Pleistocene Speciation and the Mitochondrial DNA Clock

John Klicka and Robert M. Zink (*1*) used pairwise mitochondrial DNA (mtDNA) distance data and a 2% per million year (My) "mtDNA clock" to examine whether Late Pleistocene (≤250,000 years ago) glaciations may have been an important mechanism of speciation in North American songbirds. They conclude that sequence divergence values and corresponding estimates of times of evolutionary divergence between presumptive sister pairs of North American songbirds were sufficiently large to reject a Late Pleis-tocene Origins model (LPO model, *2*) for most species. Furthermore, they conclude that the majority of North America's "young-est" species of songbirds originated in the Late Pliocene or Early Pleistocene, which would suggest that Pleistocene glaciation, in general, did not play an important role in shaping patterns of speciation in this group. These conclusions are not supported by the data in the report (*1*).

There are three major problems with the report by Klicka and Zink: (i) the authors apparently assumed that dates of divergence can be accurately estimated by dividing ob-served mtDNA divergence values, uncorrected for saturation (superimposed substitutions) (*3*), by an uncorrected rate (2% per My) of mtDNA evolution; (ii) they did not provide a measure of error associated with their esti-mated dates; and (iii) they did not provide a test of their implicit assumption that a molec-ular clock holds for their data.

In order to estimate dates of divergence from DNA sequence distance data under the assumption of a molecular clock, the number of substitutions that have occurred since two sequences diverged must be estimated under an appropriate model of nucleotide substitu-tion. This applies to the taxonomic group for which dates are to be estimated (that is, North American passerines) and also for the group or groups on which the rate calibration (the "clock") is based. Otherwise, because ob-served sequence divergence does not accu-mulate linearly over time, the rate of substi-tution will be underestimated, and estimated dates of divergence will be biased. Only when all of the distances are adequately corrected for superimposed substitutions will the effect of time (and saturation) be factored out (*4*).

To illustrate this point, we obtained Klicka and Zink's original cytochrome *b* se-quence data [available for 21 of the 35 spe-cies pairs they examined; sources "d," "k," and "l" in table 1 of (*1*)] (*5*). We then used likelihood ratio tests (LRTs) to determine the best-fit model of nucleotide substitution for the 21 species represented by at least 1000 base pairs (bp) from the cytochrome *b* gene [source "k" in table 1 of (*1*)] (*6*). The best fit model, the gamma-HKY85 (*6*), was then used to correct for superimposed substitu-tions for all 21 species pairs for which cyto-chrome *b* data were provided (*5*).

We then used this same approach to de-termine the best fit model and to correct for superimposed substitutions in the cytochrome *b* gene of two groups for which a 2% per My mtDNA clock has been proposed: primates (great apes plus humans), the group on which the original 2% per My mtDNA clock was based (*7*), and galliform birds (chicken *Gal-lus gallus* plus the partridge genus *Alectoris*) (*8*), one of the avian groups cited by Klicka and Zink as having a 2% per My rate of mtDNA evolution. With the use of published estimates for the date of divergence between chimpanzee and human (*7*), and between chicken and partridge (*8*), we were able to estimate a corrected rate of substitution for the cytochrome *b* gene in these two groups and, in the spirit of Klicka and Zink's study, use the corrected rate of substitution to re-estimate dates of divergence for the 21 species pairs of North American songbirds for which cytochrome *b* data were made available (*5*).
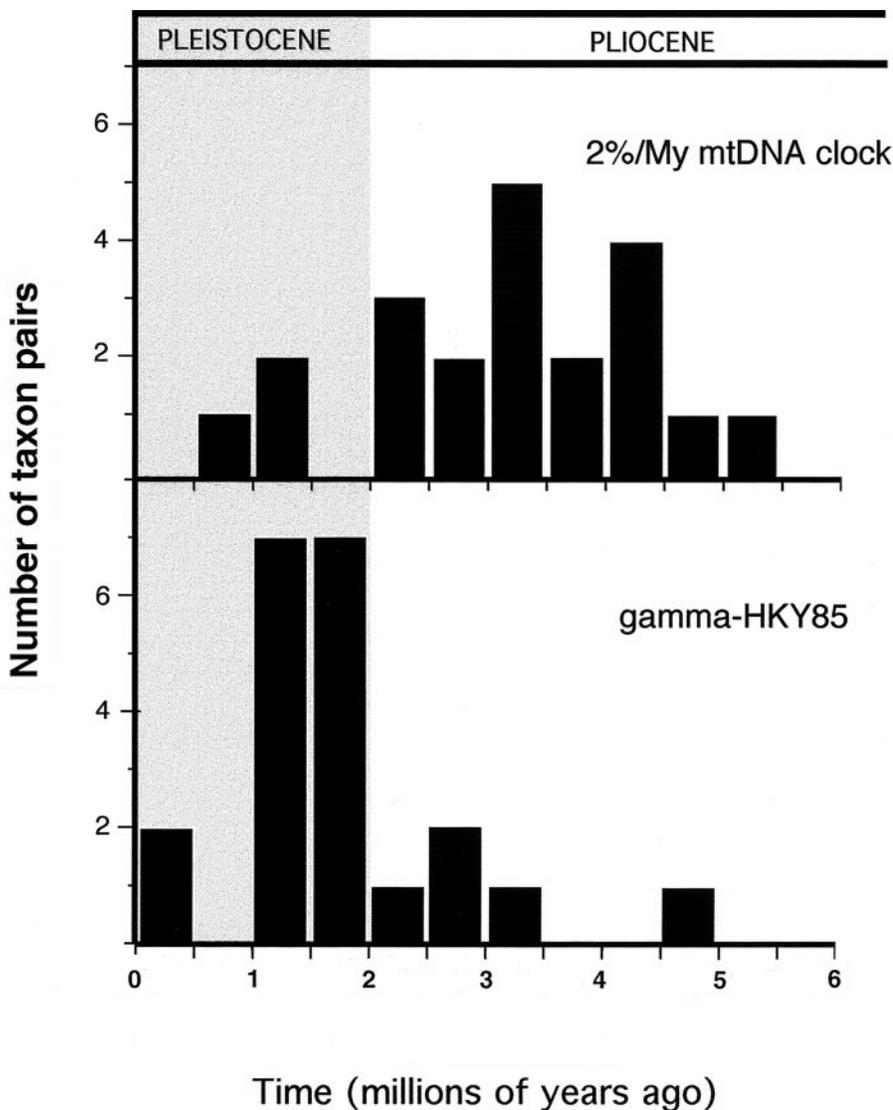
For primates and galliform birds the rate

