably the repetitive DNA), and radioactively labeling this unique-sequence DNA. Second, bringing this DNA into contact with a thousand-fold excess of unlabelled DNA from another species, and allowing it to reanneal, thus producing a sample of double-stranded DNA in which one strand from one species is labeled and the other strand from the other species is not. Third, dripping this DNA sample through hydroxylapatite (HAP), which has the property of binding double-stranded DNA, and allowing single-stranded DNA to pass through. Upon raising the temperature incrementally, DNA dissociates into single strands and passes through the column for collection. The radioactivity in each sample is a measure of the single-stranded DNA released at that specific temperature, and should ideally yield a single-peak curve when plotted against temperature.

Fourth, the information encoded in this curve must be retrieved and compared with other experiments. DNA hybrids from closely related species ought to be held together by many bonds, because their sequences are so similar, while DNA hybrids from distantly related species ought to be held together by fewer bonds, because their sequences are different. Thus, it should take less heat to separate the strands of DNA hybrids made from distantly related species. One need only calculate the differences in “melting temperatures” (or ΔT) and the result would be a scalar comparison of the similarity of genomic DNA across species.

By the early 1980s, Sibley and Ahlquist had published several papers on the phylogeny of birds, from the standpoint of DNA hybridization. Further, they began to wage a rhetorical war for the transcendence of their method of phylogenetic study over all others. The advantage of the technique, according to its proponents, is that it is genetic (rather than phenotypic), quantitative (rather than impressionistic), genomic (rather than based on a single DNA feature), and



precise or replicable (rather than idiosyncratic). Indeed, repeated in the derivative scholarly literature, DNA hybridization actually became a “holy grail” of phylogeny.⁴ There were skeptics within the ornithological community, to be sure, but they were clearly on the defensive in the face of Charles Sibley and his DNA hybridization machine.

With the obvious success of the technique in settling avian phylogeny, Yale anthropologist David Pilbeam (who moved to Harvard in 1981) suggested they apply it to a problem in primate evolution. Pilbeam’s work on *Ramapithecus* had proven to be a perfect foil for the pioneering “molecular clock” work of Vincent Sarich and Allan Wilson in 1967. A decade and a half later, Pilbeam pointed Sibley toward another problem.⁵ The molecular data consistently seemed to show the relationships among humans, chimpanzees, and gorillas as “too close to call” although the anatomy seemed to link chimpanzees to gorillas. What might DNA hybridization show?

In 1984, *The Journal of Molecular Evolution* published their result: Not only did DNA hybridization afford a “resolution of the trichotomy,” but the resolution was in favor of a human-chimpanzee connection, not gorilla-chimpanzee.⁶ It was genomic, quantitative, and precise. But was it real? Statisticians were quite vexed.⁷

The Sibley-Ahlquist data

In 1985 I was a post-doctoral researcher in a molecular genetics laboratory at the University of California at Davis. I was one of the first biological anthropologists to be sequencing DNA at the dawn of the genomics age, and was involved in a project on the evolution of alpha-globin genes in primates. As one of a small group of “molecular anthropologists” I began to be asked how credible the new work was. And like others in the field, I had no idea.

Carl Schmid ran a laboratory in the same department, working on the evolution of repetitive elements in the human genome. He actually had worked with the hydroxylapatite columns to separate repetitive from non-repetitive DNA, and was familiar with its use and whatever limitations it might have. Over a pitcher of beer one afternoon, he shared them with me.

Two issues were elevated above all others. First, the technique should not be able to work so well on very closely related species. It was predicated on the clear separation of repetitive and unique-sequence DNA, the former presumably being junk and the latter presumably being genes. Roy Britten, who had developed the technique for Sibley, had advanced a model of the genome in which repetitive DNA and unique sequence DNA

⁴ Gould, S. (1985) A clock of evolution. *Natural History*, 85 (4):12-25.

⁵ Maryellen Ruvolo, personal communication. Also: <http://www.sms.cam.ac.uk/media/754446>, approximately 29:10 into the video.

⁶ Sibley, C. G., and Ahlquist, Jon E. (1984) The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *Journal of Molecular Evolution*, 20:2-15.

⁷ Templeton, A. R. (1985) The phylogeny of the hominoid primates: A Statistical analysis of the DNA-DNA hybridization data. *Molecular Biology and Evolution*, 2:428-433. Saitou, N. (1986) On the Delta-Q test of Templeton. *Molecular Biology and Evolution*, 3:282-284. Ruvolo, M., and Smith, T. (1986) Phylogeny and DNA-DNA hybridization. *Molecular Biology and Evolution*, 3:285-289. Felsenstein, J. (1987) Estimation of hominoid phylogeny from a DNA hybridization data set. *Journal of Molecular Evolution*, 26:123-131.

were indeed compartmentalized, a generalization along the model of “satellite DNA,” which he had discovered. Schmid, however, had demonstrated that much of the repetitive DNA was not localized, like satellite DNA, but interspersed within the unique-sequence DNA. It was consequently impossible to purify the genomic unique-sequence DNA of redundancy. The genes themselves were repetitive: there were no less than seven gene sequences in the alpha-globin cluster. The presence of this interspersed redundancy might create a level of “noise” in the experiment that would render it impossible to make fine-scale phylogenetic determinations. The only way to tell would be to look carefully at the data for it.

Which brought us to the second issue: The highly-touted Sibley-Ahlquist paper included no meaningful documentation. It gave ΔT values for each pair of species, but provided only minimal discussion of how each individual value was produced. The crucial part of the analysis, after all, involved reducing a bell-shaped curve into a single number, but there were no bell-shaped curves, or DNA melting profiles, included in the paper to examine.

The only way to gauge the strength of the phylogenetic inference from the data would be to see the DNA melting profiles. The proper course of action, we decided, was to ask Sibley if we could reanalyze his data. Sibley refused. Next we approached Emile Zuckerkandl, editor of *The Journal of Molecular Evolution*, who supported Sibley’s right to sequester his data from potential critics. This particularly infuriated Schmid, who was on Zuckerkandl’s editorial board. We also began to hear grapevine stories about other workers in ornithology and molecular evolution who had asked to see some of the data, and were rebuffed on the grounds that they were potential critics.

We decided to reproduce manually a melting profile, as closely as possible to the conditions specified by Sibley and Ahlquist, and to use it as a springboard to present the potential problems and the lack of available documentation in this widely-discussed work. We prepared a manuscript and sent it to the *Journal of Human Evolution*, of which I was on the editorial board. The review was coordinated by board member Maryellen Ruvolo, who had already publicly spoken and published on behalf of the Sibley work. She sent it to Jon Ahlquist and Roy Britten to review, and joined them in unanimously deciding that the paper be rejected. I called the editor-in-chief, Eric Delson, to voice a complaint about the propriety of the review process. Delson agreed to set up a phone conversation between me and Roy Britten.

I have two recollections from the conversation with Britten. First, I remember arguing that if there were a major paper influencing opinion in his own field, and that the documentation for it were not available, would he not want that fact to be disseminated? And second, I distinctly remember feeling as though I were beating my head against a wall.

It was consequently entirely serendipitously that I received a package from Britten in early December of 1987, after having moved to Yale that summer. The good news was that the package contained about one-eighth of Sibley and Ahlquist’s raw primate data, which Britten had gotten from Sibley at a conference the year before, and now, with Sibley’s agreement, had copied for me.⁸ The bad news was that it was

⁸ The note, dated 2 Dec 1987, said, “Dr Marks, Charles said to send you the data I have and he will send you more if you need it. Roy Britten [P.S.-] Direct computer printouts!” I wrote to Sibley on 15 December 1987, but never received anything else.

approximately fifty pages of three columns of numbers, labeled as to experiment and species being compared, but with no indication of what each specific number actually meant. I copied it for Carl Schmid, and after wondering briefly about the propriety of the act, I sent a copy to Vince Sarich at Berkeley as well, with whom we had discussed the issues.

Schmid quickly deciphered the meaning of each row of numbers and we set out to calculate the values from the data we had, as closely as possible to the way Sibley and Ahlquist said they had. We quickly discovered a significant anomaly. Sibley and Ahlquist had claimed a high degree of replicability for their work. When they compared human to chimpanzee, they published the mean and range of the ΔT values; but we had one-eighth of their data, and we calculated values for several experiments that actually lay far outside the range they had given.

While we were mulling this over, Sibley and Ahlquist published their second paper on the apes. One of the defenses we had heard over the grapevine was that our questions were largely moot, since Sibley and Ahlquist had a paper in the pipeline that was going to settle the matter. When that was first articulated to me, however, I had replied, "You mean the first paper didn't settle the matter, then?"

The new paper, also in the *Journal of Molecular Evolution*, expanded the sample size, made the phylogenetic separation even clearer, and gave the particular experiment numbers and ΔT values.⁹ We were suddenly in the position of being able to match our ΔT values to theirs, for at least a portion of their experiments. And now we could see clearly that our calculations matched theirs 60% of the time, and differed by over a half-degree 40% of the time. A half-degree was a very significant discrepancy, for it was greater than the ostensible difference, in their original paper, between human-chimp DNA (1.8 degrees) and human-gorilla DNA (2.2 degrees).

It was now clear that some crucial analytic steps had been consistently omitted from their published papers, and that consequently the papers had been reviewed essentially under false pretenses – for the replicability that had anchored the work was clearly considerably overstated, at best. I wrote to Ahlquist on 16 February 1988 to ask for a clarification (the gravity of the matter now seemed to dictate keeping a paper trail; a response to this letter would ultimately arrive in early April¹⁰) and Carl Schmid, Vince Sarich, and I began to prepare two manuscripts now, for different target audiences. The first, submitted to the *Journal of Human Evolution*, was directly handled by the editors-in-chief, and went through two rounds of peer-review and a round of the lawyers at Academic Press before being published. The second was submitted to the *Journal of Molecular Evolution*, and review was coordinated by the editorial board member Allan Wilson. Wilson was a long-time collaborator of Sarich's, and Schmid was himself on the editorial board. Wilson helped rewrite the paper after a first review, and after a second round of enthusiastic reviews, accepted the paper on behalf of the journal. A few days later, we received word that Emile Zuckerkandl, editor-in-chief of the *Journal of*

⁹ Sibley, C. G., and Ahlquist, Jon E. (1987) DNA hybridization evidence of hominoid phylogeny: Evidence from an expanded data set. *Journal of Molecular Evolution*, 26:99-121.

