Chl-$Z\rho_{26}$ and Chl-ZML relationships. This important role of ocean physics indicates that the observed Chl changes primarily reflect biomass changes due to dampened or increased nutrient fluxes to the upper lit layers. Changes in the photoacclimation state of phytoplankton probably also intervene, without obscuring the global picture, however. Quantifying the respective role of both phenomena would require the parallel examination of Chl and other quantities more directly tied to biomass.

A basin-specific response of phytoplankton to large-scale climate oscillators has been shown here. This result argues for a more accurate representation of decadal regimes into global ocean models, whose predictions of the response of ecosystems to global change are still uncertain (*20*–*22*). Such improvements are crucial for a better forecast of the impact of climate change on ecosystems and carbon fluxes. Our results also show that dampening the effect of interannual variability by averaging over two decades allows the decadal variability to be revealed and analyzed. Therefore, it can be anticipated that averaging over several decades may eventually reveal longer-term trends related to subtle changes in physical forcing. This emphasizes the critical importance of reanalyzing historical data sets (*23*, *24*) and of continuing the construction of climate-quality satellite data records in the next decades (*25*).

References and Notes

1. D. Antoine, A. Morel, H. R. Gordon, V. F. Banzon, R. H. Evans, *J. Geophys. Res.* **110**, C06009 (2005).
2. W. W. Gregg, M. E. Conkright, *Geophys. Res. Lett.* **29**, 1730 (2002).
3. C. R. McClain, S. R. Signorini, J. R. Christian, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **51**, 281 (2004).
4. M. J. Behrenfeld *et al*., *Nature* **444**, 752 (2006).
5. J. A. Yoder, M. A. Kennelly, *Oceanography (Wash. D.C.)* **19**, 152 (2006).
6. C. R. McClain, G. C. Feldman, S. B. Hooker, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **51**, 5 (2004).
7. W. A. Hovis *et al*., *Science* **210**, 60 (1980).
8. Materials and methods are available as supporting material on *Science* Online.
9. T. M. Smith, R. W. Reynolds, T. C. Peterson, J. Lawrimore, *J. Clim.* **21**, 2283 (2008).
10. N. J. Mantua, S. R. Hare, Y. Zhang, J. M. Wallace, R. C. Francis, *Bull. Am. Meteorol. Soc.* **78**, 1069 (1997).
11. D. Enfield, A. Mestas-Nunez, P. Trimble, *Geophys. Res. Lett.* **28**, 2077 (2001).
12. N. J. Mantua, S. R. Hare, *J. Oceanogr.* **58**, 35 (2002).
13. J. Hua, W. Dexing, W. Xiuquan, *Chin. J. Oceanol. Limnol.* **24**, 111 (2006).
14. J. A. Carton, B. S. Giese, *Mon. Weather Rev.* **136**, 2999 (2008).
15. S. C. Doney, *Nature* **444**, 695 (2006).
16. J. J. Polovina, E. A. Howell, M. Abecassis, *Geophys. Res. Lett.* **35**, L03618 (2008).
17. D. M. Smith *et al*., *Science* **317**, 796 (2007).
18. N. Schneider, B. D. Cornuelle, *J. Clim.* **18**, 4355 (2005).
19. J. R. Knight, R. J. Allan, C. K. Folland, M. Vellinga, M. E. Mann, *Geophys. Res. Lett.* **32**, L20708 (2005).
20. L. Bopp *et al*., *Global Biogeochem. Cycles* **15**, 81 (2001).
21. J. L. Sarmiento *et al*., *Global Biogeochem. Cycles* **18**, GB3003 (2004).
22. B. Schneider, *Biogeosciences* **5**, 597 (2008).
23. R. H. Evans, H. R. Gordon, *J. Geophys. Res.* **99**, 7293 (1994).
24. P. G. Falkowski, C. Wilson, *Nature* **358**, 741 (1992).
25. C. R. McClain, S. B. Hooker, G. C. Feldman, P. Bontempi, *Eos* **87**, 337 (2006).

