proteins may share other aspects of regulation also, and that the presence of interconnected loops found here for *Neurospora*, and previously in *Drosophila* and recently in mouse (*27*), may be a general property of circadian clocks.

These observations require a revision of the way we view the *Neurospora* circadian system (Fig. 4). WC-1 and WC-2 are made in the night and form a white collar complex (WCC) (*10, 23*) through PAS-domain mediated interactions. Late in the night the WCC activates transcription of *frq*, and at dawn the WCC mediates light signaling, which results in a rapid massive induction of *frq* transcript (*7, 9*) that sets the clock; meanwhile, WC-1 levels are declining. Beginning late at night and continuing in the morning, two forms of FRQ are translated from alternative ATG start codons (*8*). Their phosphorylation begins immediately (*8, 28*) and they also rapidly enter into the nucleus (*29*) where they act to block the activity of the WCC (now at close to its lowest level) thereby turning down the expression of their own gene. In the cytoplasm, FRQ translation continues from available but declining *frq* mRNA contributing to a lag between *frq* RNA and FRQ protein peaks. Also acting either directly or indirectly, FRQ promotes the synthesis of WC-1 from existing message so that levels of WC-1 begin to rise even as phosphorylation-promoted turnover of FRQ (*28*) begins. Thus, at close to the same time, FRQ is blocking activation by the WCC while promoting WC-1 synthesis to increase the level of the WCC. Finally phosphorylation of FRQ triggers its precipitous turnover (*8, 28*), FRQ promoted synthesis of WC-1 is balanced by WC-1 degradation, and WC-1 levels peak in the night near to when FRQ levels drop to their low point; the bolus of WC-1 created by the juxtaposition of FRQ promoted WC-1 synthesis and the blockage of WCC activation creates a sharp transition with high WCC activity to initiate the next cycle.

The unexpected identification of a post-transcriptional rhythm for WC-1 synthesis, dual roles for the FRQ proteins in depressing the level of their own transcript and promoting the synthesis of their transcriptional co-activator WC-1, and the resulting identification of an interconnected feedback loop within the *Neurospora* clock, show that there is much left to be learned about the circadian system of even a relatively simple model organism. The strong sequence similarities between WC-1 and BMAL1 suggest that the regulatory themes uncovered may have broad applicability. The interlocking connections among circadian feedback loops, connections that arise when clock molecules assume opposite roles in different phases of the clock cycle, should promote both robustness to the oscillation and stability to the output.