Fig. 1. Frequency distributions of estimated dates of mtDNA divergence for 21 presumptive sister pairs of North American songbirds based on: (**top**) uncorrected mtDNA distances (*5*) and a 2% per My mtDNA clock; and (**bottom**) corrected values of mtDNA sequence divergence estimated under the best fit gamma-HKY85 model and a corrected rate of substitution for the cytochrome *b* gene of 0.025 substitutions per site per My (*9, 16*). Dates based on uncorrected values of sequence divergence are considerably younger than those based on uncorrected values (*17*).

of substitution estimated for the cytochrome *b* gene under the best-fit model (the gamma-HKY85 model in each case) was 0.0278 and 0.0252 substitutions per site per lineage per My, respectively, or more than 2.5 times faster than the rate predicted by a 2% per My mtDNA clock (0.01 substitutions per site per lineage per My) (*9*). This increase in the estimated rate of substitution is directly related to saturation. The best fit model uses a gamma distribution to incorporate among-site rate heterogeneity and predicts that the actual number of substitutions that have occurred since the divergence of chimpanzees and humans, and since the divergence of chicken and *Alectoris,* is much greater than the number that can be directly observed, and considerably more than the number predicted by models of nucleotide substitution that do not address among-site rate heterogeneity (*10*).

Correcting for saturation under the gamma-HKY85 model has a large effect on estimated dates of divergence and temporal patterns of speciation in North American songbirds (Fig. 1). Whereas a scenario based on uncorrected distances and a 2% per My mtDNA clock would suggest that 86% of the 21 avian species pairs diverged before the Pleistocene, corrected distances and a corrected rate of 0.025 substitutions per site per lineage per My predict that 76% of these species pairs have mtDNA divergences estimated to have occurred within the Pleistocene (*11*). Furthermore, because mtDNA haplotypes diverge within a common ancestral population before species are formed, species are younger than the ages estimated for the coalescence of their haplotypes (typically by several hundred thousand years in birds) (*12*). Therefore, based on our corrections, a Pleistocene origin can be safely ruled out for only a few of the 21 species pairs for which cytochrome *b* data were made available; the majority of avian species pairs would have divergences in the early-to-middle Pleistocene, a time of major glacial activity in North America (*13*). Approximately 10% of these divergences would fall within the last few hundred-thousand years.

The second problem we address is that Klicka and Zink do not provide a measure of error associated with their estimated dates of divergence. The error inherent in estimating dates of divergence using a molecular clock tends to be quite large (*14*). For example, based on the confidence limits depicted by Hillis *et al.* (*14*) for dates of divergence estimated by regression from the original primate 2% per My mtDNA clock (*7*), none of the estimated dates of divergence reported by Klicka and Zink would have a lower 95% confidence limit that did not include zero years (that is, the present). Therefore, even if we were to overlook the issue of saturation and accept Klicka and Zink's estimated dates,

there would still be little statistical support for rejecting the LPO model.

Finally, Klicka and Zink's conclusions depend on their assumption that a molecular clock holds for their data, yet they did not provide a test of this assumption. We used a likelihood ratio test (LRT) to test for a molecular clock in the 21 species of North American songbirds represented by ≥1000 bp of cytochrome *b* data. The results call for rejection of the molecular clock hypothesis (*15*). Therefore, even if Klicka and Zink had addressed the issues of saturation and error outlined above, their study would still not be valid because the songbird sequences they examined are not evolving in a clock-like manner.

Accurately estimating dates of divergence from molecular data is, at best, a challenging process. Saturation, error, and differences in the rate of molecular evolution among lineages must be addressed before strong biological conclusions can be drawn from evolutionary dates based on molecular clocks. Although the LPO model may not accurately reflect temporal patterns of speciation in North American songbirds, it cannot be rejected on the basis of the report by Klicka and Zink.