26. We thank the Agence Nationale de la Recherche (Paris) for the financial support for this work carried out within the frame of the Globphy project, NASA (U.S.) for providing the SeaWiFS global chlorophyll data, and the U.S. National Oceanographic and Atmospheric Administration National Climatic Data Center for the ERSST v3 data set. We thank D. Siegel, H. Claustre, Y. Huot, and A. Morel for the comments they provided on early versions of the manuscript, and three anonymous reviewers for their comments and helpful suggestions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5957/1253/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References

29 May 2009; accepted 24 September 2009
10.1126/science.1177012

Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly

Michael E. Mann,[1]* Zhihua Zhang,[1] Scott Rutherford,[2] Raymond S. Bradley,[3] Malcolm K. Hughes,[4] Drew Shindell,[5] Caspar Ammann,[6] Greg Faluvegi,[5] Fenbiao Ni[4]

Global temperatures are known to have varied over the past 1500 years, but the spatial patterns have remained poorly defined. We used a global climate proxy network to reconstruct surface temperature patterns over this interval. The Medieval period is found to display warmth that matches or exceeds that of the past decade in some regions, but which falls well below recent levels globally. This period is marked by a tendency for La Niña–like conditions in the tropical Pacific. The coldest temperatures of the Little Ice Age are observed over the interval 1400 to 1700 C.E., with greatest cooling over the extratropical Northern Hemisphere continents. The patterns of temperature change imply dynamical responses of climate to natural radiative forcing changes involving El Niño and the North Atlantic Oscillation–Arctic Oscillation.

Considerable progress has been made over the past decade in using climate "proxy" data to reconstruct large-scale trends in past centuries, and in using climate models to assess the roles of natural and anthropogenic forcing in those trends (*1*). Owing in part to the sparseness of the available proxy data, less progress has been made in identifying the underlying spatial patterns of those changes, let alone the causal factors behind them. Yet a better understanding of past patterns of climate change and their causes (e.g., the role of past changes in the El Niño–Southern Oscillation, or ENSO) may be even more important for validating the regional-scale projections, which are paramount in assessing future climate change impacts.

Patterns of past climate change can be estimated through the simultaneous analysis of multiple spatially distributed proxy records. Such analyses have been performed via statistical reconstruction (*2*–*8*) and model assimilation approaches (*9*), but available proxy networks have been insufficient for estimating spatially resolved large-scale temperature reconstructions beyond the past few centuries (*2*, *4*, *7*).

Here, we employ a diverse multiproxy network previously used to estimate global and hemispheric mean annual temperature trends (*10*) to reconstruct global patterns of surface temperature changes over the past 1500 years. We use a climate field reconstruction (CFR) approach (*11*) that has been rigorously tested with synthetic "pseudoproxy" networks generated from forced climate model simulations (*12*). We interpret the resulting reconstructions in the context of results from climate model simulations forced by estimated past changes in natural (solar and volcanic) radiative forcing.

We employ the global proxy data set used by (*13*) comprising more than a thousand tree-ring, ice core, coral, sediment, and other assorted proxy records spanning the ocean and land regions of both hemispheres over the past 1500 years. The surface temperature field is reconstructed by calibrating the proxy network against the spatial information contained within the instrumental annual mean surface temperature field (*14*) over a modern period of overlap between proxy and instrumental data (1850 to 1995) using the RegEM CFR procedure (*12*) with additional minor modifications. Further details of

[1]Department of Meteorology and Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA 16802, USA. [2]Department of Environmental Science, Roger Williams University, Bristol, RI 02809, USA. [3]Department of Geosciences, University of Massachusetts, Amherst, MA 01003–9298, USA. [4]Laboratory of Tree-Ring Research, University of Arizona, Tucson, AZ 85721, USA. [5]NASA Goddard Institute for Space Studies, New York, NY 10025, USA. [6]Climate Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO 80305, USA.

*To whom correspondence should be addressed. E-mail: mann@meteo.psu.edu

the reconstruction procedure, associated statistical validation and skill assessments, uncertainty estimation procedures, data used, and MATLAB source codes for the analysis procedures are provided in the Materials and Methods.

Earlier proxy-based large-scale surface temperature reconstructions (*2*, *15*) resolved only a single statistical degree of freedom before the 15th century, precluding the possibility of investigating spatial patterns of surface temperature variation in earlier centuries. By contrast, the current reconstructions resolve multiple degrees of freedom in the surface temperature field back through the 6th century, allowing us to meaningfully interpret spatial features in the