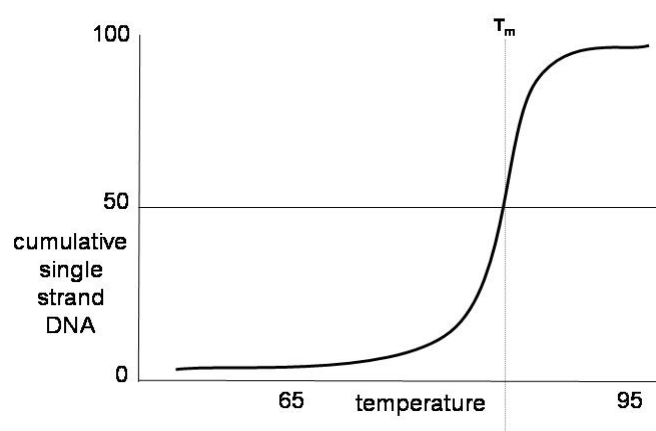
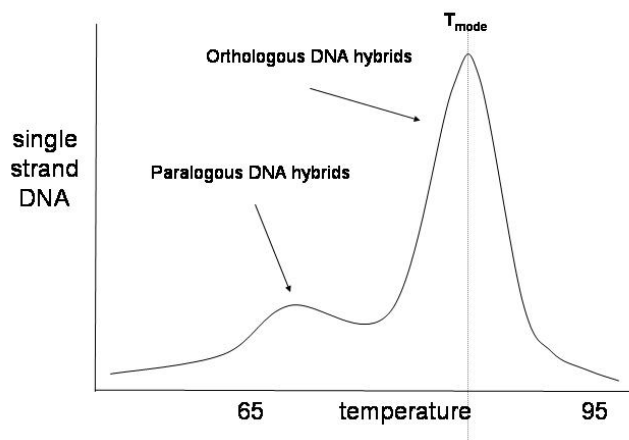
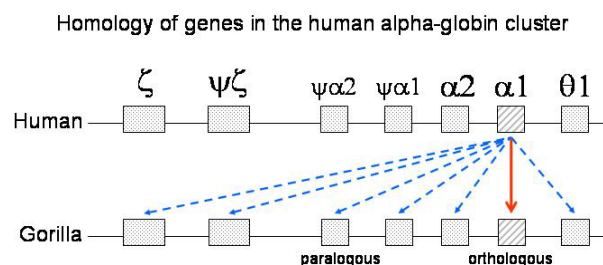
¹⁰ Ahlquist to Marks, dated 21 March 1988, but postmarked 29 March 1988.

Molecular Evolution, had un-accepted it. It was subsequently published in the journal *Cladistics*.¹¹

Why DNA Hybridization Doesn't Work: Paralogy

The ubiquity of repetitive DNA means that there will always be a small proportion of DNA able to bond, however imperfectly, to the “wrong” partner in the other species. The amount of pairing of these serial homologs in different species, or paralogs, will depend upon the features of any particular experiment. Where the crucial distinction is between a few tenths of a degree, a bit of poorly-paired paralogous DNA may artificially deflate the melting temperature, because of the way that value is calculated.

The presence of paralogous DNA would show up on the melting profile as a small satellite peak to the left of the main one. Being less perfectly matched to their DNA partners, the paralogs would be less thermally stable and elute off the HAP column at a lower temperature. If we measure the “melting temperature” as the temperature corresponding to the highest point on the melting curve, the melting temperature would be unaffected by the size of this secondary peak. We can call this measurement of the thermal stability of the DNA hybrids T_{mode} . Its merit is that it is insensitive to the paralogous DNA; but its disadvantage is that it is difficult to calculate precisely, because the temperature is raised in increments of 2.5 degrees, and the data are collected discontinuously. Some subjective “curve fitting” is thus necessary to estimate the temperature precisely, which is just what practitioners claimed to be avoiding.



To obtain a more precise value, we can transform the melting profile into a cumulative curve, and interpolate to find the temperature at which 50% of the DNA is

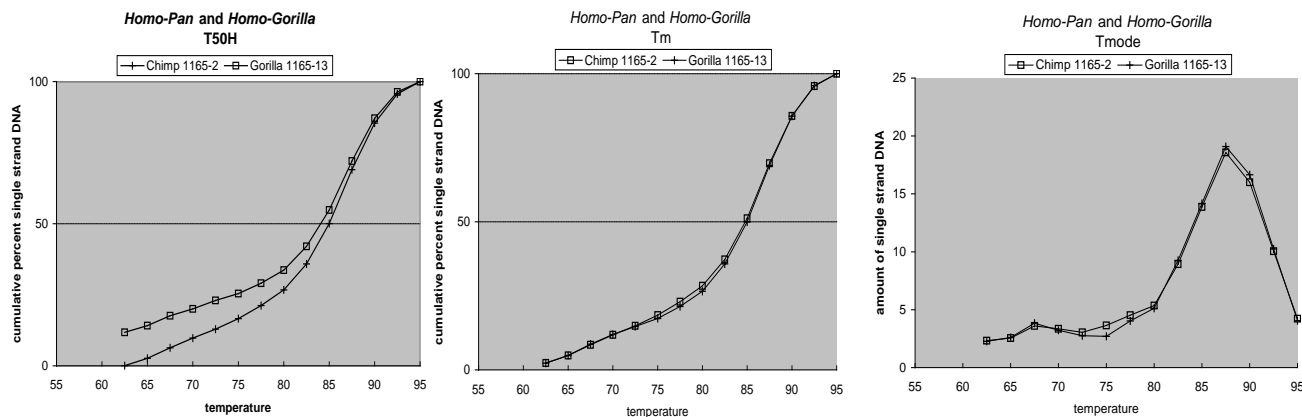
¹¹ Allan C. Wilson to Vince Sarich, Carl W. Schmid, and Jon Marks, 26 August 1988. Emile Zuckerkandl to Vincent M. Sarich, 19 September 1988. Sarich, V. M., Schmid, C. W., and Marks, J. (1988) DNA hybridization as a guide to phylogeny: A critical analysis. *Cladistics*, 5:3-32.

single-stranded, or the median value. However, if there is a small fraction of DNA eluting as paralogs – say 5% of the sample – then the 50% point will necessarily incorporate that, and we will record a value that does not actually reflect the thermal stability of the orthologs, which is what we are presumably interested in. If 5% of the DNA is paralogs, then the DNA of interest comprises only 95% of the total; the cumulative curve itself will be shifted to the left slightly, and we will record a 50% value when actually only 45% of the orthologs have melted (along with all of the paralogs, comprising the first 5%). So this measurement, the simple median or T_m , has the advantage of precision – being straightforward to produce – but has the disadvantage of inaccuracy, depending upon the particulars of the DNA preparation, expressed in the melting curve.

Neither of those, however, was what Sibley and Ahlquist measured. Preferring the more precise measurement, they discarded the T_{mode} . They also wanted a measurement of the melting temperature of the DNA hybrids that could be used to compare species quite distantly related to one another. So they adopted a median (50%) value, but not of the DNA that actually formed hybrids. To return to the formation of the hybrid DNA itself, we initially assumed that all of the labeled DNA found a partner in the other species' DNA. But for distantly related species, not only will the thermal stability of their DNA hybrids be lower, but DNA segments are more likely *not to be able to find a partner at all* in the other species' DNA sequences.

So the percentage of DNA that actually forms hybrids is itself a measure of the genetic similarity of the two species, albeit a rather crude one. This gives us two different measures of genetic similarity: the amount of DNA that was able to pair up between the two species, and the thermal stability of the DNA hybrids. Sibley and Ahlquist combined these into a single value, as follows. Their equipment ran the thermal elutions of hybrid DNA in batches of 25, the first always being the DNA of one species hybridized to itself, that is to say, the homoduplex control – from which the melting temperatures of the inter-species DNA hybrids or heteroduplexes would be subtracted to yield the ΔT . The first data point (collected at the lowest temperature) of each thermal elution through the hydroxylapatite column would contain any unhybridized DNA, which would be discarded or ignored in the calculation of T_{mode} and T_m . Sibley and Ahlquist, however, retained this information, and assuming that the control had hybridized to 100%, would calculate a normalized percent of hybridization (NPH) for each experiment. If a particular experiment had an NPH of 90%, then their cumulative curve, from which they would interpolate the median value, would begin with 10% of the hybrid DNA being single-stranded, rather than 0%. This median value, which incorporates the extent of hybridization into the calculation of thermal stability, is the T_{50H} . Its benefit is that it permits measurements to be made on more distantly related DNA samples; its disadvantage is that it adulterates the measurement of the thermal stability of the DNA hybrids, and adds an additional source of experimental variability. Not only would the individual values be affected by variation in thermal stability, but as well by variation in the extent of hybridization. And more importantly, two experiments producing DNA hybrids with identical thermal stabilities might appear spuriously to be different, if their extent of hybridization were different.

That is precisely the situation that exists for these two experiments from the Sibley-Ahlquist work: 1165-2 (Human to Chimpanzee) and 1165-13 (Human to Gorilla).



If we compare their T_{50H} values (left), we see the Human-Chimp temperature (at 50% single-stranded, the median) to be considerably higher than the Human-Gorilla value. But the Human-Gorilla curve begins (at 62.5°) with about 10% of the DNA hybrids already single stranded. If we eliminate this information and study only the thermal stability of the DNA hybrids, that is to say, the T_m (center), we see virtually no difference at all in the melting profiles. And if we look at the T_{mode} curve (right), we see that they directly overlies one another (80° - 95°), the only slight difference between them lying apart from the main curve, in the region of paralogous DNA (75°).

The percent of hybridization thus introduces a spurious difference between the melting temperatures, if one adopts the T_{50H} statistic. Of course, it could be the case that the difference in the extent of hybridization is consistent and phylogenetically informative. But empirically it is not: We had 15 experiments comparing Human-Chimp, in which the percent hybridization ranged from 90.7 to 102.3 (that is to say, the Human-Chimp DNA hybridized more extensively than the Human-Human control!), with a mean of 96.1%. We had 20 Human-Gorilla experiments, ranging from 90.2 to 97.9, with a mean of 95.2%. There is no statistical difference between them, and as we had expected, introducing the extent of hybridization simply makes the actual DNA comparisons even cruder.