References and Notes

1. J. C. Dunlap, *Cell* **96**, 271 (1999).
2. J. J. Loros, *Curr. Opin. Microbiol.* **1**, 698 (1998).
3. H. Iwasaki and J. C. Dunlap, *Curr. Opin. Microbiol.* **3**, 189 (2000).
4. U. Schibler, *Nature* **404**, 27 (2000).
5. A. L. Scully and S. A. Kay, *Cell* **100**, 297 (2000).
6. B. D. Aronson, K. A. Johnson, J. J. Loros, J. C. Dunlap, *Science* **263**, 1578 (1994).
7. S. K. Crosthwaite, J. C. Dunlap, J. J. Loros, *Science* **276**, 763 (1997).
8. N. Garceau, Y. Liu, J. J. Loros, J. C. Dunlap, *Cell* **89**, 469 (1997).
9. S. C. Crosthwaite, J. J. Loros, J. C. Dunlap, *Cell* **81**, 1003 (1995).
10. H. Linden, P. Ballario, G. Arpaia, G. Macino, *Adv. Genet.* **41**, 35 (1999).
11. R. Allada, N. E. White, W. V. So, J. C. Hall, M. Rosbash, *Cell* **93**, 805 (1998).
12. T. K. Darlington *et al.*, *Science* **280**, 1599 (1998).
13. N. Gekakis *et al.*, *Science* **280**, 1564 (1998).
14. H. Hao, D. L. Allen, P. E. Hardin, *Mol. Cell. Biol.* **17**, 3687 (1997).
15. K. Kume *et al.*, *Cell* **98**, 193 (1999).
16. P. Ballario, C. Talora, D. Galli, H. Linden, G. Macino, *Mol. Microbiol.* **29**, 719 (1998).
17. D. King *et al.*, *Cell* **89**, 641 (1997).
18. J. B. Hogenesch, Y.-Z. Gu, S. Jain, C. A. Bradfield, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5474 (1998).
19. B. L. Taylor and I. B. Zhulin, *Microbiol. Mol. Biol. Rev.* **63**, 479 (1999).
20. S. Honma *et al.*, *Biochem. Biophys. Res. Commun.* **250**, 83 (1998).
21. C. Lee, K. Bae, I. Edery, *Neuron* **21**, 857 (1998).
22. N. J. R. Glossup, L. C. Lyons, P. E. Hardin, *Science* **286**, 766 (1999).
23. C. Talora, L. Franchi, H. Linden, P. Ballario, G. Macino, *EMBO J.* **18**, 4961 (1999).
24. D. Denault, J. C. Dunlap, J. J. Loros, unpublished data.
25. C. Luo, thesis, Dartmouth University, Hanover, NH (1998). Culture conditions for rhythmic expression of FRQ and WC-1 were as described previously (*6–9*).
26. In the course of this work, we identified several frame-shift errors in the original published WC-1 sequence [P. Ballario *et al.*, *EMBO J.* **15**, 1650 (1996)] that resulted in lower BLAST scores. These were also independently identified and corrected by those authors.
27. L. P. Shearman *et al.* [*Science* **288**, 1013 (2000)] reported the existence of interlocking molecular feedback loops in the mammalian circadian clock after submission of this manuscript.
28. Y. Liu, J. Loros, J. C. Dunlap, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 234 (2000).
29. C. Luo, J. J. Loros, J. C. Dunlap, *EMBO J.* **17**, 1228 (1998).
30. Supplemental material is available at www.sciencemag.org/feature/data/1051806.shl
31. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
32. J. J. Loros and J. F. Feldman, *J. Biol. Rhythms* **1**, 187 (1986).
33. M. Merrow, M. Bruner, T. Roenneberg, *Nature* **399**, 584 (1999).

34. We thank the members of our laboratories for assistance on statistics and densitometric analyses, and for critical reading of the manuscript. Supported by grants from NIH (R37-GM 34985 to J.C.D, MH44651 to J.C.D. and J.J.L.), NSF (MCB-9307299 to J.J.L.), and the Norris Cotton Cancer Center core grant at Dartmouth Medical School.

12 April 2000; accepted 2 June 2000

An Empirical Assessment of Taxic Paleobiology

Jonathan M. Adrain[1]* and Stephen R. Westrop[2]

The analysis of major changes in faunal diversity through time is a central theme of analytical paleobiology. The most important sources of data are literature-based compilations of stratigraphic ranges of fossil taxa. The levels of error in these compilations and the possible effects of such error have often been discussed but never directly assessed. We compared our comprehensive database of trilobites to the equivalent portion of J. J. Sepkoski Jr.'s widely used global genus database. More than 70% of entries in the global database are inaccurate; however, as predicted, the error is randomly distributed and does not introduce bias.

The publication of J. J. Sepkoski Jr.'s (*1*) factor-analytical description of the marine fossil record was an epochal event in modern evolutionary paleobiology. The compilation of marine families on which it was based (*2*, *3*) has served as the raw material for many influential papers, including studies of extinction (*4–7*), the periodicity of mass extinction (*8–10*), and evolutionary rates (*11–15*). Recognizing the need for a more detailed level of analysis, Sepkoski (*16*) began a more ambitious compilation of fossil genera (*17*), which now serves as the foundation for the majority of current work in the field.

There are critics of taxic paleobiology (*18*). Some have pointed out that taxa of a particular Linnean rank have no natural equivalence (*19*), and others (*20*, *21*) that traditional taxonomy contains a large number of polyphyletic or paraphyletic groups, which hamper the estimation of large-scale pattern (*22*). The most widespread complaint (*23*, *24*), however, has been that the basic accuracy of global databases is suspect because they are compiled by workers who are not systematic specialists. Because

[1]Department of Geoscience, University of Iowa, Iowa City, IA 52242, USA. [2]Oklahoma Museum of Natural History and School of Geology and Geophysics, University of Oklahoma, Norman, OK 73019, USA.