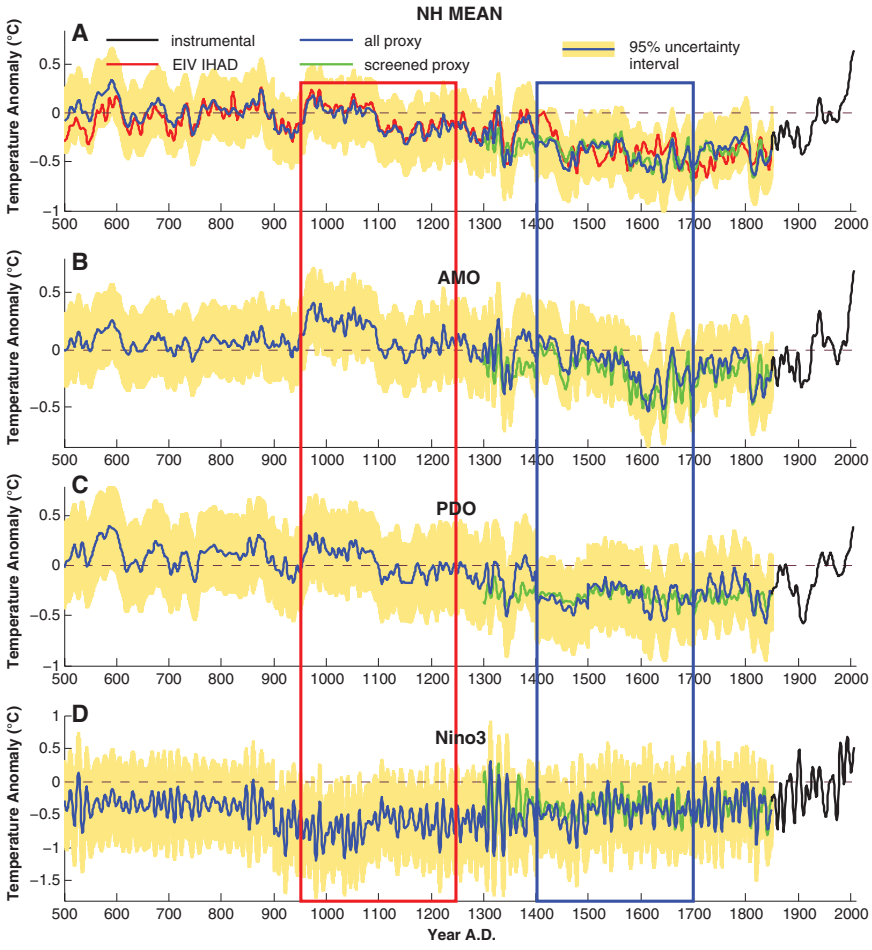

Fig. 1. Decadal surface temperature reconstructions. Surface temperature reconstructions have been averaged over (**A**) the entire Northern Hemisphere (NH), (**B**) North Atlantic AMO region [sea surface temperature (SST) averaged over the North Atlantic ocean as defined by (*30*)], (**C**) North Pacific PDO (Pacific Decadal Oscillation) region (SST averaged over the central North Pacific region 22.5°N–57.5°N, 152.5°E–132.5°W as defined by (*31*)], and (**D**) Niño3 region (2.5°S–2.5°N, 92.5°W–147.5°W). Shading indicates 95% confidence intervals, based on uncertainty estimates discussed in the text. The intervals best defining the MCA and LIA based on the NH hemispheric mean series are shown by red and blue boxes, respectively. For comparison, results are also shown for parallel ("screened") reconstructions that are based on a subset of the proxy data that pass screening for a local temperature signal [see (*13*) for details]. The Northern Hemisphere mean Errors in Variables (EIV) reconstruction (*13*) is also shown for comparison.