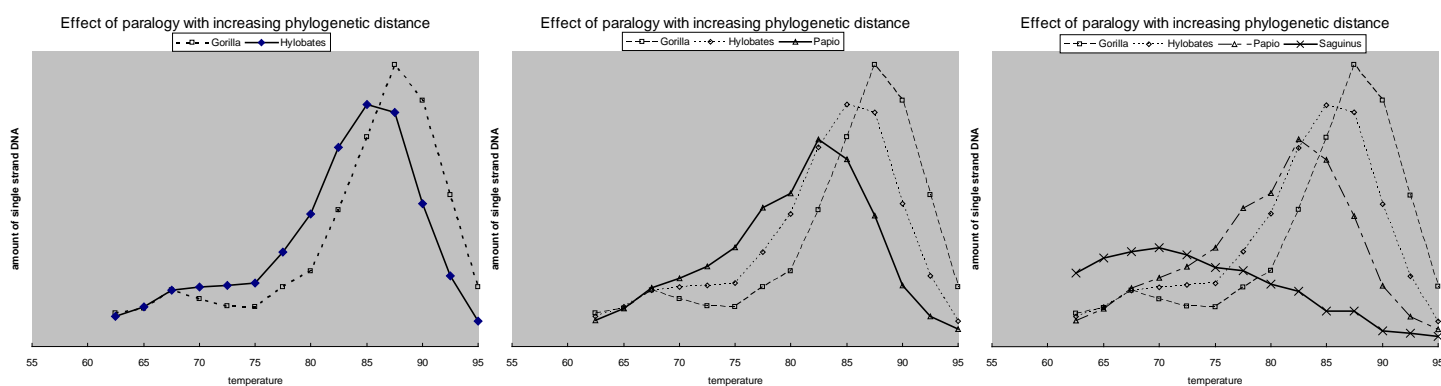
This difference between T_m and T_{50H} will become crucial to understanding subsequent arguments. T_m measures the median melting temperature of the DNA that formed hybrids; T_{50H} measures the median melting temperature of the DNA that conceivably could have formed hybrids. Since T_{50H} incorporates two variables, it should have a greater scatter than T_m ; and the two measures can only coincide if the percent hybridization of the particular experiment is exactly 100%, which is virtually never is.

Thus far, we have established that a small and variable fraction of imperfectly-paired paralogous DNA hybrids could have a significant blurring effect upon the measurement of thermal stability for closely related species. Essentially, the contribution of paralogous DNA to the calculation of a median temperature can simply outweigh any small differences in actual thermal stability of orthologous DNA pairs, which are presumably what the experiment strives to study. For more distantly related species, the same cause produces a different effect.

Consider two species that diverged from one another 10 million years ago. Their orthologous DNA hybrids will encode that divergence. Suppose, however, that the gene duplications within the genome happened, on the average, 100 million years ago. When you incubate the DNA from the two species to allow them to form hybrids, the amount

that hybridizes will be high (since the species are closely related), the orthologs will be very thermally stable (for the same reason), and the orthologous pairings, being roughly *10 times more similar* to one another than the paralogous pairings, will be strongly favored by the thermodynamics of the experiment, so only a small proportion of the hybrid DNA might be expected to be paralogs. Now take two species that diverged 30 million years ago. Not only is there less hybridization, and the orthologous DNA is more divergent, but because the genes themselves still duplicated 100 million years ago, the orthologous DNA is now only *3 times more similar* than the paralogous DNA. In other words, there will be proportionately more paralogous hybrid DNA in the experiment itself, and it is more difficult to distinguish from the orthologous DNA. And for species that diverged 50 million years ago, fairly little DNA hybridizes, the orthologs are quite divergent, and are only favored over the paralogs *by a factor of 2:1*. The melting curve, which began as a large peak of orthologs melting at 86 and separate from a small satellite peak of paralogs melting at 70, becomes progressively lower, wider, and indistinguishable from the paralogs with increasing phylogenetic distance. And that is exactly what the Sibley-Ahlquist data show.

From experiment 1165, using human DNA bound to gorilla and then to gibbon (*Hylobates*), we see a curve that is flatter and wider for the human-gibbon hybrids, reflecting DNA that hybridizes less, is less thermally stable, and competes less successfully against paralogs. Adding the baboon, an Old World monkey, shows the trend continuing; and examining the tamarin, a New World monkey, yields a curve in which the paralogs and the orthologs blend into one another.



What these data show is that DNA hybridization indeed works, but only for a restricted phylogenetic range. For closely related species, the amount of separation is commonly too small in relation to the noise introduced by paralogous DNA to be meaningful; and for distantly related species, the paralogs can't be distinguished from the orthologs.

Correction:Falsification::Freedom-fighter:Terrorist

For the apes at least, we now knew that (1) DNA hybridization should not be able to resolve this problem; and (2) it did not resolve this problem. Indeed, the technique had been applied by other researchers prior to Sibley and Ahlquist without resolving the

problem.¹² So the next question was: How had Sibley and Ahlquist actually gotten it to work so well?

In a third paper on the subject, published in the *Journal of Molecular Evolution* in 1990, Sibley and Ahlquist finally acknowledged the application of three “correction methods”.¹³ The first is a correction for fragment length disparities. The experiment begins with the genomic DNA sheared to an average length of about 500 bp. However, DNA is easier to pull apart at the ends than in the middles, and the smaller the average fragment length, the higher is the ratio of ends to middles. One can correct for this, with an equation derived from the physical chemistry of DNA, if one measures the fragment length.¹⁴ Although Sibley et al. (1990) noted that they had “corrected” some experiments this way, there is no evidence that they actually measured the fragment lengths and could therefore know how much of a correction to apply. That, of course, makes it unlikely *a priori* that the correction was applied in any sense legitimately.

More significant, however, is their own discussion of this correction factor, which we will call Correction #1. Although they claim to have applied this “Correction for Driver Length” only to a handful of experiments, they did not apply it with any consistency. They applied it to the Chimp-Human experiment 785A-10 to reduce the value of 2.7 (much higher than the mean) down to 1.5 (just below the mean) with the explanation that the human DNA “was faulty, apparently [*sic!*] short-stranded.” And yet, the Gorilla-Human experiment 785B yielded a value of 2.2 (right around the published mean for those species) and underwent no alteration in spite of their statement that “the same faulty [human DNA] used in Exp. 785A was also used in Exp. 785B.”

Even worse, the same “correction” was applied to experiment 843-6, to reduce the value of a Chimp-Human experiment from 2.6 to 1.6, according to Table 7 of their 1990 paper.¹⁵ However, in Ahlquist’s correspondence with me, he wrote that that particular experiment “just has a lower percent of hybridization than the others and therefore can be corrected.”¹⁶ But that is actually Correction #3: we haven’t gotten to that one yet.

What I will call Correction #2 is the “Linear Correction” of Sibley et al. (1990). We deduced something like it from the results of Series 1165, which seemed to have an entire run of numbers off by about 1.2° from the published values. They argue, quite possibly correctly, that the labeled DNA samples became degraded over time. They infer that this degradation affects the melting temperature, but most curiously, not of all of the experiments using that degraded DNA. They conclude that only the first experiment in each series – the homoduplex control, from which each particular T₅₀H value will be subtracted to obtain a ΔT – was affected and required correcting. “The correction

¹² Kohne, D. E., Chiscon, J. A., and Hoyer, B. H. (1972) Evolution of primate DNA sequences. *Journal of Human Evolution*, 1:627-644. Benveniste, R. E., and Todaro, George J. (1976) Evolution of type C viral genes: Evidence for an Asian origin of man. *Nature*, 261:101-107.

¹³ Sibley, C. G., Comstock, John A., and Ahlquist, Jon E. (1990) DNA hybridization evidence of hominoid phylogeny: A reanalysis of the data. *Journal of Molecular Evolution*, 30:202-236. Although we had deduced the existence of “correction factors,” they were first mentioned in a letter I received from Jon Ahlquist early April of 1988, and then orally at a hastily-convened meeting on DNA Hybridization in southern California in 1989.

¹⁴ This is the correction used by Caccone and Powell (below), who applied it much more creatively.

¹⁵ Sibley, C. G., Comstock, John A., and Ahlquist, Jon E. (1990) DNA hybridization evidence of hominoid phylogeny: A reanalysis of the data. *Journal of Molecular Evolution*, 30:202-236. Quotations taken from pp. 203 and 204.

¹⁶ Ahlquist to Marks, 21 March 1988.

procedure is to substitute the value of a good homoduplex for the faulty one and to calculate the delta values based on the substituted homoduplex.”¹⁷

The substitution of controls across experiments, however, would be sufficient to flunk an undergraduate science student in most college laboratory courses. One hardly needs to say more.

Once again, however, there is an even worse side to this. Run 1165 included 5 Human-Chimp experiments, 8 Human-Gorilla experiments, 1 Human-Orangutan (1165-14), and 1 Human-Gibbon (1165-15). In fact, the melting curves of 1165-14 and 1165-15 are nearly identical, with the gibbon DNA being slightly more similar to the human than the orangutan. We calculated the ΔT_{50H} values to be 3.1 for Human-Orangutan, and 3.0 for Human-Gibbon. In their 1987 paper, which first allowed us to compare our calculations to theirs for the same experiments, Sibley and Ahlquist gave the results of these experiments as 3.1 for Human-Orangutan and 4.0 for Human-Gibbon, in line both with the established phylogenetic relationships, and with their published mean values – and without comment. In other words, they adjusted every experiment in this series upwards by over 1 degree, and skipped the Human-Orangutan experiment 1165-14, whose value was already in line with the expectation.

In the 1990 paper, they say they adjusted this series upwards by 1.2 degrees, by substituting the control. In a footnote, they add, “1165-14 should have been corrected to 4.3 in Sibley and Ahlquist (1987),” but make no mention of the Human-Gibbon experiment (1165-15), which was altered upwards, and thereby brought into line with the expected value. In other words, they did not actually even change the values for the whole series; they only changed the values within the series that they didn’t like.