*To whom correspondence should be addressed. E-mail: jonathan-adrain@uiowa.edu

nonspecialists have no means to evaluate or correct taxonomic data, they might compile considerable error, and this error might significantly bias the result. This charge has been answered on conceptual grounds by proponents of taxic paleobiology, who claim that statistical treatment of the data in itself presumes that error is present and accounts for it (*18*), and that even large amounts of error, if randomly distributed, are unlikely to bias the underlying pattern (*25*). Others have pointed to the fact that major revisions of large databases have had little impact on the results produced by the use of those databases (*26*). A seminal study by Sepkoski himself (*27*) demonstrated that significant additions and corrections to his family compilation had minimal effect on diversity patterns. To date, however, there has never been an empirical assessment of either the amount of error present in large databases or of its effect.

We tested the accuracy of the Sepkoski genus database by comparing a portion of it to our own independent database of trilobite genera (*28*). Our database has been assembled over a period of years by systematic specialists and represents a taxonomically standardized (*29*), critically evaluated compilation that we regard as essentially complete and exhaustive for the Ordovician and Silurian periods. Although it might still contain errors and omissions, it is considered "correct" data against which to compare the equivalent portion of the Sepkoski database. Taking the study interval from the beginning of the Ordovician through the end of the Lower Silurian (Wenlock), the Sepkoski compilation yields a sample size of 941 trilobite genera. Trilobites are among the most common fossils of the Lower Paleozoic (*28*) and are likely to be a good proxy for other components of the global genus database.

Our sampling intervals for the Lower Silurian, the five global standard stages, are identical to Sepkoski's (*2*, *3*). We used 9 intervals for the Ordovician, in contrast to Sepkoski's 13. Problems of global correlation, regional biostratigraphic resolution, and reported resolution make his more detailed scheme impossible to apply to records from many parts of the world, and our scheme is based on widely correlated biohorizons (*30*). This creates no difficulty, because Sepkoski's scheme is easily and directly transferable to our own.

Of Sepkoski's trilobite data, the stratigraphic ranges of about one-third of the records are not resolved to his sampling scheme (*31*). Sepkoski (*16*, *32*) had explicit strategies to distribute the known error introduced by this. We followed Sepkoski in distributing the resolution error by assigning the bin-occurrence of poorly resolved taxa in proportion to that of the fully resolved taxa, and we calculated range error with known errors both included and excluded (that is, using only the fully resolved data).

We compared the databases by checking each genus in the Sepkoski compilation and assessing its validity and stratigraphic range relative to our estimate. We also tracked its stratigraphic interval occurrence, both as presence and absence, and scaled to include the distribution of uncertainty in stratigraphic resolution. If a genus was for some reason not accepted (for example, if it was a junior synonym of a previously established taxon), all of its assigned occurrence was treated as error. If a genus was accepted, then the accuracy of its stratigraphic first appearance (FA) and last appearance (LA) was recorded. If either was in error, the direction and magnitude of that error was tabulated.

Our database contains 1383 historically proposed trilobite genera with occurrence in the study interval, of which 389 are junior synonyms, leaving 994 valid genera. The Sepkoski database lists 941 genera for the same interval, of which 681 contain error of some kind (Fig. 1): 202 are invalid records for various reasons and 479 are valid genera with errors in FA, LA, or both. Only 260 of 941 records are valid genera with accurate ranges (*33*). Figure 2 shows the distribution of error in reported FA and LA among fully resolved records.

It is evident that most of the data in the Sepkoski trilobite data set are erroneous. The database captures only 74% of the valid genera for the interval and includes an additional 20.3% of taxonomic "noise," while 55.2% of raw reported FA (46.5% of fully resolved FA) and 53.6% of reported LA (45.0% of fully resolved LA) are erroneous. The magnitude of raw individual sampling bin error (the reported occurrence of a genus in a sampling interval that is incorrect, or failure to report a valid genus in an interval in which it occurs) is 52.9% that of the total taxon-bin occurrence. And finally, only 27.6% of the records in the database comprise valid genera with correct stratigraphic ranges.