Brian S. Arbogast
Department of Biology,
Wake Forest University,
Winston-Salem, NC 27109, USA
E-mail: arbogbs5@wfu.edu

Joseph B. Slowinski
Department of Herpetology,
California Academy of Sciences,
San Francisco, CA 94118–4599, USA

References and Notes

1. J. Klicka and R. M. Zink, *Science* **277**, 1666 (1997).
2. We use Klicka and Zink's (*1*) definition of the LPO model throughout our paper.
3. Contrary to note 14 in the report (*1*), percent sequence divergence values shown in their table 1 are not corrected for superimposed substitutions under any model of nucleotide substitution (J. Klicka, personal communication).
4. J. B. Slowinski and B. S. Arbogast, *Syst. Biol.*, in press.
5. J. Klicka and R. M. Zink (*1*) provided their cytochrome *b* data for our re-analysis. These data represented a total of 33 species and 21 presumptive species pairs. In a few cases, the sequence-divergence values presented in table 1 in (*1*) were incorrect as a result of errors in the original data matrix (J. Klicka, personal communication). These errors were subsequently corrected by J. Klicka and R. M. Zink; our analyses are based on the amended data set provided to us by those authors. Sources "e," which consisted of only 288 bp of cytochrome *b* sequence, and "m," which is unpublished data, are not Klicka and Zink's original data, and were not provided to us by the authors. Therefore, we did not include these two sources of data in our analyses.
6. For all phylogenetic analyses, we used a β-tester version of the program PAUP* 4.0 [D. L. Swofford, *PAUP*: Phylogenetic Analysis Using Parsimony (*And Other Methods)*, Version 4.0 (Sinauer, Sunderland, MA, 1998)]. This program estimates pairwise sequence divergence between taxa and infers phylogenetic trees under parsimony, maximum-likelihood, and distance criteria. Use of LRTs to determine best-fit models of nucleotide substitution and to test for a

molecular clock followed the procedures outlined by J. Huelsenbeck and B. Rannala [*Science* **276**, 227 (1997)]. For each data set, we evaluated the fit of ten models, five models each evaluated with and without among-site rate variation incorporated via a discrete gamma distribution with four rate categories as described by Z. Yang [*Mol. Biol. Evol.* **10**, 1396 (1993)]. These models were "JC" [T. H. Jukes and C. R. Cantor, in *Mammalian Protein Metabolism*, H. N. Munro, Ed. (Academic Press, New York, 1969), pp. 21–132], "F81" [J. Felsenstein, *Evolution* **35**, 1229 (1981)], "K2" [M. Kimura, *J. Mol. Evol.* **16**, 111 (1980)], "HKY85" [M. Hasegawa, H. Kishino and T. Yano, *ibid.* **22**, 32 (1985)], and "GTR" [C. Lanave, G. Preparata, G. Serio, *ibid.* **20**, 86 (1984); Z. Yang, *ibid.* **39**, 105 (1994)]. Because it is unrealistic to assume that any nucleotide site is wholly invariant [Z. Yang, *Mol. Biol. Evol.* **10**, 1396 (1993)], all sites were assumed to be potentially variable. The majority of these models require direct sequence data, which prevented us from re-analyzing RFLP data presented in table 1 of (*1*). In order to reduce sampling error and to avoid among-gene rate heterogeneity, we used only those 21 species of North American passerines represented by ≥1000 bp of cytochrome *b* sequence data to determine a best fit model and to test for a molecular clock (*14*). The gamma-HKY85 model (-Ln likelihood = 6779.1872; α = 0.20826) provided the best fit to these data with the fewest parameters (*11*).
7. W. M. Brown, M. George Jr., and A. C. Wilson, *Proc. Natl. Acad. Sci. USA* **76**, 1967 (1979); W. M. Brown *et al.*, *J. Mol. Evol.* **18**, 225 (1982)]. We use the same calibration point for the chimpanzee-human divergence as these authors (about 5 My B.P.).
8. We used the cytochrome *b* data presented by E. Randi [*Mol. Phylogenet. Evol.* **6**, 214 (1996)] and the estimated date of divergence for chicken-*Alectoris* of about 17 My B.P. presented in K. M. Helm-Bychowski and A. C. Wilson [*Proc. Nat. Acad. Sci. USA*, **83**, 688 (1986)].
9. Corrected rates of substitution were estimated by $\lambda = d_{HKY85}/2t$, where d_{HKY85} is the estimated value of pairwise-sequence divergence under the best fit model, and *t* is the divergence date for two lineages. The corrected rate of substitution estimated for primates was based on complete cytochrome *b* sequences for the great apes (*Homo sapiens, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus*, and *Hylobates lar;* NCBI accession numbers X93334, X93335, X93347, X97717, and X99256, respectively). The phylogeny that we recovered by maximum-likelihood was identical to the well-established phylogeny for the group: ((((*H. sapiens, P. troglodytes*) *G. gorilla*) *P. pygmaeus*) *H. lar*). The gamma-HKY85 model (-Ln likelihood = 3337.70770; α =0.211427) was determined to provide the best fit to these data with the fewest parameters. On the basis of a calibration point of 5 My B.P. (*7*) for the chimpanzee-human divergence, λ = 0.02774 substitutions per site per lineage per My. The gamma-HKY85 model (-Ln likelihood = 3140.48705; α = 0.080588) also provided the best fit to E. Randi's (*8*) cytochrome *b* data for chicken, *Gallus gallus*, and partridges, *Alectoris* spp. The phylogeny that we recovered via maximum-likelihood had a topology identical to that presented by Randi (*8*). Based on the average corrected pairwise distance between species of *Alectoris* and chicken (corrected under the best fit gamma-HKY85 model and an estimated date of divergence between chicken and *Alectoris* of about 17 My B.P. (*8*), the rate of substitution in the cytochrome *b* gene for this group is approximately 0.0252 substitutions per site per lineage per My.
10. D. L. Swofford *et al.* [in *Molecular Systematics*, D. M. Hillis, C. Moritz, B. K. Mable, Eds. (Sinauer, Sunderland, MA, 1996), pp. 407–514] review the assumptions and performance of various models of molecular evolution. Comparisons of models that assume there is no variation in the rate of substitution among nucleotide sites versus those that assume that substitution rates follow a gamma distribution suggest that the former may often underestimate the real number of substitutions that differentiate two DNA sequences. The gamma-HKY85 model provided the best fit to the North American passerine, great ape,