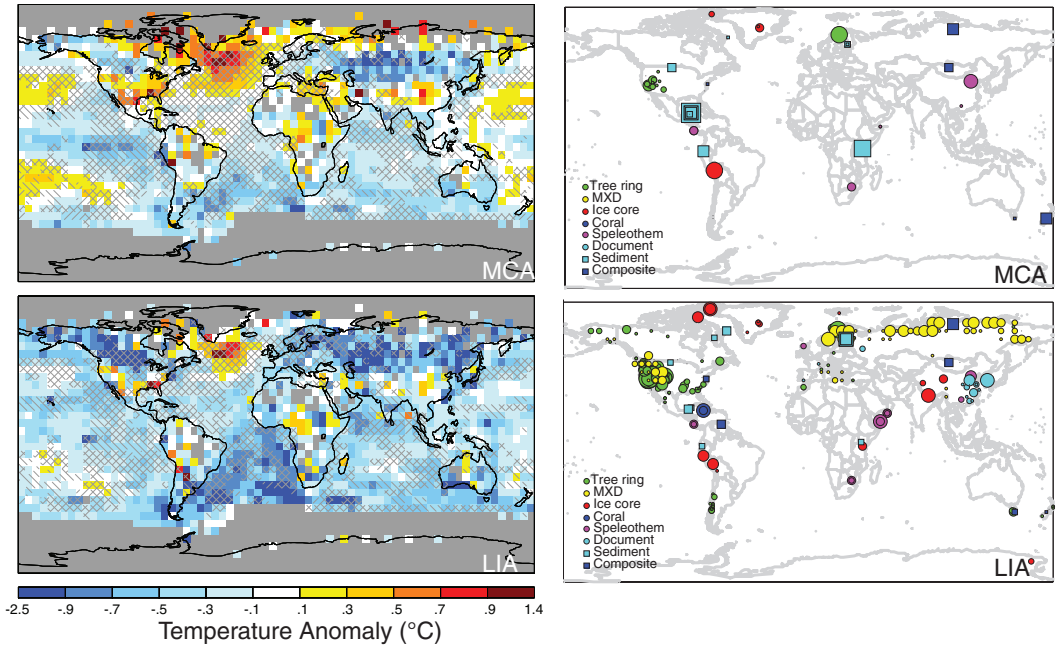

Fig. 2. Reconstructed surface temperature pattern for MCA (950 to 1250 C.E.) and LIA (1400 to 1700 C.E.). Shown are the mean surface temperature anomaly (left) and associated relative weightings of various proxy records used (indicated by size of symbols) for the low-frequency component of the reconstruction (right). Anomalies are defined relative to the 1961–1990 reference period mean. Statistical skill is indicated by hatching [regions that pass validation tests at the $P = 0.05$ level with respect to *RE* (*CE*) are denoted by / (\) hatching]. Gray mask indicates regions for which inadequate long-term modern observational surface temperature data are available for the purposes of calibration and validation.

reconstructions. Nonetheless, certain caveats must be kept in mind in interpreting the proxy-based surface temperature reconstructions. Before 1600 C.E., the low-frequency component of the surface temperature reconstructions is described as a linear combination of just two leading patterns of temporal variation, so that regional features in the temperature field are represented by a spatiotemporally filtered approximation. Moreover, as decadal-resolution proxy data were used in addition to annual-resolution data, only interdecadal and longer-term variations are meaningfully resolved; i.e., the details of individual years and even individual decades should not be emphasized. Thus, it is the longer-term, and larger-scale, variations resolved by the reconstructions that are most meaningful.

The large-scale surface temperature reconstructions, when spatially averaged, e.g., over the Northern Hemisphere, yield a long-term history very similar to the hemispheric mean reconstructions of (*13*) (Fig. 1A). However, the spatial reconstructions can also be averaged to yield other indices of interest [Fig. 1, B to E; other regional average series are shown in the Supporting Online Material (SOM) Text]. Though there are relatively few distinct patterns of variation resolved by the reconstructions, particularly before 1600 C.E., there are notable differences of behavior among the various diagnosed indices. The Atlantic Multidecadal Oscillation (AMO) series, for example, is marked by substantial multidecadal variability, consistent with previous proxy studies of North Atlantic variability [e.g., (*16*)]. The high-frequency fluctuations of the Niño3 series are consistent with the oscillatory nature of ENSO. The Niño3 index suggests strong and persistent La Niña conditions around 1000 years ago, as discussed further below.

Our reconstructions span two climatologically interesting periods, the so-called Little Ice Age (LIA) and Medieval Climate Anomaly (MCA). For the purpose of investigating the associated spatial patterns (Fig. 2), we defined the LIA and MCA in terms of distinct three-century-long intervals (1400 to 1700 C.E. and 950 to 1250 C.E., respectively), which both correspond to relative cold and warm hemispheric conditions, respectively (Fig. 1), and are distinct with regard to the estimated external radiative forcing of the climate (*1*, *17*). The observed patterns are not, however, sensitive to the precise time intervals used to define these periods (fig. S9). The MCA pattern is based on a smaller number of predictors than the LIA pattern (Fig. 2) and, accordingly, on fewer resolved spatial degrees of freedom (SOM Text).

The reconstruction skill diagnostics suggest that the MCA and LIA reconstructions are most reliable (Fig. 2) over the Northern Hemisphere and tropics, and least reliable in the Southern Hemisphere, particularly in the extratropics. To assess if the larger-scale features of the earlier MCA pattern are robust, we used only the more restricted network of proxy data available back through the beginning of the MCA interval to reconstruct temperatures for the LIA interval. This analysis gave a reconstruction very similar to the LIA reconstruction based on the full data set (fig. S10).