Does this error result in bias? Figure 3 plots the cumulative genus diversity by sampling interval as indicated by both the raw Sepkoski data and our own data, as well as the respective percentage changes in diversity from interval to interval and the distribution of taxonomic "noise" by interval in the Sepkoski data set. The raw Sepkoski data describe a curve nearly identical to, and almost coincident with, our own. A G test of goodness of fit between the two curves demonstrates that they are not significantly different ($P > 0.70$). The Sepkoski data track both the direction and magnitude of diversity changes almost exactly. Cumulative diversity curves differ most at the upper and, especially, lower ends of the Ordovician-Wenlock study interval, and it is clear that this is due to higher levels of taxonomic noise (Fig. 3B). Noise levels are elevated at each end of the distribution by the inclusion of genera whose stratigraphic durations actually fall in entirely older or younger strata. Such errors are less likely to

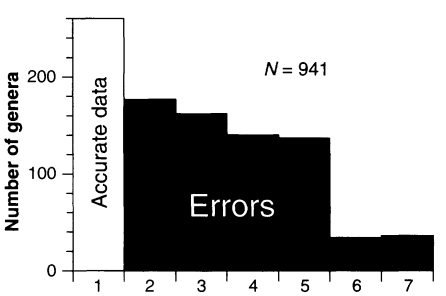

Fig. 1. Accuracy of genus records in the Sepkoski trilobite database (N = 941 total genera). 1, valid genera with accurate stratigraphic ranges; 2, valid genera with incorrect stratigraphic FA but correct LA; 3, valid genera with incorrect LA but correct FA; 4, valid genera with incorrect FA and LA; 5, junior synonyms; 6, valid genera with LA before or FA after the study interval; 7, other errors: uninterpretable genera restricted to types, misspelled repetitions, and nomina nuda.

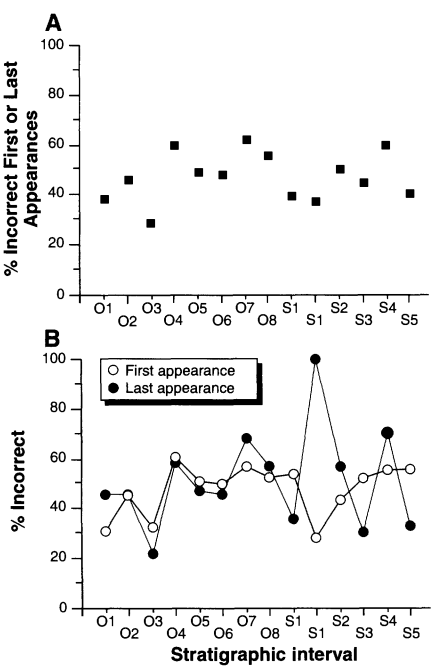

Fig. 2. (**A**) Proportion of all range data among fully resolved records that is incorrect, by interval. (**B**) Proportion of total reported FA and LA among fully resolved records that is incorrect, by interval.

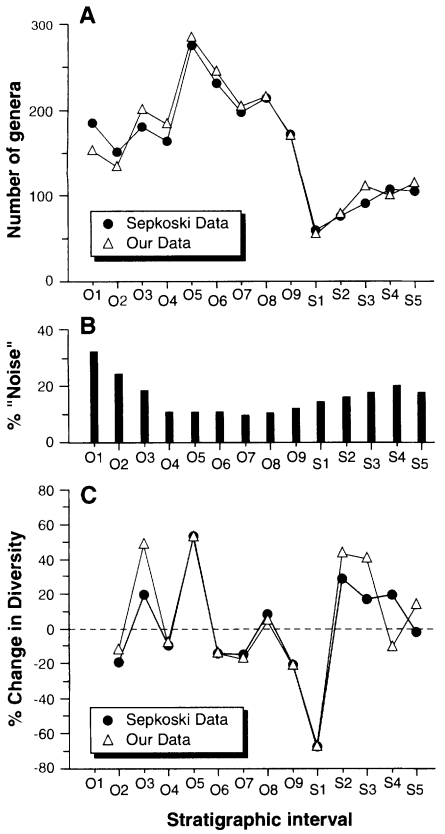

Fig. 3. (**A**) Cumulative genus diversity by interval, as indicated by the Sepkoski data and by our data. (**B**) Distribution of taxonomic "noise" in the Sepkoski data set. (**C**) Direction and magnitude of the percent change in diversity from interval to interval as indicated by either data set.