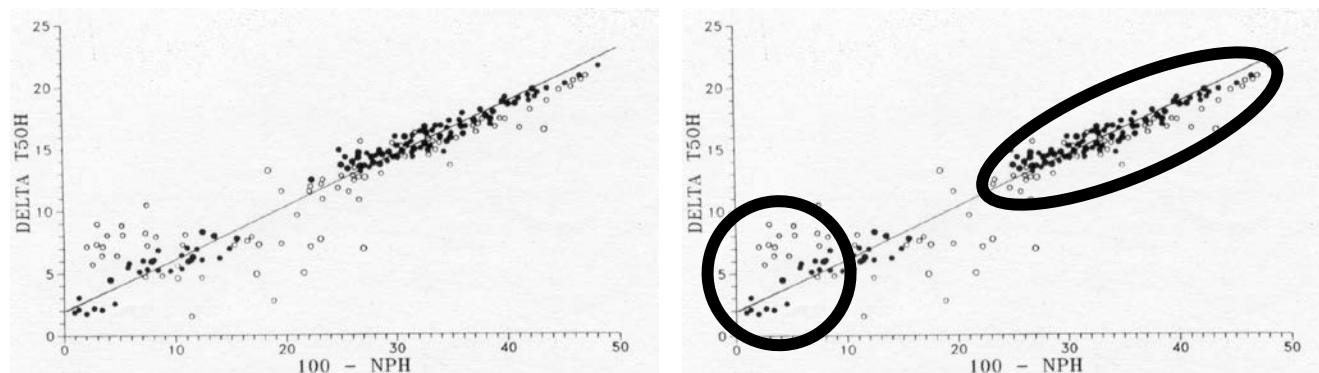
Correction #3 is the one they claimed to have used most extensively, applied to nearly 40% of their experimental values: the “Proportional Correction”. Their justification involves the fact that the percent of hybridization (NPH) is a crude measure of relatedness. They plot the T_{50H} values against the NPH values and find them to be highly correlated (which is trivial, since T_{50H} incorporates NPH), and observe that “aberrant hybrids differ from those that lie close to the line” – although with no other specific criteria for judging an experiment to be aberrant other than the value it produced.

And how do they deal with it? “The correction is made by moving the aberrant point to the linear regression, determining the new NPH value, and calculating the new T_{50H} value.” They do not say whether they move “the aberrant point” horizontally, vertically, or orthogonally. What they do appear to be saying is that they correlated two variables and moved the data into the regression line describing them. Even if this were valid, it would negate all the statistical treatments, which had been assuming the datapoints were independent of one another. The low standard deviations they reported were consequently, at best, nonsense.

And the “even worse”: Sibley et al. (1990) did not even bother to illustrate this point with the primate data, which was the subject of the paper. Rather, they illustrated it with bird data. And the illustration is not particularly straightforward: rather than showing the correlation of T_{50H} and NPH (as actually described in the text), they showed ΔT_{50H} and 100-NPH. Here is their illustration (left). Not only is it difficult to relate to the text, but it is also unduly complex, and not derived from the dataset under discussion. But there is some crucial sense to be made from this illustration, and once again, it does

¹⁷ Ibid., p. 228.

not cast their data analysis in a good light. First, orientation: the points at the upper right are of distantly-related species, and the ones at the lower left are of closely-related species. Now, notice that the correlated points fall into two general groups: the ones tightly clustered around the regression line (X-axis values for 100-NPH ranging from 25-



50), and the ones that seem to be uncorrelated (X-axis values for 100-NPH ranging from 0-25), as grouped in the identical illustration at right. Remember, however, these are bird data, concerning species that are far more distantly related to one another than the Human-Chimp-Gorilla triad under discussion. The well-correlated points hybridized from 50-75%; but Human-Chimp-Gorilla invariably hybridized above 90%. In other words, the only points on the graph that are directly relevant to the issue at hand are those on the extreme left side, where 100-NPH ranges from 0-10. And for those points, there is no relationship between thermal stability and extent of hybridization at all: the scatter indicates a very poor correlation between the two variables in this range. Consequently, even their rationalization for this adjustment makes no sense.

I invoked the most significant conclusion of their 1990 paper as an epigram for this paper, however. “If the corrections had not been used, it is virtually certain that Sibley and Ahlquist would have concluded that *Homo*, *Pan*, and *Gorilla* form a trichotomy.” In other words, their data did not actually bear out their conclusion, as we had suspected from the outset.

Was it fraud? Of course it is beyond our capability to prove that, for it would necessitate establishing an intent to deceive – the *mens rea* or guilty mind, which is virtually impossible to establish. Instead, I will lay out two alternatives, each of which may explain what happened. It is important to recall, however, that there are three exceedingly non-normative, and presumably rare, behaviors co-occurring here.

First, this entire episode really began with the refusal of Sibley and Ahlquist to provide documentation of their work to their peers. We were not the only ones who had asked and been rebuffed; we were simply the ones who pursued it the farthest. Second, if the data “corrections” that drove the conclusions had been described in the papers, it seems unlikely that the papers would have been published in their present form. And even if reviewers had not paused to consider critically the nature of these “data corrections,” certainly readers would have been in a position to do so. The failure to include a crucial part of the data analysis in a scientific paper is quite irregular, and in this case obviously prevented reviewers and readers from being able to assess the validity of the results accurately. In fact, it is even more irregular than refusing to allow interested colleagues to examine the data from a published paper. And third, how many practicing

scientists at respectable institutions do not know that substituting controls across experiments, and moving correlated points into the regression line describing them, are illegitimate operations? Not too many, I should hope.

The best-case scenario, then, would be that these three nested, exceedingly non-normative, and presumably rare behaviors all occurred independently of one another. They sequestered their data, and coincidentally omitted the critical analytic methods from their papers, and coincidentally did not know that what they were doing to the data was illegitimate.

And obviously, the worst-case scenario would be that they sequestered their data, because they had omitted the critical analytic methods from their papers, because they knew that what they were doing to the data was illegitimate. We assumed the work would be readily, formally, and unambiguously identified as scientific fraud by any impartial third party. Our only problem was that we never actually found that third party.

The Best Defense

The boxer Jack Dempsey reputedly was fond of saying that “the best defense is a good offense”. The offensive took shape in 1988-89, and was the most surprising part of this affair, not least of all for its ultimate success.

The controversy surrounding the Sibley-Ahlquist work was certainly no secret. After hearing Vince Sarich discuss it at the meetings of the Society for the Study of Evolution, journalist Bruce Fellman wrote it up in *The Scientist* (“the newspaper for the science professional”) in the issue of 13 June 1988. The 25 August 1988 issue of *Nature* contained an exchange between Jared Diamond and myself, which at least brought the issues, as they stood at the time, out into the public eye.

Diamond had previously written two “News and Views” features for *Nature* extolling Sibley and Ahlquist’s work. After his third essay on their work, “DNA-based phylogenies of the three chimpanzees” (which prefigured his 1992 science bestseller, *The Third Chimpanzee*), *Nature* published a letter from me, which said:

[The Sibley-Ahlquist] study, unfortunately, is still not fully documented. In my experience the melting profiles, per cent hybridizations, melting temperatures and even the analytical procedures on which Sibley and Ahlquist based their conclusions are not available to interested colleagues. Sibley and Ahlquist, since first publicizing their conclusions, have consistently failed to meet the burden of proof which falls upon all investigators. Until that burden is met, it is gratuitous to assume the interpretations are valid, or to draw conclusions from it.

Diamond responded:

While Marks refers to the study of Sibley and Ahlquist as undocumented, these authors gave detailed descriptions of their methods in many earlier papers and presented their hominoid data in two lengthy papers, of which ref. 3 gives 514 DNA/DNA hybridization values. At Marks’ urging, J. Powell and A. Caccone at a recent meeting of the Society or the Study of Evolution redetermined hominoid DNA/DNA hybridization values by a different method, using some samples provided

by Sibley and Ahlquist as well as others obtained with Marks' help, and obtained results concordant with those of Sibley and Ahlquist.

Let us take these two sentences up in turn. The first sentence, obviously, hangs on the assertion that the “values” and the “detailed descriptions of their methods” that Sibley/Ahlquist had presented were accurate, complete, and valid – and we now knew that not to be the case. Since neither of our papers had been published yet, however, the assertion could not be documented, and was only known as hearsay. Obviously Diamond was “in the loop” because his second sentence invokes unpublished results and *ad hominem* assertions (urging? help?) in order to sidestep the fundamental issues.

Evolutionary biologist Jeffrey Powell was a senior faculty member in the biology department at Yale. I had been appointed a junior member of the anthropology faculty in 1987, with a courtesy joint appointment in biology. I knew that Powell and his protégé, Adalgisa (Gisella) Caccone, also did DNA hybridization to determine fruitfly phylogenies, and shortly after I arrived, I visited his laboratory, and I did indeed lay out to him the problems with the Sibley work, and Powell suggested that we redo it. (No urging was necessary.) I did help get some ape DNA,¹⁸ and assumed – quite naively, as it turned out – that I would be a collaborator, and actually participate in the collection and analysis of the data. I did not hear again about the work, however, until it was completed – and then, only indirectly. In the Spring of 1988 I received a phone call from a colleague at another university, who told me that David Pilbeam had told him that Charles Sibley had told him that Jeff Powell had told him that Sibley's work had been completely “vindicated” by Powell. Clearly, there was a different loop, which I was out of.

Now, the journal *Science* took an interest. Their deputy news editor, Roger Lewin, contacted us to tell us he was going to do a story on the controversy. Since our two papers had not yet been published, we tried to convince him that a news story was premature, but Lewin insisted on doing the story immediately, with our participation or without it. Lewin had, in fact, been touting Sibley's work in *Science* on a nearly annual basis. Even worse, I had just published a somewhat critical review of his book *Bones of*

¹⁸ I do not actually recall how much help I gave in them in acquiring DNA samples. A two-page document from Powell and Caccone titled “Summary of Sources of Primate DNA” and dated 25 May 1988 does not mention me. I have two drafts of their podium talk from the 1988 Evolution meetings, both dated 7 June 1988. The earlier one (containing typos corrected in the other draft) says: “We have used independently obtain [sic] DNA as well as DNA generously supplied by Professor Sibley.” The corresponding part of the second draft, the one they circulated, says: “With the help of Jon Marks in our Anthropology Department, we have obtained fresh tissue from which we extracted DNA. In addition Professor Sibley generously sent us samples of the DNA used in his studies.” I take that to be the source of Diamond's assertion of my involvement with the work. The paper that they ultimately published in *Evolution* in 1989 acknowledges me as being “also helpful in obtaining samples and in encouraging us to undertake this project” and for “comments and helpful suggestions concerning the manuscript”. My comments, dated 5 November 1988, ran to more than 7 single-spaced pages, of which the most salient was to urge Powell to take his head out of Sibley's noose, in almost exactly those words. Retrospectively, from the way my relationship to the work was subsequently invoked, I suspect that I was only mentioned to pre-emptively undermine any criticisms I might raise.