The reconstructed MCA pattern is characterized by warmth over a large part of the North Atlantic, Southern Greenland, the Eurasian Arctic, and parts of North America, which appears to

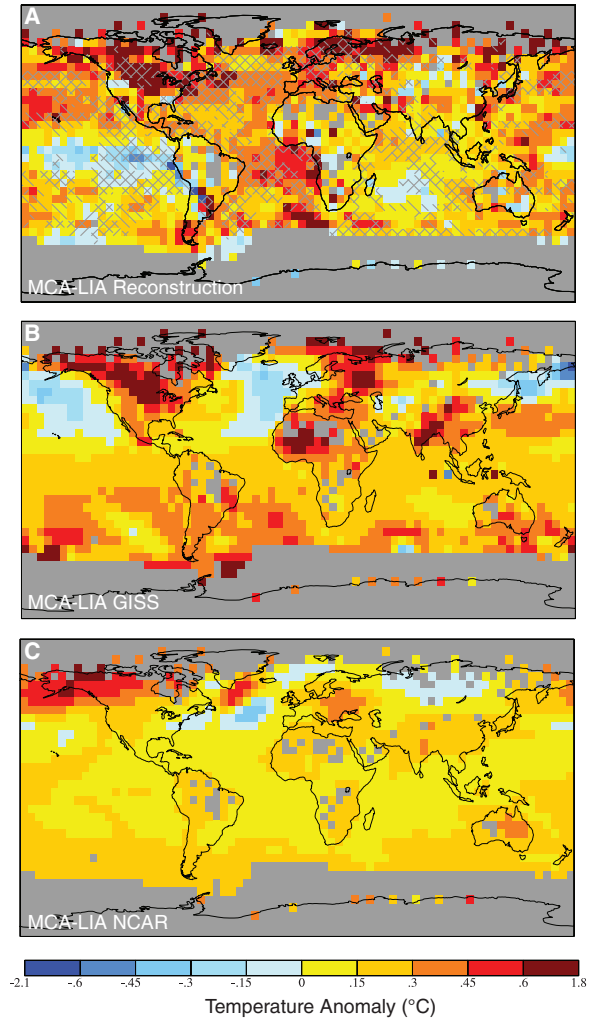

Fig. 3. Spatial pattern of MCA-LIA surface temperature difference in reconstructions and model simulations. (**A**) Proxy-based temperature reconstructions, (**B**) GISS-ER (using the same solar forcing difference used in the NCAR simulation—shown is the ensemble mean; see the SOM for example results from one of six realizations), and (**C**) NCAR CSM 1.4 simulation (using the same MCA and LIA time intervals as defined above). The observational mask has been applied to both model patterns for ease of comparison. Statistical skill for (A) is indicated with the same conventions as in Fig. 2 (statistical significance here indicates that the particular test statistic independently passed during both the MCA and LIA intervals).

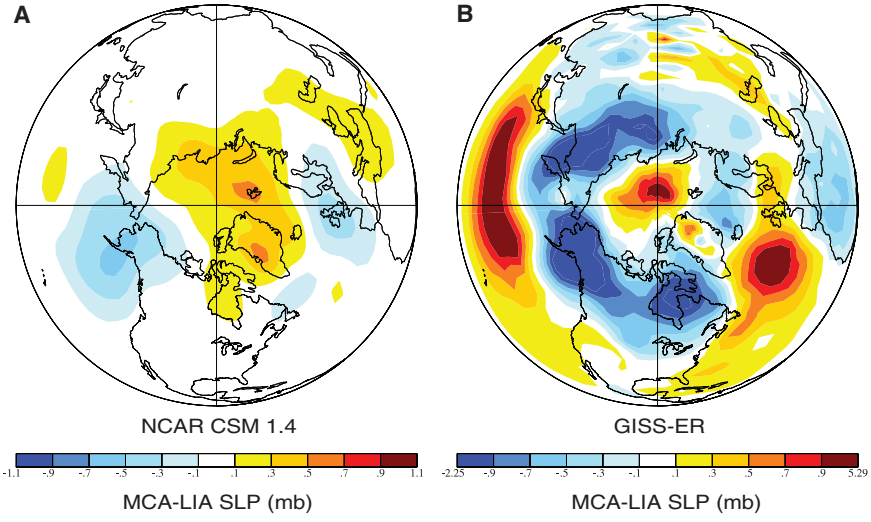

Fig. 4. Spatial pattern of MCA-LIA sea-level pressure difference in model simulations. (**A**) NCAR CSM 1.4 and (**B**) GISS-ER. For the NCAR model, a single run was available (*23*). For the GISS-ER coupled model, we show the ensemble mean of six realizations; see SOM section 5 for further details.