and chicken-partridge data sets that we examined in this study (in each case, the more parameter-rich gamma-GTR model provided a slightly better fit to the data, but the improvement was not significant based on LRTs). The low value of α estimated under the best fit gamma-GTR model for each data set (6, 9) indicates that there is a highly skewed distribution of rates of substitution in the cytochrome b sequences we examined. In other words, a few nucleotide sites within each rate category are evolving quite rapidly, but most sites are evolving relatively slowly. This means that a few nucleotide sites are likely to receive superimposed substitutions very soon after divergence. Because the equal-rates models (which assume α = infinity) used to estimate the various 2% per My mtDNA clocks cannot address this type of among-site rate variation, they will often underestimate the true number of substitutions that have occurred since two sequences diverged, which in turn leads to an underestimate of the true rate of evolution.

11. For our re-estimates of dates of divergence, we used all 21 species pairs for which cytochrome b data were available. Dates of mtDNA divergence (t) were calculated as $t = d_{HKY85}/2\lambda$, where d_{HKY85} is the estimated value of sequence divergence between species pairs under the best fit model for the 21 species represented by $\geq$1000 bp of cytochrome b data (gamma-HKY85; α = 0.20826) (6).

12. S. V. Edwards, in *Avian Molecular Systematics and Evolution*, D. P. Mindell, Ed. (Academic Press, New York, 1997), pp. 251–278.

13. C. B. Cox and P. D. Moore, *Biogeography: An Ecological and Evolutionary Approach* (Blackwell, Oxford, ed. 5, 1993).

14. D. M. Hillis *et al.*, in *Molecular Systematics*, D. M. Hillis, C. Moritz, B. K. Mable, Eds. (Sinauer, Sunderland, MA, 1996), pp. 515–543. When using a molecular clock to estimate dates of divergence there

are two important sources of error that must be addressed: (i) expected error in the measurement of molecular divergence, and (ii) error associated with predicting dates of divergence from a regression-based rate calibration (that is, the residual of regression). The former, although often large, is trivial compared to the residual error of the regression. However, Klicka and Zink only reported mean estimated dates of divergence and did not address either of these sources of error. Since a regression of avian cytochrome b divergence on time is not available (the calibrations cited in note 10 of the report cannot be used to perform such a regression because they are based on various gene regions of the mtDNA and are not comparable), we used the confidence limits depicted by Hillis *et al.* for the original 2% mtDNA clock to emphasize that even if the other problems we address are ignored, the mean dates of divergence presented by Klicka and Zink would be highly imprecise.

15. We tested for a molecular clock by using an LRT to compare the likelihood scores for these data under the gamma-HKY85 (α = 0.20826, transition:transversion = 3.7135), with and without enforcing a molecular clock. The results of the LRT were highly significant in rejecting a molecular clock (-Ln likelihood with clock enforced = 6934.3154, without clock enforced = 6779.1872; χ^2 = 310.26, degrees of freedom = 19; $P < 0.0001$).