Contention in *The Journal of Human Evolution* and he was known to be very unhappy about it.¹⁹

Lewin ended up publishing a two-part article in September 1988, reproducing illustrations from the (yet-unpublished) *Cladistics* paper without permission, and taking up more than six full pages in the leading science journal in the U.S.²⁰ Lewin made it clear that indeed, Sibley's data (1) had not been available, and (2) had been subjected to unreported manipulations whose nature was still unknown and whose existence had only been discovered serendipitously.

Even worse, Vince Sarich had put Lewin on to Fred Sheldon, who had done his graduate work with Sibley at Yale, and was now a post-doc at the Philadelphia Academy of Sciences. Sheldon had been invited to co-author the second (1987) paper by Sibley and Ahlquist, but had declined – again, a very unusual act for a young scientist, whose rational interests lie in developing a publication record. When Lewin inquired of Sibley and Ahlquist about their data, he was told it was impossible.

Even if Sibley had been willing to give Marks the human/ape data, for practical reasons he would have found it difficult to comply. Not only had he and Ahlquist recently left Yale for two different universities, thus engendering the organizational confusion that such a move typically entails, but also the data themselves were in a state of some disarray. "Our method of recording and working with the data had switched several times, as the Yale computer system changed, first in one way then another," says Ahlquist. "The data were in several different forms, and, when Marks asked for them, in several different places." In short, a mess.

In cases of possible scientific fraud, this is the equivalent of "the dog ate my homework." They were consequently able to assure Lewin, when asked what proportion of the numbers had actually been altered, that it was far less than the 40% that our sample seemed to indicate. "Sibley and Ahlquist told *Science* that the figure was more like 20%," said Lewin. But Fred Sheldon had actually retained the data, and was therefore able to answer the question authoritatively. "Sibley agreed to *Science*'s request to have Sheldon run the human/ape data through a T50H program, so that a comparison could be made. The result is that a little more than 40% of Sibley and Ahlquist's published T50H numbers have apparently been altered by 0.5° or more," wrote Lewin.

So all of our charges as of September, 1988 were substantiated. The data had indeed been withheld; unreported transformations had been applied; reviewers and readers had been misled as to their existence, nature, and effects. And where Sibley and Ahlquist's word could be independently assessed, it turned out to be false. Lewin had scrupulously avoided using the word "fraud", as had we.

¹⁹ Lewin, R. (1984) DNA reveals surprises in human family tree. *Science*, 226:1179-1182. Lewin, R. (1985) Molecules vs. Morphology: Of Mice and Men. *Science*, 229:743-745. Lewin, R. (1987) My close cousin the chimpanzee. *Science*, 238:273-275. Lewin, R. (1988) Molecular clocks turn a quarter century. *Science*, 239:561-563. Marks, J. (1988) Review of *Bones of Contention* by Roger Lewin. *Journal of Human Evolution*, 17:267-270. In fact I had sent Lewin a draft of the review as a professional courtesy, and he had responded by attempting to persuade the *Journal of Human Evolution*'s book review editor, Lawrence Martin, not to publish it.

²⁰ Lewin, R. (1988) Conflict over DNA clock results. *Science*, 241:1598-1600. Lewin, R. (1988) DNA clock conflict continues. *Science*, 241:1756-1759.

Nevertheless, Lewin also presented a defense of Sibley that consisted of three elements. First, mistakes were made, but no fraud was committed. The work was now “sloppy” and “embarrassing” – although these were quite different sorts of adjectives than had been claimed for the work only months earlier, and which had nevertheless recently managed to get Sibley elected to the National Academy of Sciences. (This is the “incompetence defense” successfully mounted by Thereza Imanishi-Kari slightly later, in which extraordinary levels of ineptitude are invoked to deflect the charges of dishonesty, which come with much steeper penalties.) Second, Lewin invoked Jeffrey Powell’s work as validating Sibley’s work. We pointed out to him in the phone interviews that Powell’s work was irrelevant to evaluating the integrity and honesty of Sibley’s, especially since Powell’s work was itself unpublished. (The same argument a few months later made more sense to John Horgan at *Scientific American*, who deleted mention of the unpublished Powell work in his article.²¹) Third, Lewin discredited us (or rather, allowed Sibley to discredit us, through him) with the accusation of ungentlemanly behavior and ulterior motives. A callout quote from Roy Britten actually said that our two unpublished manuscripts “are not scientific articles, they are weapons with political purposes.” And so, in spite of having validated our accusations, Lewin’s parting shot was directed at us: “But the very combative and partisan tone with which the challenges have been made has not advanced Sarich and his colleagues’ stated concern with scientific integrity.”

The last sentence actually stung, for if anything, we felt as though we had bent over backwards to be decorous. All we had wanted to do was to see whether the data actually supported the conclusions. We had tried to keep the focus as much as possible on the data – its availability, analysis, and interpretation. And after being unsuccessful through cordial and informal channels, we wrote and published a criticism of the work, which included the serendipitous discovery that the work in question appeared to have passed peer review and entered the scholarly literature essentially under false pretenses. Is there a nicer way to accuse someone of fraud?

Powell’s Impossible Claims

Caccone and Powell’s work on the apes was finally published in 1989,²² but quite extraordinarily for a public debate that turned on melting curves, their paper also presented none. The technique was different, in that it did not rely upon elution through hydroxyapatite, but instead took place in a solution of tetraethyl ammonium chloride, or TEACl. TEACl attempts to correct for an error that we hadn’t even bothered with: base composition. Since G-C pairs in DNA are held together by three hydrogen bonds, and A-T pairs are held by only two, a stretch of DNA (or a genome) that is GC-rich will dissociate at a higher temperature than one that is A-T rich, with the same extent of base-pair mismatch. TEACl equilibrates this.

In this method, the DNA of one species is sonicated to a mean of about 500 bp, then labeled and hybridized to that of another species. Instead of studying a single sample as the temperature is gradually raised, this method takes aliquots of the sample, stabilizes them at different temperatures, digests any unpaired DNA enzymatically, and

²¹ Horgan, J. (1989) Time bomb: War breaks out in the field of evolutionary biology. *Scientific American*, (March): 24-25.

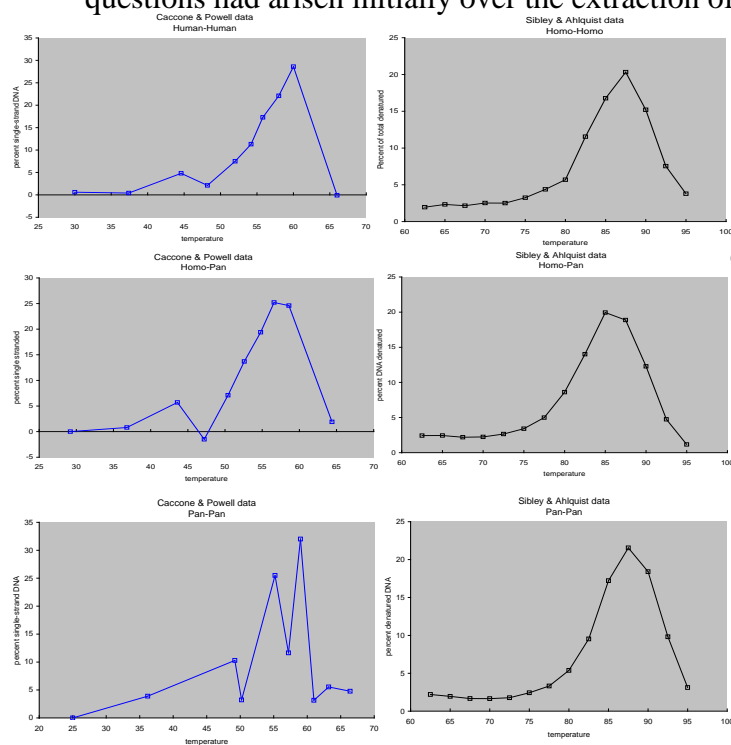
²² Caccone, A., and Powell, Jeffrey R. (1989) DNA divergence among hominoids. *Evolution*, 43:925-942.

analyzes that free, labeled DNA as the amount of single-stranded DNA present at that particular temperature. The data are presented as a T_m , rather than as a T_{50H} . I strongly advised Powell to keep his claims conservative, and I found it difficult to understand why he made the claims he did. The claims were (1) that their data resolved the phylogeny into human-chimp, as Sibley had claimed; (2) that the values on which they based their claim matched Sibley's values; and (3) that the Sibley work was thereby validated. Caccone and Powell called particular attention to "the remarkable congruence of our ΔT_m 's and Sibley and Ahlquist's (1984, 1987) ΔT_{50H} 's," and reiterated that "our data are so similar to Sibley and Ahlquist's." Obviously, it was hard to miss the connection they were establishing, as they had been establishing it long before actually publishing their work.

The most telling part of the story concerns their extended discussion of the artifacts affecting the Sibley/Ahlquist work. "Since we obtained results nearly identical to those of Sibley and Ahlquist when we studied the same taxa, it would seem that these possible technical problems do not significantly affect the results."

The point I made to Powell is that his claim to match Sibley's numbers is illogical, since ΔT_m s don't match ΔT_{50H} s except at exactly 100% hybridization; and moreover, we don't even know what Sibley's real numbers were, because it was clear that the published numbers weren't them.²³ And indeed, the difference between the real numbers and the published numbers was the heart of the matter for Sibley and Ahlquist. I was also interested to see just what Powell's melting curves looked like, since the questions had arisen initially over the extraction of a single number from a curve.

Powell never did share any of his melting curves with me, but he gave three to Vince Sarich – which enabled us to see whether Powell's raw data were in fact of higher quality than Sibley's. If Sibley's data did not in fact resolve the phylogeny of the apes, then Powell's would presumably have to be of considerably superior quality in order to do so. Here are the three melting curves we saw from Powell's work, paired up against comparable melting curves from the limited Sibley dataset we had. On the left of each pair is the Caccone/Powell melting curve, and on the right is the corresponding Sibley/Ahlquist experiment. The three comparisons are Human-Human DNA, Human-



Chimp DNA and Chimp-Chimp DNA. Two things are easily seen. First, because TEACl weakens the G-C bonds, the entire melting curve is shifted to a lower temperature in the Caccone-Powell work, while the melt takes place over a comparable range. And

²³ Marks to Powell, 5 November 1988.

second, the Sibley-Ahlquist curves are uniformly more symmetrical, and more unimodal. Not only do they seem to have less in the way of paralogous DNA at low temperatures, but in all the experiments we saw, we never saw anything as weird-looking as the Chimp-Chimp melt from the Caccone-Powell series (lower left).