substantially exceed that of the modern late–20th century (1961–1990) baseline and is comparable to or exceeds that of the past one-to-two decades in some regions. This finding is consistent with that of a recent tree-ring–based study of high-latitude Eurasian temperatures (*18*). Relative warmth in the central North Pacific MCA is consistent with the expected extratropical signature of the strong observed La Niña–like pattern in the tropical Pacific (strong cooling in the east and warming in the west). Certain regions, such as central Eurasia, northwestern North America, and (with less confidence) parts of the South Atlantic, exhibit anomalous coolness. The LIA pattern is characterized primarily by pronounced cooling over the Northern Hemisphere continents, but with some regions—e.g., parts of the Middle East, central North Atlantic, Africa, and isolated parts of the United States, tropical Eurasia, and the extratropical Pacific Ocean—displaying warmth comparable to that of the present day. In some places, e.g., northern Labrador, apparent LIA warmth is a product, at least in part, of the relatively cool nature of the 1961–1990 reference period in the region.

For comparison with model simulation results, it is useful to eliminate the influence of the choice of modern reference period by examining the pattern of the MCA-LIA difference itself (Fig. 3). The MCA-LIA pattern highlights the extent to which the MCA is both more "La Niña–like" [e.g., (*17*, *19*–*21*)] and, with enhanced warmth over interior North America and the Eurasian Arctic, and cooling over central Eurasia, suggestive of the positive phase of the North Atlantic Oscillation (NAO) and closely related Arctic Oscillation (AO) sea-level pressure (SLP) pattern (*17*, *22*), as discussed further below.

We examined results for two different coupled model simulations of the past millennium, driven with those factors (solar irradiance changes and stratospheric aerosols from explosive volcanic eruptions) that can most plausibly explain the climate changes of the past millennium (*17*): (i) the National Center for Atmospheric Research (NCAR) Climate System Model (CSM) 1.4 coupled model driven with estimated solar plus volcanic forcing over the past millennium [see (*23*) for details]; and (ii) the Goddard Institute for Space Studies–ER (GISS-ER) coupled model with solar (but no volcanic) forcing (SOM Text), scaled for an MCA-LIA solar radiative forcing at the tropopause of 0.37 W/m^2 (equivalent to the MCA-LIA solar forcing difference used in the NCAR simulation). Both simulations give very similar estimates of the global mean MCA-LIA temperature difference (0.16° and 0.24°C for NCAR and GISS, respectively; the latter is identical to the proxy reconstructed mean surface temperature difference of 0.24°C). The spatial patterns of response for the two models (Fig. 3), however, are quite different, as discussed further below.

The La Niña–like nature of the MCA-LIA pattern is not reproduced in either of the two different coupled model simulations analyzed. On the other hand, such a pattern is reproduced in simulations (*19*) using the low-order Cane-Zebiak (*24*) model of the tropical Pacific coupled ocean-atmosphere system. The discrepancy in the model responses may arise because the tropical Pacific "thermostat" mechanism (*25*) is not active in either the NCAR or GISS simulations. In (*19*), this mechanism is responsible for the La Niña–like response to the positive tropical radiative forcing of the MCA that arises from a combination of relatively high solar irradiance and inactive tropical volcanism. Although there is still a vigorous debate regarding the nature of the response of the tropical Pacific to anthropogenic radiative forcing [e.g., (*26*)], paleoclimate evidence examined here, as elsewhere [e.g., (*19*, *27*)], appears to support a thermostat-like response, at least for natural radiatively forced climate changes in past centuries.

The NCAR simulation also does not reproduce the enhanced warming over the Eurasian Arctic, high-latitude North Atlantic, and North American region evident in the reconstructed MCA-LIA pattern. As discussed previously, this surface temperature pattern is consistent with a relative positive (negative) NAO-AO atmospheric circulation anomaly during the MCA (LIA), associated with annular bands of positive (negative) SLP anomalies in the subtropics and midlatitudes, and negative (positive) SLP anomalies in the subpolar latitudes. Such a pattern has been inferred in paleoclimate studies of the past millennium (*5*, *17*, *22*, *28*, *29*), and the negative phase of this pattern has been produced as a dynamical response to decreased solar radiative forcing during the LIA using a previous version of the NASA GISS model that incorporates the effects of ozone photochemistry on the vertical structure of the atmosphere (*28*, *29*). These effects are not accounted for in the NCAR simulation, which is limited to 36 km in vertical extent. The GISS-ER model used here extends to ~80 km and does incorporate these processes and, indeed, reproduces roughly the observed pattern of enhanced North American, high-latitude North Atlantic, and Arctic Eurasian warming, as a dynamical response to the imposed radiative forcing. These surface temperature changes are, in turn, associated with an annular atmospheric circulation response (Fig. 4) reminiscent of the positive phase of the NAO-AO pattern, though with some differences [in particular, (i) the high- and low-pressure regions in the North Atlantic sector are somewhat asymmetric and geographically shifted relative to the conventional pattern—hence, for example, the relative absence of warming in western Europe; and (ii) there is a positive SLP anomaly over Northern Greenland and part of the Eurasian Arctic Ocean that is absent in the conventional pattern]. Comparisons over the Pacific sector and neighboring regions, by contrast, are of limited utility, given the inability of the GISS-ER model to reproduce the aforementioned La Niña–like feature of MCA-LIA pattern, which strongly affects the Pacific basin. There is no evidence of a positive NAO-AO response in the NCAR simulation (Fig. 4).