16. We used a conservative estimate of 0.025 substitutions per site per lineage per My for the rate of cytochrome b evolution under the gamma-HKY85 model. For each of the 21 species pairs we examined the corrected cytochrome b divergence and estimated time of mtDNA divergence (in My B.P.) are shown parenthetically: *Polioptila melanura-P. nigriceps* (6.72%; 1.34), *Polioptila melanura-P. california* (6.56%; 1.31), *Piranga olivacea-P. ludoviciana* (8.14%; 1.63), *Passerina*

cyana-P. amoena (9.58%; 1.92), *Passerina cyana-P. versicolor* (9.35%; 1.87), *Sialia sialis-S. mexicana* (6.73%; 1.35), *Cardinalis cardinalis-C. sinuatus* (13.9%; 2.78), *Calcarius lapponicus-C. mccownii* (12.9%; 2.60), *Calcarius lapponicus-C. ornatus* (15.1%; 3.02), *Oporornis philadelphia-O. toimiei* (2.26%; 0.450), *Oporornis philadelphia-O. agilis* (11.4%; 2.28), *Spizella pallida-S. breweri* (8.76%; 1.75), *Pheucticus ludovicianus-P. melanocephalus* (5.76%; 1.15), *Cyanocitta cristata-C. stelleri* (24.8%; 4.96), *Pipilo aberti-P. crissalis* (6.13%; 1.23), *Pipilo aberti-P. fuscus* (9.50%; 1.90), *Toxostoma rufum-T. longirostre* (9.95%; 1.99), *Toxostoma lecontei-T. redivivum* (6.53%; 1.31), *Toxostoma lecontei-T. crissale* (7.78%; 1.56), *Toxostoma redivivum-T. crissale* (6.47%; 1.29), *Toxostoma bendirei-T. cinereum* (1.24%; 0.248).

17. The dates of divergence presented in Fig. 1 emphasize the impact of saturation on estimates of dates of evolutionary divergence. Even the dates based on corrected values of mtDNA sequence divergence rely on the assumption that the songbird sequences are evolving in a clock-like manner, which does not seem to be the case.

18. We thank J. Klicka and R. M. Zink for providing their data matrix and subsequent corrections. We thank an anonymous reviewer for helpful remarks about testing for a molecular clock and evaluating estimated dates of divergence. We thank D. J. Anderson, R. A. Browne, J. W. Demastes, S. V. Edwards, M. S. Hafner, H. Hoekstra, K. A. Kron, M. Sanderson, A. J. Schwandt, T. A. Spradling, P. O. Lewis, B. R. Riddle, M. D. White, D. Zelmer, and C. Zeyl for their assistance in the preparation of this manuscript. We also thank D. L. Swofford for permission to publish results based on the β-tester version of the computer program PAUP* 4.0.

Response: In our report (1), we challenged the conventional notion that a previously defined set of North American songbird (order Passeriformes) species pairs originated as a consequence of being isolated during the last one [100,000 years before the present (B.P.)] or two (250,000 years B.P.) cycles of North American glaciations (2). Mitochondrial DNA (mtDNA) sequence divergences calculated for 35 such pairs of sister species differed on average by 5.1%. This value is an order of magnitude greater than the amount of divergence expected of species that originated within the last 250,000 years (the Late Pleistocene as we defined it). On this basis, and two other lines of evidence (3), we rejected the prevailing model of "Late Pleistocene Origins" (LPO) for this particular group of birds. We are gratified that Arbogast and Slowinski's "reanalysis" supports our main conclusion. Even accepting their recalibration, for the moment, 19 of the 21 (90%) species pairs that they examined diverged over 1 million years ago [figure 1 (bottom) of the comment]. There are, however, problems with their analysis.

Our original conclusions were derived from three main assumptions regarding North American songbirds of recent origin: (i) that rates of sequence evolution are constant (clock-like) among species pairs; (ii) that uncorrected molecular distances provide a reasonably accurate measure of molecular evo-

lution; and, (iii) the 2% per My rate (4) of evolution is a reasonable divergence rate (5) for this set of species. Arbogast and Slowinski contend that these assumptions are biased such that our study fails to provide a test of the LPO. We disagree.

The first challenge raised by Arbogast and Slowinski concerns our assumption that our sequences are evolving in a clock-like manner. They purport to demonstrate that our data show among-taxon rate heterogeneity, thus invalidating our study. In fact, the highly significant likelihood ratio test (LRT) that Arbogast and Slowinski report is an erroneous result of enforcing the molecular clock assumption on an improperly rooted phylogeny. When rooted correctly the results of the LRT for a molecular clock are not significant (6), indicating that the assumption of homogeneous rates among taxa is valid for this data set (7, 8). Thus, their conclusion that among-taxon rate heterogeneity negates our test of the LPO is spurious.

A second disagreement concerns the time at which significant levels of saturation (multiple substitutions over time at the same base position) occur. The maximum likelihood model Arbogast and Slowinski used (gamma-HKY85) is presumed to estimate more accurately the true number of substitutions that have accrued between two DNA sequences since they diverged from a common ancestor. For example, the average uncorrected diver-

gence for the three *Passerina* buntings in our common data set is 6.63%. The gamma-corrected divergence estimate for these same three closely related species averages 8.76%, an increase of 32% due entirely to putative saturation effects. As another example, the human-chimpanzee observed distance of 11% has a gamma-HKY85 distance of 27.4%. Thus, the method used by Arbogast and Slowinski would indicate that saturation is substantial, if not enormous, even at relatively low levels of divergence (9).