This may explain why Powell chose not to publish his melting curves, when these had been the centerpiece of the debate from the outset. Nevertheless, it raises a bizarre question, to wit: If the Caccone-Powell data are not as good quality as the Sibley-Ahlquist data, how can they possibly produce more robust conclusions?

Even worse, what can it possibly mean that Powell's ΔT_m s match Sibley's altered ΔT_{50H} s? Surely – unless you believe it is justifiable to move correlated points into the regression line describing them without clear criteria or methods, subsequently treat them as independent data points, and occasionally substitute controls across experiments – it would be more appropriate to compare them to Sibley's unaltered ΔT_m s! Those numbers were not in the public domain when Caccone and Powell published their paper, but were included in the 1990 paper by Sibley, Ahlquist, and Comstock. And unsurprisingly, those numbers neither resolve the phylogeny, nor match the Caccone-Powell numbers.

The first column gives the 1987 Sibley-Ahlquist values, on which the phylogenetic argument hinged. *Homo-Pan* are the genetically most similar pair, with *Gorilla* being symmetrically more distant. The means and standard deviations give statistically significant results. The Powell numbers (middle column) are similar to the published Sibley numbers for all three pairs, except that they exaggerate the separation of the gorilla, an additional 0.2 degrees away.

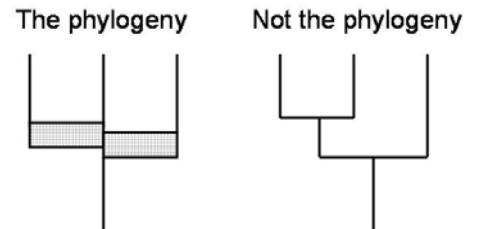
	Sibley ΔT_{50H} (altered)	Powell ΔT_m (claimed match)	Sibley ΔT_m (real)
<i>Homo-Pan</i>	1.6 ± 0.2	$1.59 \pm .16$	1.4 ± 0.8
<i>Pan-Gorilla</i>	2.3 ± 0.2	$2.55 \pm .24$	1.7 ± 0.4
<i>Homo-Gorilla</i>	2.3 ± 0.2	$2.50 \pm .10$	1.8 ± 0.8

Caccone-Powell interpreted this result throughout the paper as validating the work of Sibley-Ahlquist (“our results are so similar to those previously obtained by Sibley and Ahlquist”), but the logic of the claim is elusive. The numbers shouldn't match, and the structure of the phylogeny is irrelevant to evaluating the merits of the Sibley-Ahlquist data. And their results can't reasonably be interpreted to mean that it is now okay to do the things that Sibley and Ahlquist were now claiming they did, because that would violate elementary standards of laboratory protocol and data analysis, and would necessarily call into question one's own standards.

When compared to the unaltered ΔT_m values given by Sibley et al. (1990), two things become clear. First, the Sibley values do not resolve the phylogeny; and second, aside from being in the same ballpark, they don't match the Powell numbers at all. The scatters are much wider and the means are not much different from one another.

This is actually a rather important point. The real Sibley numbers are very consistent with the bulk of molecular data, which show the split among the three genera to be “too close too call”, with sampling errors at least as large as the separation (or internode) itself. In other words, the most consistent feature of the molecular data is not how distinct the gorilla is, but how similar it is to human-chimp, and whether the resolving power of the technique in question is adequate to see the separation among them, which is very close. The original (1984) paper found a separation of 0.4 degrees; in the face of criticism, this separation grew to 0.7 degrees in the 1987 paper, and was now (in Powell's hands) over 0.9 degrees, or over 60% of the total human-chimp value.

In finding the gorilla so distinct, the Powell data actually are discordant from the general consensus of the data, indeed discordant even from the information encoded in the real Sibley data. While the two diagrams at the right may share a similar apparent branching sequence, in fact they encode very different bio-historical narratives. On the left is a demographically complex story entwining microevolutionary and macroevolutionary processes – the one in evidence in nearly all of the molecular data bearing on the problem.²⁴ On the right is a model invoking successive simple bifurcations, according to Caccone-Powell.



The key to making sense of the Caccone-Powell work is that it came out before the 1990 paper by Sibley, Ahlquist and Comstock, which acknowledged and described the “corrections,” and which gave the uncorrected numbers. Caccone-Powell therefore did not have access to their presentation of the bizarre, ad hoc adjustments, nor to the acknowledgment that without them the phylogenetic “story” evaporated.

Somehow Caccone and Powell managed to produce a match to a series of non-comparable, falsified numbers. But how and why?

The “how” is actually quite simple, and can be easily seen in their published paper. As noted earlier, DNA strands can dissociate at different temperatures partly because of differences in fragment length. Although Sibley-Ahlquist did not bother much (if at all) with this, this correction is the core of the Caccone Powell work. The data in their paper make it possible to calculate the average ΔT_m values before and after correction for fragment length, to gauge their effect.

And indeed, it quickly becomes clear that the correction for fragment length introduces the phylogenetic resolution and the match to the published Sibley-Ahlquist numbers.

	Before	After
Human-Chimp	2.4 ± 1.1	1.6 ± .2
Chimp-Gorilla	3.0 ± .6	2.6 ± .2
Human-Gorilla	4.1 ± 1.1	2.5 ± .1

Like the unaltered Sibley numbers, the Powell numbers before adjustment are asymmetrical, are not easily interpretable phylogenetically, and have a high statistical scatter. After adjustment, the numbers are symmetrical and phylogenetic and are congruent with the Sibley numbers. The fragment-length adjustment introduces the phylogenetic information into the mean values, introduces the concordance with the Sibley-Ahlquist numbers, and reduces the standard deviations by up to an order of magnitude.

In practice, the fragment-length adjustment was used to change one Human-Chimp value from 4.10 to 1.36, and another from 4.64 to 1.39. The difference between the two experimental results was therefore changed from 0.54 to 0.03 degrees. The reported fragment length of the first experiment was 99 bp; for the second, it was 90.

²⁴ Which indeed geneticists periodically rediscover. Fischer, A., Wiebe, V., Paabo, S. and Przeworski, M. (2004) Evidence for a Complex Demographic History of Chimpanzees. *Molecular Biology and Evolution*, 21:799-808. Kumar, S., Filipowski, A., Swarna, V., Walker, A. and Hedges, S. B. (2005) Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proceedings of the National Academy of Sciences, USA*, 102:18842-18847. Patterson, N., Richter, D. J., Gnerre, S., Lander, E. S. and Reich, D. (2006) Genetic evidence for complex speciation of humans and chimpanzees. *Nature*, 441:1103-1108.

This raises the obvious question of whether such a number could be determined so precisely, and if so, how. The Caccone-Powell paper does not indicate how it would be possible to discriminate between 99 and 90 in a mixture of genomic DNA, following sonication, long-term incubation, and enzymatic digestion. If the extraction of a single number from what ought to be a “smear” of DNA has such a critical effect on these experiments, then we probably have a simple answer to the methodological question of just how Caccone and Powell managed to produce ΔT_{ms} that replicated the falsified Sibley ΔT_{50Hs} .

The second question is more difficult: Why did they bother? Or to put it more anthropologically, Why would they see their interests to lay in defending Sibley? Indeed, a few years later Powell would claim that the Sibley-Ahlquist work had “been repeated in my own laboratory” and “is good science inasmuch as it is repeatable and independently corroborated.”²⁵

The answer was clarified for me in 1990, in a slightly different context. The DNA hybridization community was never very large, but young scholars could apply the technique to produce quantitative, precise, genomic answers to some range of phylogenetic questions. But in the hands of an expert, it could produce quantitative, precise, genomic answers to any phylogenetic question, and the answers would be sufficiently mystified as to be largely untraceable. In June of 1990, I was asked to peer-review a long chapter on marsupial DNA hybridization for an edited volume. I wrote a three-page review recommending cuts to the least well-documented parts and better documentation of other parts, especially when the authors were drawing conclusions by distinguishing between a ΔT_{50H} of 21 degrees and a ΔT_{50H} of 22 degrees.

But rather than make the suggested revisions, the senior author indignantly withdrew the paper, to the astonishment of both the editor and me. That was the point at which I began to realize what was at stake. DNA hybridization was essentially a circle of friends, reviewing each other’s grants and manuscripts, producing data that was in some senses both authoritative and unchallengeable; and although often it wasn’t necessary, the data could be, and often were, readily massaged. In brief, you could use it to generate unimpeachable evidence for any plausible phylogenetic hypothesis, and the only people who might be able to impeach it, weren’t going to ask the questions. We had written early on, “We look forward to a time when the normal standards of scientific rigor will be applicable to DNA hybridization,”²⁶ without fully grasping the implication of a science that resisted quality control. The practitioners had every reason to be very angry at us: We were killing a goose that had been laying golden scientific eggs for them. DNA hybridization was effectively Paul Feyerabend’s nightmare:²⁷ a science in which indeed, “anything goes” in the collection, presentation, and interpretation of research.

I articulated the methodological, epistemological, and simply logical problems with the claims of the Caccone-Powell work in the *American Journal of Physical Anthropology* in 1991. I made the point again in 1994, for the last time in the primary literature, with an analogy I hoped would be resonant: “It would be, after all, one thing to

²⁵ Powell, J. R. (1993) Reviewing misconduct? *American Scientist*, 81:408.

²⁶ Marks, J., Schmid, C.W. and Sarich, V.M. (1989) Response to Britten. *Journal of Human Evolution*, 18:165-166.