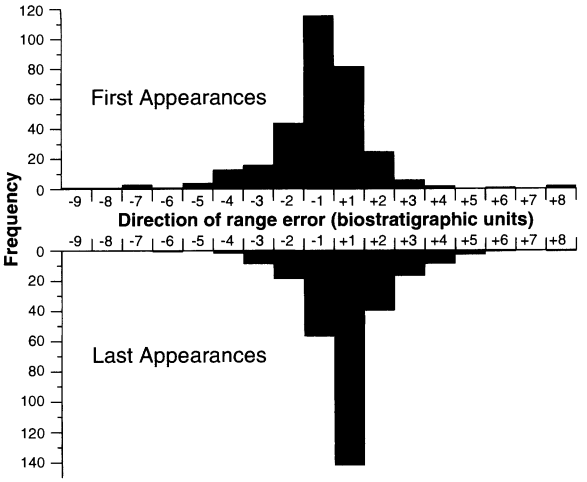

Fig. 4. Frequency distribution of magnitude and direction of range error among valid Sepkoski genera.

be a factor toward the midpoint of the distribution, where noise levels are reduced. The bias toward more noise at the lower end of the distribution (that is, the earliest Ordovician) reflects historical uncertainty in the position and correlation of the Cambrian-Ordovician boundary.

Stratigraphic range error appears to be a less important factor. With the exception of a few sample bins, the frequency of resolved range error is essentially constant (Fig. 2A) and plots in a narrow band around a mean value of 47%. Errors in raw first or last appearances are symmetrically distributed (Fig. 4). Stratigraphic ranges are commonly overestimated, but in similar amounts in either direction, so that no pervasive bias is introduced. These results indicate that studies in analytical paleobiology that are based on large compilations of data are likely to be highly resilient to error.

We conclude that Sepkoski's global genus database, although rife with error and of little value for low-level systematic studies, is adequate for its intended application. As far as can be determined, it accurately estimates the large-scale patterns of Phanerozoic biodiversity, and its widespread use in current studies of analytical paleobiology is justified.