This conclusion, however, conflicts with empirical evidence. Of the mtDNA distance estimates obtained in our study, 94% (33 of 35) differed by less than 10%. Most studies suggest that saturation would not bias rate calibrations until uncorrected sequence divergences exceed this value (10) and all plots of avian mtDNA genetic distances (by codon position) versus time, of which we are aware, are linear within this range. For avian cytochrome b data, the evidence suggests that "progress toward transition saturation accelerates between 10% and 18% divergence" (11). More compelling, a plot of distances derived from cytochrome b (α = 0.22) versus those for a nuclear intron (β-fibrinogen intron 7, α = 0.89) for woodpeckers (order Piciformes) is linear to approximately 13% (12). Because noncoding introns are thought to be relatively less biased measures of time, this correlation is strong support for the linear

relationship of cytochrome *b* divergence and time at the evolutionary level we considered. In contrast, the maximum likelihood model of Arbogast and Slowinski suggests significant non-linearity by 5% sequence divergence. The discrepancy between the empirical evidence and their conclusions is likely due, in part, to the maximum likelihood model that they used overestimating saturation, at least at low levels of molecular divergence (*13*).

A third issue concerns calibration—"setting" the clock. Arbogast and Slowinski derive a "universal" vertebrate substitution rate of roughly 5% per My from primate and galliform data sets. Arbogast and Slowinski did not justify the extrapolation of substitution and rate parameters derived from older and unrelated taxa onto recently evolved songbirds (*14*). That primates display significantly heterogeneous rates of mtDNA evolution has been established elsewhere (*8*). Sequences not evolving in a clock-like manner would seem to be a questionable source with which to calibrate a general vertebrate clock. The partridge (*Alectoris*)-chicken (*Gallus*) calibration also does not inspire confidence. Curiously, Arbogast and Slowinski used the original partridge data of E. Randi (*15*), but not the age of the fossil that Randi, after examination of all available data, considered correct for calibration purposes. It would seem that Arbogast and Slowinski chose, from among a range of potential values (8 to 20 My), the "fossil" date that yielded a substitution rate most similar to the one obtained for primates. The date that they did choose to represent the time of partridge-chicken divergence (17 My B.P.) in fact represents an indirect "provisional estimate" (*16*) that was obtained from restriction mapping of nuclear genes. Calibrations based on other molecular markers are generally considered inappropriate (*17*).

If possible, calibrations should be derived from within the group of organisms for which they are used (*17*). Arbogast and Slowinski do not mention the only relevant calibration available (*18*), that of the Hawaiian honeycreepers. This study has relevance to our work in that (i) it considers songbirds of similar body size and generation length (*19*), (ii) these species have recent origins, and (iii) the calibration dates (emergence times for three different islands) are recent and well established. With the use of cytochrome *b* sequence data and similar analytical methods (a maximum likelihood model with a gamma correction) to obtain divergence estimates, Fleischer *et al.* (*18*) obtained a substitution rate of 0.008 per site per lineage per million years (1.6% per My), a rate very different from those (over 5%) derived from primates and fowl by Arbogast and Slowinski. This songbird calibration suggests that the

plot of divergence values [figure 1 (top) of the comment] would be pushed slightly to the right (older), not to the left as Arbogast and Slowinski's reanalysis [their figure (bottom)] would indicate (*20*). In sum, the difference in the two histograms (figure 1 in the comment) stems from (i) recomputed mtDNA distances corrected for saturation, and (ii) a calibration of these distances based on primates/fowl. Both aspects are not correct.

Arbogast and Slowinski note that stochastic error associated with a molecular clock may be nontrivial. A general regression of separation times on sequence divergence for birds is lacking for the reasons they suggest. Although regression error values are typically large, this is, in part a statistical artifact resulting from an inadequate number of calibration points (that is, accurate fossil dates). We agree that more independent and recent fossil calibrations are needed, but this discussion detracts from our main focus on songbird diversification during the most recent 250,000 years. It is difficult to envision a plausible clock correction that would compress 5% sequence divergence into the last 250,000 years. However, a relevant regression would be constructed using songbird divergences. In their study of Hawaiian honeycreepers, Fleischer *et al.* (*18*) compared gamma-corrected cytochrome *b* distances with island emergence times in a regression analysis. The tight linearity of their plot (Mantel matrix *r* = 0.995, *P* = 0.018) implies the existence of a predictable rate of molecular evolution in recently evolved songbirds with a higher degree of precision than Arbogast and Slowinski recovered by using the primate regression (*21*).

The conclusions that can be supported by the reanalysis in the comment differ little from our own. Both analyses falsify the LPO model of speciation, and both support (see figure 1 of the comment) our contention that (*1*, p. 1668) "the majority of the 'youngest' songbird species have late Pliocene or early Pleistocene origins." Arbogast and Slowinski state that we (*1*) suggest that "Pleistocene glaciation, in general, did not play an important role in shaping patterns of speciation in this group." In fact we (*1*, p. 1668) stated, "Periodic glacial cycles may have strongly influenced the diversification of the North American songbird fauna. . ." albeit over a more extended period (*22*). We stand by our original conclusion (*1*) that the LPO model of North American songbird evolution is not correct.