²⁷ Or daydream? Feyerabend, P. (1975) *Against Method*. London: Verso.

find a new fossil with a human cranium and an orangutan's jaw, but quite another to claim that it thereby validated Pliocene Man."²⁸

Fallout: My Career as Whistleblower

In my readings I had come upon a letter published in *Science* in 1954 by the geneticist Conway Zirkle. Zirkle had been struck by the fact that decades after Paul Kammerer killed himself, a new generation had rediscovered his work on the inheritance of acquired characteristics – either ignorant of, or deliberately overlooking, the fact that the nuptial pads on Kammerer's midwife toad's forelimb were actually ink injections. Zirkle was vexed by a problem that has never been satisfactorily resolved: How do you get bogus work out of the literature?²⁹

Ann Gibbons, who had replaced Roger Lewin as *Science's* news writer on human evolution, wrote in 1990:

DNA hybridization is ... tricky to do, tricky to interpret, and less precise than having the actual sequences. Partly as a result, the Sibley and Ahlquist results were subjected to a withering bombardment from Marks and Vincent Sarich of the University of California at Berkeley. Marks and Sarich questioned their methods of data analysis and even charged that Sibley and Ahlquist had falsified data.³⁰

We had avoided the words “fraud” and “falsification” in our primary publications, of course, but it was clear to us that they had taken place – quite egregiously! – and would be readily identified as such by anyone sufficiently interested to have a look at the actual work, particularly in relation to the claims made on its behalf. By now, the technique of DNA hybridization had devolved into being doubly “tricky” – but more significantly, the outstanding charge of data falsification was there in black-and-white in the leading science journal in America. It seemed as though nothing more needed to be said for the “wheels of justice” to begin turning. Yet they didn't.

In 1993, I was asked by *The Journal of Human Evolution* to review Jared Diamond's book, *The Third Chimpanzee*. Noting that the book's “hook” was based on the Sibley-Ahlquist work, which Diamond was still touting uncritically, I said:

Perhaps you recall Sibley and Ahlquist. In a nutshell, their results were: (1) chimp-gorilla DNA hybrids were more thermally stable than chimp-human hybrids; (2) the differences were insignificant; and (3) reciprocity was very poor when human DNA was used as a tracer. Unfortunately, the conclusions they reported were: (1) chimp-human was more thermally stable than chimp-gorilla; (2) differences were significant; and (3) reciprocity was near-perfect. And they got from point A to point B by (1) switching experimental controls; (2) making inconsistent adjustments for variation in

²⁸ Marks, J. (1991) What's old and new in molecular phylogenetics. *American Journal of Physical Anthropology*, 85:207-219. Marks, J. (1994) Blood will tell (won't it?): A century of molecular discourse in anthropological systematics. *American Journal of Physical Anthropology*, 94:59-80, quotation from p. 66.

²⁹ Zirkle, C. (1954) Citation of fraudulent data. *Science*, 120:189-190.

³⁰ Gibbons, A. (1990) Our chimp cousins get that much closer. *Science*, 250:376-376.

DNA length, which was apparently not even measured; (3) moving correlated points into a regression line; and (4) not letting anyone know. The rationale for (4) should be obvious; and if (1), (2) and (3) are science, I'm the Princess of Wales. This work needs to be treated like nuclear waste: bury it safely and forget about it for a million years.³¹

Our first paper on the subject had also been published by *The Journal of Human Evolution*, and what I said – especially in light of its sarcastic tone – was sufficiently well-known as to be common knowledge on the field. Daniel Dennett would later quote that book review in *Darwin's Dangerous Idea*, to liken me to Bishop Wilberforce, Thomas Huxley's creationist nemesis.³²

At about the same time, I was invited by *American Scientist* to review four books on scientific fraud, which they published as a Lead Review in their issue of July-August 1993. I was quite interested in scientific fraud, of course, but I was a biological anthropologist. The earlier book review I had written for them was on human evolution, not scientific fraud. Quite possibly they had some reason to believe I was knowledgeable about the subject of scientific fraud.

I concluded my review with a discussion of power in the Sibley business, and its lack of resolution – all the details of which had already been extensively discussed in the primary and derivative literature. We knew everything we were ever going to know about the Sibley work, and yet:

[i]n spite of the fact that the language used in the published papers wildly misled reviewers and readers as to the nature of the data analysis and the robustness of the conclusions, we have not seen a retraction of the work, an inquiry by the National Academy into the nature of the research of its member, nor a public repudiation of the work by the senior community of molecular evolutionists. Rather, those who have spoken out have taken the position that the data alterations were a bad idea (although not bad enough to jeopardize his membership), their absence from the previously published work was a sad coincidence, the subsequent withholding of documentation was a crying shame, and the serendipitous discovery of the data alterations by others in 1988 was a thorough embarrassment.³³

Although the review had been vetted by two levels of editors, and it said nothing that wasn't already widely known, Sibley threatened them with litigation, and prevailed upon them to publish six letters in his defense – from Sibley and Ahlquist themselves, Britten, Powell, Morris Goodman, and two others. These made the same points already noted, and in addition (paradoxically) speculated quite freely on my own motives for launching a “personal attack” on Sibley, from under the very noses of the editors. One letter, quoted at the beginning of this paper, made the utterly false claim that the “corrections” hadn't made a difference. Jeff Powell claimed that Sibley's work was

³¹ Marks, J. (1993) Review of *The Third Chimpanzee* by Jared Diamond. *Journal of Human Evolution*, 24:69-73.

³² Dennett, D. C. (1995) *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Schuster, p. 338.

³³ Marks, J. (1993) Scientific misconduct: Where 'Just Say No' fails. *American Scientist*, 81:380-382.

“good science” – and that his replication had made it so. Roy Britten clumsily tried to make the Sibley-Ahlquist “correction” sound better by simply restating it in the passive voice:

Dr. Marks charges that Dr. Sibley moved “correlated points on a scatter plot into the regression line describing them.” This is untrue. The fact is that regression-line plots were used to detect outliers and make corrections on runs with large or small extents of hybridization....³⁴

Vince Sarich pointed out to the editor, “[Y]ou’re missing a big story here. Sibley and Ahlquist have perpetrated perhaps the biggest fraud in the history of science, and it is fascinating to see so many falling all over themselves to keep that fraud under wraps,” but to no avail. The editor privately had written contritely to Sibley, “you will find that I have severely limited Dr. Marks’s reply.... Jonathan Marks, by the way, will not be reviewing books for *American Scientist*.”³⁵

Sibley bit on the National Academy issue, though. “Dr. Marks urges the National Academy to conduct an investigation into our alleged crimes against science. We shall suggest such an investigation to the National Academy of Sciences Home Secretary.” Vince Sarich and I quickly wrote (separately) to Peter Raven, Home Secretary of the National Academy of Sciences, with the specific charges and documentation.³⁶ With both sides ostensibly wanting an inquiry, you might think the National Academy would undertake one to see if a member had indeed been elected on the basis of largely falsified research. Here is the full text of Raven’s letter to me, dated 25 October 1993:

Dear Dr. Marks:

Thank you very much indeed for your letter and the enclosures. I was extremely interested in what you had to say in reading the enclosures. It is obviously a very complex case and, as I am sure you understand, the National Academy of Sciences would not undertake to conduct a formal review of the activities of its members as a matter of general principle, lacking the judicial machinery to do so properly. I would add, however, that no one is elected to the Academy for a single piece of work, and thus it is incorrect, as a matter of principle, to say that “this is the work that ultimately resulted in Sibley's election to the National Academy of Sciences.....”. In summary, I was very interested in the material that you sent. We will be conducting no investigation.

Yours sincerely,
Peter H. Raven
Home Secretary

³⁴ Britten, R. J. (1993) Reviewing misconduct? *American Scientist*, 81:408.

³⁵ Vincent Sarich to Rosalind Reid, 21 September 1993. Rosalind Reid to Charles Sibley, 31 August 1993, kindly sent to me by Ralph Estling. *American Scientist* is published by Sigma Xi, which also distributes a booklet on “Honor in Science”. The booklet does not take up the issues of retaliation or blacklisting.

³⁶ Sibley, C. G., and Ahlquist, Jon (1993) Reviewing misconduct? *American Scientist*, 81:407-408. A British writer became convinced, after corresponding with Sibley, that his call for an investigation to clear his name was entirely disingenuous. Ralph Estling to Vincent Sarich, 11 October 1993.

The National Academy of Sciences had also awarded Sibley and Ahlquist their Daniel Giraud Eliot medal in 1988. Perhaps they didn't want it back.

Jeff Powell's *American Scientist* letter also contained the suggestion of a "self-serving or vindictive motivation" on my part, and of my having leveled "false accusations". This is itself an actionable accusation of misconduct, and troubled me, especially since Powell was a senior professor at my institution. In 1991, as I came up for reappointment to Associate Professor (untenured or "on term", a Yale thing), Powell turned up on the Term Appointments Committee, and did not recuse himself. My department chair hastily recruited the support of the Director of the Peabody Museum for the meeting, and the promotion went through without incident. I had, in fact, managed to develop an impressive record of research and scholarship in my free time away from the Sibley business! But with an accusation of possible misconduct directed at me, I consulted my senior colleagues about pressing Yale to clear my name by actually adjudicating the Sibley case.

Fortunately, they talked me out of it. Yale, of course, was in a strange position vis-à-vis Sibley and Ahlquist, since neither man was still at Yale – so there was nothing disciplinary at stake, aside from some possible needless self-flagellation. If NSF, which had funded the research, decided that they wanted their money back, there might be a reason for Yale to investigate, but NSF didn't seem to have any interest either. And I was savvy enough to appreciate that although some Yale administrators were protecting me, the more I rocked the boat, the less inclined they would probably be to continue.

I emailed John Yellen, the NSF Anthropology program director, on 25 July 1994, to inquire if there was indeed any interest in the case on their part, especially since Sibley had acknowledged NSF grants in the compromised articles. Yellen responded that Anthropology had indeed co-funded the research, and that he had forwarded my note to Peggy Fischer in the office of the Inspector General, and added, "If you wish to proceed with this issue, you should contact her directly." I thanked him and said, "I don't have any interest in personally pressing it with NSF, but it does seem to me that with all that has already come out in the literature, it's a bit surprising that the issue hasn't pressed itself. I am certainly happy to supply anyone with the relevant literature and correspondence, and provide any assistance they may want."

And that was the end of that.

Unfortunately, with fraud – and especially with white-collar fraud – one often also commonly faces cover-up and retaliation. In this case, there were two ultimate consequences. Without any formal investigation by any adjudicating body, the work would still be "out there" and thus could stay open as an interpretive debate. In science, however, one would ideally like to distinguish clearly between "You say po-tay-to, I say po-tah-to" and "You say po-tay-to, but I don't see any potato". The second consequence was that funding and ink indeed quickly dried up for DNA hybridization, and it is now simply not done or written about much any more, despite the advantages it appeared to offer rhetorically in the 1980s. In that sense, "the system" worked.