The observed patterns of change, even when averaged over multicentury intervals, are unlikely to be entirely forced in nature, as there is also a potentially important role for purely internal, natural variability (*9*). Consistent with this view, we find that individual realizations of the GISS-ER transient response to the MCA-LIA solar forcing difference yield patterns that differ modestly in their details. For at least one realization, for example, the reconstructed warm anomaly over Western Europe is reproduced. In most cases, the basic features discussed above are nonetheless evident (SOM Text).

The paleoclimate reconstructions presented here hold important implications for future climate change. For example, if the tropical Pacific thermostat response suggested by our analyses of past changes applies to anthropogenic climate change, this holds profound implications for regional climate change effects such as future drought patterns. Continued refinement of paleoclimate reconstructions through expanded proxy databases and refinements of CFR methodology, improved estimates of past radiative forcing, and a better understanding of the influence of radiative forcing on large-scale climate dynamics should remain priorities as we work toward improving the regional credibility of climate model projections.

References and Notes

1. E. Jansen *et al.*, in *Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, New York, 2007), pp. 433–497.
2. M. E. Mann, R. S. Bradley, M. K. Hughes, *Nature* **392**, 779 (1998).
3. K. R. Briffa, P. D. Jones, F. H. Schweingruber, T. J. Osborn, *Nature* **393**, 450 (1998).
4. K. R. Briffa *et al.*, *J. Geophys Res. Atmos.* **106**, 2929 (2001).
5. J. Luterbacher *et al.*, *Geophys. Res. Lett.* **26**, 2745 (1999).
6. J. Luterbacher, D. Dietrich, E. Xoplaki, M. Grosjean, H. Wanner, *Science* **303**, 1499 (2004).
7. M. N. Evans, A. Kaplan, M. A. Cane, *Paleoceanography* **17**, 1007 (2002).
8. S. Rutherford *et al.*, *J. Clim.* **18**, 2308 (2005).
9. H. Goosse, H. Renssen, A. Timmermann, R. S. Bradley, M. E. Mann, *Clim. Dyn.* **27**, 165 (2006).
10. See Dataset S1 in the Materials and Methods.
11. The synthetic proxy data in these tests are constructed to have noise characteristics similar to those estimated for actual proxy data. The calibration process, as with real-world reconstructions, is performed over a modern interval that is subject to anthropogenic forcing. The ability of the method to reproduce the earlier variations is then objectively assessed. See (*12*) for further details.
12. M. E. Mann, S. Rutherford, E. Wahl, C. Ammann, *J. Geophys. Res. Atmos.* **112**, D12109 (2007).
13. M. E. Mann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13252 (2008).
14. P. Brohan, J. J. Kennedy, I. Harris, S. F. B. Tett, P. D. Jones, *J. Geophys. Res. Atmos.* **111**, D12106 (2006).
15. M. E. Mann, R. S. Bradley, M. K. Hughes, *Geophys. Res. Lett.* **26**, 759 (1999).