John Klicka
Robert M. Zink
J. F. Bell Museum of Natural History,
University of Minnesota,
St. Paul, MN 55108, USA
E-mail: rzink@biosci.cbs.umn.edu

References and Notes

1. J. Klicka and R. M. Zink, *Science* **277**, 1666 (1997).
2. This widely accepted model of North American songbird diversification is derived in large part from the works of Mengel [R. M. Mengel, *Living Bird* **3**, 9 (1964); R. M. Mengel, *Univ. Kansas Dept. Geol. Special Publ.* **3**, 279 (1970)] and Hubbard [J. P. Hubbard, *Living Bird* **12**, 155 (1973)].
3. The distribution of distances for the 35 species pairs does not differ from 13 "randomly" selected species pairs, showing that what ornithologists had assumed was a special class of recently evolved species was in fact not the case. Second, the species thought to be of recent origin did not exhibit the absence of reciprocal monophyly as we found for the Timberline Sparrow (*Spizella taverneri*) and Brewer's Sparrow (*Spizella breweri*). Such a result should be common for species that diverged in the last 250,000 years. Arbogast and Slowinski do not account for these additional pieces of evidence.
4. Independently calibrated (mostly from fossils) clocks from six different avian orders [note 11 in (*1*)] all converge upon the 2.0% per My rate that we used, lending credibility to the idea that some level of molecular rate constancy is discernible across avian taxa.
5. We acknowledged [note 10 in (*1*)] many of the caveats surrounding the use of molecular clocks. A controversy receiving much recent attention is that of among-site nucleotide substitution rate variation [for example, J. Wakeley, *J. Mol. Evol.* **37**, 613 (1993); Z. Yang, N. Goldman, A. Friday, *Mol. Biol. Evol.* **11**, 316 (1994); J. Sullivan and D. L. Swofford, *J. Mamm. Evol.* **4**, 77 (1997)].
6. The appropriate hypothesis calls for mid-point rooting or for the correct outgroups to be specified. With the methodology used by Arbogast and Slowinski and a tree rooted in either of these ways, the gamma-HKY85 model (α = 0.209914, Ts:Tv = 3.7092) of sequence evolution yields a -Ln likelihood of 6773.20406 without the clock and 6787.98634 with the clock enforced. χ² = 29.56, degrees of freedom = 19, *P* > 0.05. The data set used for this test is available from either of us, and the original data set (*1*) contained only one "error," although subsequent sequencing resolved some unknown base positions.
7. Several comments warrant mention. The true songbirds (oscine passerines) are a speciose group of some 4,561 species [C. G. Sibley and B. L. Monroe Jr., *Distribution and Taxonomy of Birds of the World*, (Yale Univ. Press, New Haven, CT, 1990], nearly half of all known birds. Arbogast and Slowinski's test of the molecular clock using only 21 species (<1% overall species diversity) is dubious. Studies of molecular rates [for example, M. Robinson, M. Gouy, C. Gautier, D. Mouchiroud, *Mol. Biol. Evol.* **15**, 1091 (1998)] stress the need for complete taxon sampling. Furthermore, "failing" the test does not mean that the rate differences are large enough to bias our results, or that they are uniform across the birds sampled (that is, not concentrated in one or a few lineages). In addition, the general validity of the χ² likelihood test for a molecular clock has been questioned [N. Goldman, *J. Mol. Evol.* **36**, 182 (1993)].
8. An informative method of assessing rate heterogeneity is the construction of linearized trees [N. Takezaki, A. Rzhetsky, M. Nei, *Mol. Biol. Evol.* **12**, 823 (1995)] in which the exact numbers of problematic taxa and their locations (nodes) are identified.
9. We thank Arbogast and Slowinski for pointing out that most fossil-based calibrations represent older and thus increasingly saturated lineages. If these calibrations suggest a molecular evolutionary rate of 2% sequence divergence per My, then the true rate (that is, unsaturated, as expressed in very young lineages) likely exceeds 2%. We find no fault in this logic and agree that some correction for among site rate variation will, in general, improve temporal resolution. However, we question the magnitude of the correction that they propose (5.04 to 5.06% per My versus 2.0% per My). Improved maximum likelihood models will be useful in this regard.
10. M. Nei, *Molecular Evolutionary Genetics*, (Columbia Univ. Press, New York, 1987); A. Meyer, *Trends Ecol.*

Evol. **9**, 278 (1994); W. S. Moore and DeFilippis, *in Avian Molecular Systematics and Evolution*, D. P. Mindell, Ed. (Academic Press, New York, 1997) pp. 83–119.

11. C. Krajewski and D. G. King, *Mol. Biol. Evol.* **13**, 21 (1996).

12. W. S. Moore, S. M. Smith, T. Prychitko, in *Proc. 22 Int. Ornithol. Congr.*, N. Adams and Slowtow, Eds. (Univ. of Natal, Durban, South Africa), in press.