References and Notes

1. J. J. Sepkoski Jr., *Paleobiology* **7**, 36 (1981).
2. _____, *Milw. Public Mus. Contrib. Biol. Geol.* **51**, 1 (1982).
3. _____, *Milw. Public Mus. Contrib. Biol. Geol.* **83**, 1 (1992).
4. D. M. Raup and J. J. Sepkoski Jr., *Science* **215**, 1501 (1982).
5. D. Jablonski, *Science* **231**, 129 (1986).
6. M. L. McKinney, *Paleobiology* **11**, 227 (1985).
7. _____, *Nature* **325**, 143 (1987).
8. D. M. Raup and J. J. Sepkoski Jr., *Proc. Natl. Acad. Sci. U.S.A.* **81**, 801 (1984).
9. _____, *Science* **231**, 833 (1986).
10. J. J. Sepkoski Jr., *J. Geol. Soc.* **146**, 7 (1989).
11. K. W. Flessa and D. Jablonski, *Nature* **313**, 216 (1985).
12. G. E. Boyajian, *Geology* **14**, 955 (1986).
13. N. L. Gilinsky and R. K. Bambach, *Paleobiology* **13**, 427 (1987).
14. M. Foote, *Paleobiology* **14**, 258 (1988).
15. N. L. Gilinsky, *Paleobiology* **20**, 445 (1994).
16. J. J. Sepkoski Jr., in *Global Bio-Events*, O. H. Walliser, Ed. (Springer-Verlag, Berlin, 1986), pp. 47–61.
17. The compilation remained unpublished at the time of Sepkoski's death but had been freely circulated since its inception.
18. Reviewed by M. J. Benton, in *Numerical Palaeobiology*, D. A. T. Harper, Ed (Wiley, Chichester, UK, 1999), pp. 249–283.
19. B. D. Mishler, in *Species: New Interdisciplinary Essays*, R. A. Wilson, Ed. (MIT Press, Cambridge, MA, 1999), pp. 307–315.
20. C. Patterson and A. B. Smith, *Nature* **330**, 248 (1987).
21. A. B. Smith and C. Patterson, *Evol. Biol.* **23**, 127 (1988).
22. See, however, J. J. Sepkoski Jr. and D. C. Kendrick, *Paleobiology* **19**, 168 (1993).
23. D. V. Ager, in *Extinction and Survival in the Fossil Record*, Systematics Association Special Volume No. 34, G. P. Larwood, Ed. (Clarendon Press, Oxford, 1988), pp. 89–97.
24. C. W. Stearn, *Palaeontol. Electron.* **2**, issue 1, 2 pp., 15 KB (1999).
25. D. M. Raup, in *Analytical Paleobiology*, Short Courses in Paleontology Number 4, N. L. Gilinsky and P. W. Signor, Eds. (Paleontological Society, Knoxville, TN, 1991), pp. 207–216.
26. W. D. Maxwell and M. J. Benton, *Paleobiology* **16**, 322 (1990).
27. J. J. Sepkoski Jr., *Paleobiology* **19**, 43 (1993).
28. J. M. Adrain, R. A. Fortey, S. R. Westrop, *Science* **280**, 1922 (1998). The version of the database used here has been subject to ongoing refinement (for example, from 6 sampling intervals to 14) and expansion by J.M.A. The total number of genera has increased by 11.4%, but the number of accepted genera by only 5.2%, indicating that most recent additions have been junior synonyms culled from the more obscure primary literature. A current version of the database is available on request from J.M.A.
29. S. J. Culver, M. A. Buzas, L. S. Collins, *Paleobiology* **13**, 169 (1987).
30. Our sampling intervals for the Ordovician are based on recent work on potential global subdivisions and key faunal markers, as summarized by B. D. Webby [*Newsl. Stratigr.* **36**, 1 (1999)]. Terms given are based on Webby's figure 4, left column; the usage of these terms is in some cases considerably emended from their traditional meaning. The corresponding intervals from the Sepkoski compilation are given in parentheses. O1, lower Tremadocian (Trem-l); O2, upper Tremadocian (Trem-u); O3, "Latorpian" (Aren-l); O4, "Volkhovian" (Aren-u); O5, Darriwilian (Llvi-l, Llvi-u, and Llde-l); O6, lower "Caradocian" (Llde-u and Cara-l); O7, upper "Caradocian" (Cara-m); O8, lower "Ashgillian" (Cara-u and Ashg-l); and O9, upper "Ashgillian" (Ashg-m and Ashg-u). Silurian sampling intervals are the standard Llandovery and Wenlock global stages: S1, Rhuddanian (Ldov-l); S2, Aeronian (Ldov-m); S3, Telychian (Ldov-u); S4, Sheinwoodian (Wenl-l); and S5, Homerian (Wenl-u).
31. This matches an earlier estimate for the entire database given by Sepkoski (*16*). The lack of precision results mainly from dependence on secondary data sources such as the *Treatise on Invertebrate Paleontology* and *Zoological Record*, where stratigraphic data is often given only to series or subsystem. Precise ranges can always be determined through reference to the primary literature (a time-consuming task beyond the scope of Sepkoski's compilation). All of our data are resolved to our sampling scheme, and no distribution of error was necessary.
32. J. J. Sepkoski Jr. and C. F. Koch, in *Global Events and Event Stratigraphy*, O. H. Walliser, Ed (Springer-Verlag, Berlin, 1996), pp. 21–34.
33. This number overestimates the accuracy of the Sepkoski database because some range error will have been masked by "dashing down" from his more resolved to our less resolved Ordovician stratigraphic scheme.

34. We thank the late J. J. Sepkoski Jr. for providing us with his trilobite genus data; R. A. Fortey for advice during the compilation of an earlier (1998) version of the trilobite database; and M. J. Benton, G. D. Edgecombe, R. A. Fortey, and R. Lupia for comments on the manuscript. Supported by NSF grant EAR 9973065.

30 March 2000; accepted 19 May 2000