I did suffer one bit of successful retaliation at Yale. I began my appointment in anthropology at Yale with a nominal "Courtesy Joint Appointment" in the biology department. In 1992, as I was being reappointed, the biology department set a precedent by (very discourteously) revoking my "Courtesy Joint Appointment" – a purely symbolic gesture that even stunned the department secretary, who called me to apologize!

Morris Goodman, a senior member of the molecular evolution community, was invited to co-author a review article with me on ape genetics by the editor of *Current Opinion in Genetics and Development* in 1991, but refused, citing his friendship with Charles Sibley. I asked him to take a good hard look at the claims and counterclaims, and to consider the implications of defending fraud in molecular evolution, after having worked so long to legitimize the field. He would not. I began to appreciate that I would have to develop more eclectic professional interests, as I probably did not have much of a future in molecular evolutionary studies.

Charles Sibley died in 1998, somewhat sullied, but nevertheless a member in good standing of the National Academy of Sciences. His biographical memoir notes,

One could argue that the methods of data analysis were not as rigorous as they might have been, and there were certainly differences of opinion among the members of Charles' own research group on how best to quantify and summarize the data, but that does not constitute fraud.³⁷

No, of course it doesn't. But that hardly gets to the heart of the matter. It wasn't merely the systematic alteration of the data to make "outliers" into "inliers" and thereby to produce spurious statistical distinctions in the data. It was also the failure to acknowledge or explain it, the impediments to other scholars interested in evaluating the work, and the entirely serendipitous discovery that the alterations had in fact been made. The data did not in fact say what Sibley and Ahlquist said, and were just crudely and clandestinely changed to make it appear as though they did. That would constitute fraud.

I will give Jon Ahlquist the last word, from a memoir he published of Sibley.

Publication of our results immediately generated opposition. It came from a variety of sources and was fueled by the acquisition of some of our poor-quality data by those who immediately claimed fraud.... [T]hese very data found their way into the hands of our antagonists and were publicized *without peer review* as fraud and bad science. In retrospect, the phrase "bad science" was little more than thinly veiled euphemism for character assassination and a specific political agenda. The matter could have been resolved with a civil phone call asking "How did you guys analyze these data, anyway?" Instead, a trial was carried out using innuendo, tabloid journalism, and licentiousness – that would make the host of a TV talk show blush. There was empty talk about "truth," while any attempt at rational discussion was shouted down [emphasis in original].³⁸

Conclusions

³⁷ Brush, A. H. (2003) Charles Gald Sibley, 1917–1998. *National Academy of Sciences, Biographical Memoirs*, 83:1-24. The same sentences appear in: Corbin, K. W. and Brush, A. H. (1999) In Memoriam: Charles Gald Sibley, 1917-1998. *The Auk*, 116:806-814.

³⁸ Ahlquist, J. (1999) Charles G. Sibley: A Commentary on 30 Years of Collaboration. *The Auk*, 116:856-860. Quotation from p. 858.

1. Whistle-blowing is ugly business. Don't get involved in it. It's like a wart on the nose of science: Nobody wants to know about it, nobody wants to look at it, and if you force them to, it will only make them angry at you.

2. If you do blow a whistle, have your guns drawn and blazing from the outset. There is no nice way to do it, and you will be accused immediately of being nasty. You might as well be. Although we tried to talk about the data as much as possible, the return fire involved free speculations about our underlying motives for doing so. This was sometimes accompanied by a dismissal of our criticisms, on the ostensible grounds that we were merely arguing *ad hominem*. I think I learned that practicing scientists cannot readily distinguish between strong criticism and a personal attack.

3. The exposure of fraud is worse than the commission of fraud. The latter is perceived as a blip – the insane uncle locked away in the attic – but the former involves calling undue attention to that blip.

4. As a whistle-blower, you are a prosecuting attorney without powers of subpoena or a judge to compel compliance on the part of the defendants. You nevertheless must have the complete case to present before coming forward.

5. There is widespread confusion, which is easily exploited, between the common law injunction of the presumption of innocence in the face of criminal charge, and the obligation that falls upon a scientific investigator to meet a burden of proof with rigor and transparency. Although it can coincide with legal proceedings, scientific fraud is not itself illegal. Making a false claim in a scientific paper is not the same thing as making a false claim under oath, which is perjury. Claiming to meet a burden of proof, but not actually meeting it, is not necessarily in itself either Fabrication, Falsification, or Plagiarism. It is, however, directly relevant to the critical evaluation of the conclusions, which is in turn central to the process of science.

6. Failure to meet the burden of proof may be caused by error, rather than fraud. Because of the lesser penalties associated with error, the suspicion of fraud will invite a plea-bargain, that the observed discrepancies do not constitute deceit, but are “merely” mistakes. This is surely a post-modern moment, when a scientist can be defended with the charge of gross incompetence. The difference, however, lies in the ability to establish an intent to deceive, and is rarely worth the effort.

7. All scientists potentially involved in controversies, as social creatures, make cost-benefit calculations about their allegiances. It is in nobody's interest to get involved with a fraud case; nor is it in anyone's interests to align themselves against power. After the story became public, various senior biologists privately shared with me their suspicions that Sibley was “dirty”. But the first generation of molecular evolutionists had come up through the ranks together and had worked hard, both scientifically and rhetorically, to establish the legitimacy of the field. In fact, Allan Wilson and Vince Sarich had harbored suspicions about the quality of Sibley's work in relation to the stridency of his claims, and Sarich carried on a cordial but frustrating correspondence with Sibley about the work in the mid-1980s. But when the evidence became available, Sibley was able to recruit many first-generation molecular evolutionists to share a “siege mentality”, much as David Baltimore would a few years later. Be prepared to go it alone; fortunately, I didn't have to.

8. A song from the musical “Chicago” advises, “When you're in trouble, go into your dance.” One of the most sobering things for me was the ease with which “the other

side” tried, and often succeeded, to deflect attention away from what we thought were the central issues – the meaning of the Sibley-Ahlquist data, and its apparent misrepresentation by the authors. Whether it was Roy Britten attempting to legitimize a bizarre and recondite “correction procedure” by restating it in the passive voice, or Jeff Powell’s impossible claim that he had “matched” Sibley and Ahlquist’s altered numbers using a different statistic, or John Kirsch avowing that the data alterations were irrelevant even though Sibley had already admitted the opposite, we felt as though we were striving for transparency largely in vain. And as much as we were participating in a debate over the nature and quality of the molecular evidence pertaining to ape phylogeny, we were always a step behind the documentation. We were told that our initial manuscript was unnecessary because Sibley and Ahlquist would soon (1987) be publishing a paper that would render our criticisms moot. Then we were told that Sibley’s undocumented work was verified by Powell’s unpublished work. And when Powell’s work was finally published, it lacked precisely the same documentation – the melting curves – that had started the row over the Sibley work. Maryellen Ruvolo, an early and consistent defender of Sibley, wrote in 1995: “the DNA hybridization data still support a closest association between *Homo* and *Pan* (work in preparation by Caccone, Powell, myself, and others).”³⁹ But no such work subsequently appeared, and Powell himself had already quietly abandoned the technique. The same year, David Pilbeam wrote, “I do not discuss the Sibley et al. (1990) data set here, although I will do so elsewhere, except to note that it is of poorer quality than that of Caccone and Powell.”⁴⁰ Likewise, over a decade later, the promised discussion has not appeared. The irony is that Pilbeam had been among the first to promote the DNA hybridization work, and was consistently incapable of evaluating its quality properly – which effectively replicated his mistake on *Ramapithecus* vis-à-vis the molecular clock a decade earlier, but from the other side of the evolutionary fence.

9. In science, the ends cannot justify the means. Good science has only two attributes – competence and honesty – and these concern methods, not results. The idea that questions about scientific integrity can be mooted by the results of another body of work is not only ridiculous, it is mildly offensive. David Baltimore raised that line of argument as a smoke screen and was curtly rebuffed by the biochemist Paul Doty: “The scientific literature would become irredeemably corrupted if this became accepted practice. The essential standard is that the evidence presented in a scientific paper is the bedrock on which interpretations and conclusions are built. If this connection is violated so that speculations drawn depend on subsequent investigations to prove them right or wrong, then the reporting of research would be reduced to a lottery.”⁴¹ But Sibley’s supporters were to a considerable extent successful at muffling the fraud issue with the argument that there really is a consensus of genetic data supporting human-chimp. “[T]he problem of hominoid phylogeny can be confidently considered solved,” wrote Maryellen Ruvolo in a 1997 meta-analysis, conspicuously citing the Powell work in lieu of the Sibley work without discussion, and with Jeff Powell acting as “reviewing

³⁹ Ruvolo, M. (1995) Seeing the forest through the trees: Replies to Marks; Rogers and Commuzzie; Green and Djian. *American Journal of Physical Anthropology*, 98:218-232.

⁴⁰ Pilbeam, D. (1996) Genetic and morphological records of the Hominoidea and hominid origins: A synthesis. *Molecular Phylogenetics and Evolution*, 5:155-168.

⁴¹ Doty, P. (1991) Baltimore’s unanswered questions. *Nature*, 353:495.

editor”.⁴² Morris Goodman chooses the opposite route to the same destination, blithely continuing to cite the Sibley work as if it were fine.⁴³ But if you take the view that scientists have the freedom to do anything they like as long as it gives them the correct answer, there are two consequent questions. First, the reliability of your own work is implicitly impeached; do you hold your own work up to a higher standard than Sibley’s? And second, how real is the consensus – are there rhetorical constraints on presenting it, or does the principle that “anything goes” apply here as well? Obviously consensus can look more convincing if data are presented or criticized selectively.

10. At lastly, talk about an “old boys’ network”! Charles Sibley’s friends were the gatekeepers of molecular evolution. And whether it was Emile Zuckerkandl overturning the decision of his reviewers and associate editor to publish our paper; or having the *American Scientist* publish half-a-dozen letters defending Sibley’s egregious data falsification; or just ensuring that the matter would not be subjected to formal adjudication – it was sobering to experience such institutional clout in action. Successful fraud cases are generally pressed against junior and relatively powerless researchers; but Sir Cyril Burt, for a notable example, was actually able to get away with it.

⁴² Ruvolo, M. (1997) Molecular phylogeny of the hominoids: Inferences from multiple independent DNA sequence data sets. *Molecular Biology and Evolution*, 14:248-265.

⁴³ Goodman, M., Grossman, L. I. and Wildman, D. E. (2005) Moving primate genomics beyond the chimpanzee genome. *Trends in Genetics*, 21:511-517. Goodman died in 2011.