13. Levels of among-site rate heterogeneity vary widely among different genes, regions of genes, and organisms [D. L. Swofford *et al.*, in *Molecular Systematics*, D. M. Hillis, C. Moritz, and B. K. Mable, Eds. (Sinauer, Sunderland, MA, 1996) pp. 407–514]. Our own work (unpublished) on various songbird data sets suggests that a single gamma estimate often is not applicable even to closely aligned taxonomic groups. Furthermore, gamma estimates for first and second codon positions are typically much lower than that for third position sites [S. V. Edwards, in *Avian Molecular Systematics and Evolution*, D. P. Mindell, Ed. (Academic Press, New York, 1997) pp. 251–278; G. Voelker, S. V. Edwards, *Syst. Biol.*, in press]. Thus the maximum likelihood model used by Arbogast and Slowinski probably provides biased estimates of divergences at low levels. Also, because standard errors for gamma HKY85 distances are not given in the comment, the 95% confidence interval stated for the molecular rate of 5% is unknown.

14. A difference in mtDNA evolution among taxa considered is suggested by the Ts:Tv ratios calculated under the gamma-corrected HKY85 models. The older and more saturated primate data yield a Ts:Tv ratio of 10.3:1 yet a ratio of only 3.7:1 is obtained for the songbirds of more recent origin. In our experience with songbird data, younger lineages always display Ts:Tv ratios that are larger than related but older lineages due to increasing effects of saturation through time.

15. E. Randi, Mol. *Phylogenet. Evol.* **6**, 214 (1996).

16. K. M. Helm Bychowski and A. C. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 688 (1986).

17. D. M. Hillis, B. K. Mable, C. Moritz, in *Molecular Systematics*, D. M. Hillis, C. Moritz, B. K. Mable, Eds. (Sinauer, Sunderland, MA, ed. 2, 996) pp. 515–534.

18. R. C. Fleischer, C. E. McIntosh, C. L. Tarr, *Mol. Ecol.* **7**, 533 (1998).

19. Arbogast and Slowinski do not account for the problem of among-lineage variation in substitution rates. Such variation may reflect differences in body size and correlates such as metabolic rate and generation time [C.-L. Wu and W.-H. Li, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1741 (1985); A. P. Martin and S. R. Palumbi, *ibid.*, **90**, 4087 (1993); P. Canatore *et al.*, *J. Mol. Evol.* **39**, 589 (1994); A. P. Martin, *Mol. Biol. Evol.* **12**, 1124 (1995)] as well as unique population histories (for example, bottleneck frequency) or phylogenetic constraints on mutation and fixation [D. P. Mindell and C. E. Thacker, *Ann. Rev. Ecol. Syst.* **27**, 279 (1996)]. We analyzed only North American oscine passerines (true songbirds) for this specific reason. Arbogast and Slowinski criticize our work for not testing for rate heterogeneity among songbird lineages, yet they assume that songbirds are evolving at the same rate as primates and Galliformes.

20. In earlier work on these birds [C. L. Tarr and R. C. Fleischer, *Auk* **110**, 825 (1993)] a substitution rate of 2 to 2.4% was calculated from the entire mtDNA molecule using RFLP data. Fleischer *et al.* (*18*) attribute this difference to the fact that the cytochrome *b* gene evolves slightly below the genome-wide rate.

21. We note that Arbogast and Slowinski do not apply the same criteria to their own study. For example, they point out that estimated songbird divergence dates, when plotted against a regression of primate divergence estimates, have a lower 95% confidence limit that includes zero. If indeed the error of regression invalidates our study as they suggest, it is unclear why the critical chimpanzee-human fossil date (which is a data point contributing to the regression line in the figure they cite) that also is not statistically distinguishable from zero does not render their own work invalid. At the least, it would render their discussion of the relative distribution of pre- and post-Pleistocene divergence dates (figure 1 in the comment) meaningless.

22. The relevance of the difference in the distributions of divergence dates (figure 1 of the comment) is unclear. Because of variation accompanying genetic distance estimates we did not tabulate and compare exact numbers of Pleistocene versus Pliocene divergences. Arbogast and Slowinski do provide such an accounting. Traditionally, the Pleistocene-Pliocene boundary has been thought to correspond to a time of profound, worldwide climatic change [L. D. Martin, *Inst. for Tertiary-Quaternary Studies Symp. Ser. 1*, 33 (1985)]. The original framers of the LPO model [for example, see (*2*)] stated that the Pleistocene was one My in duration with only four glacial cycles. With new understanding of the paleoecological record, the value assigned to the Pleistocene-Pliocene boundary has increased. We now know that climatic oscillations sufficient to produce major changes in the biota are dated to well within the tertiary [G. M. Hewitt, *Biol. J. Linn. Soc. London* **58**, 247 (1996); T. Webb and P. J. Bartlein, *Annu. Rev. Ecol. Syst.* **23**, 141 (1992)]. The 2-My date that Arbogast and Slowinski cite as a critical temporal reference point may be viewed instead as a relatively arbitrary point in time.

23. We thank R. Sullivan, M. Robinson, W. Moore, A. Simons, A. Kessen, and S. Lanyon for insights and assistance in the preparation of this note.

Science

Pleistocene Speciation and the Mitochondrial DNA Clock

Brian S. Arbogast and Joseph B. Slowinski

ARTICLE TOOLS	http://science.sciencemag.org/content/282/5396/1955.1
REFERENCES	This article cites 25 articles, 5 of which you can access for free http://science.sciencemag.org/content/282/5396/1955.1#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service