16. T. L. Delworth, M. E. Mann, *Clim. Dyn.* **16**, 661 (2000).
17. P. D. Jones, M. E. Mann, *Rev. Geophys.* **42**, RG2002 (2004).
18. K. R. Briffa *et al.*, *Philos. Trans. R. Soc. B* **363**, 2271 (2008).
19. M. E. Mann, M. A. Cane, S. E. Zebiak, A. Clement, *J. Clim.* **18**, 447 (2005).
20. K. M. Cobb, C. D. Charles, R. L. Edwards, H. Cheng, M. Kastner, *Nature* **424**, 271 (2003).
21. N. E. Graham *et al.*, *Clim. Change* **83**, 241 (2007).
22. V. Trouet *et al.*, *Science* **324**, 78 (2009).
23. C. M. Ammann, F. Joos, D. Schimel, B. L. Otto-Bliesner, R. Tomas, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3713 (2007).
24. S. E. Zebiak, M. A. Cane, *Mon. Weather Rev.* **115**, 2262 (1987).
25. A. C. Clement, R. Seager, M. A. Cane, S. E. Zebiak, *J. Clim.* **9**, 2190 (1996).
26. G. A. Meehl *et al.*, in *Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, New York, 2007), pp. 747–845.
27. J. Brad Adams, M. E. Mann, C. M. Ammann, *Nature* **426**, 274 (2003).
28. D. T. Shindell, G. A. Schmidt, M. E. Mann, D. Rind, A. Waple, *Science* **294**, 2149 (2001).
29. D. T. Shindell, G. A. Schmidt, R. L. Miller, M. E. Mann, *J. Clim.* **16**, 4094 (2003).
30. R. A. Kerr, *Science* **288**, 1984 (2000).
31. N. J. Mantua *et al.*, *Bull. Am. Meteorol. Soc.* **78**, 1069 (1997).

32. M.E.M. and Z.Z. gratefully acknowledge support from the ATM program of the National Science Foundation (grant ATM-0542356). R.S.B. acknowledges support from the Office of Science (BER), U.S. Department of Energy (grant DE-FG02-98ER62604). M.K.H. and F.B.N. were supported by the National Oceanic and Atmospheric Administration (grant NA16GP2914 from CCDD). D.T.S. and G.F. acknowledge support from NASA's Atmospheric Chemistry, Modeling, and Analysis Program.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5957/1256/DC1
Materials and Methods
SOM Text
Figs. S1 to S11
Tables S1 to S5
References
SOM Data

4 June 2009; accepted 5 October 2009
10.1126/science.1177303

Extensive, Recent Intron Gains in *Daphnia* Populations

Wenli Li,[1]* Abraham E. Tucker,[1]* Way Sung,[2] W. Kelley Thomas,[2] Michael Lynch[1]†

Rates and mechanisms of intron gain and loss have traditionally been inferred from alignments of highly conserved genes sampled from phylogenetically distant taxa. We report a population-genomic approach that detected 24 discordant intron/exon boundaries between the whole-genome sequences of two *Daphnia pulex* isolates. Sequencing of presence/absence loci across a collection of *D. pulex* isolates and outgroup *Daphnia* species shows that most polymorphisms are a consequence of recent gains, with parallel gains often occurring at the same locations in independent allelic lineages. More than half of the recent gains are associated with short sequence repeats, suggesting an origin via repair of staggered double-strand breaks. By comparing the allele-frequency spectrum of intron-gain alleles with that for derived single-base substitutions, we also provide evidence that newly arisen introns are intrinsically deleterious and tend to accumulate in population-genetic settings where random genetic drift is a relatively strong force.

Introns are noncoding sequences that interrupt eukaryotic exons and are removed from premature mRNAs by the spliceosomal machinery before translation (*1*–*3*). Intron colonization affects the evolution of gene structure and is a factor in the emergence of genomic and organismal complexity, as newly arisen introns are thought to be intrinsically deleterious owing to the increased mutational target that they impose on their host genes (*4*, *5*). The number of introns in a genome is determined by the relative rates of intron gain and loss over evolutionary time, which differ among lineages. Across eukaryotes, intron numbers range from >100,000 per vertebrate genome to only two in *Giardia lamblia* (*6*, *7*). The fundamental causes of this variation remain controversial (*8*, *9*), partly because of a lack of population-level analyses with the power to infer the properties of recent gain or loss alleles.

The early eukaryotic progenitor has been assumed to be intron-rich on the basis of the presence of introns in homologous positions of orthologous genes of widely divergent eukaryotes (*10*–*12*) and the likely presence of a complex spliceosome in the eukaryotic ancestor (*13*). In this context, intron-poor lineages are assumed to reflect a long-term history of intron loss (*14*). Alternatively, moderate ancestral intron density followed by lineage-specific gains (*15*) may have occurred, even at orthologous positions in divergent taxa (*16*). However, most comparative studies of introns have examined only a small subset of highly conserved genes between deeply divergent lineages, and although some studies have documented unambiguous examples of intron gain (*17*–*19*) and some statistical procedures allow an indirect inference of parallel gains and/or losses (*20*, *21*), comparative studies of taxa with extreme sequence divergence have essentially no possibility of directly inferring parallel intron gains.

Because they potentially retain the molecular signatures of the process of intron origin, intron presence or absence alleles segregating in natural populations provide material to infer gain or loss mechanisms and to estimate taxon-specific turnover rates. Such polymorphisms do exist. A standing intron presence/absence polymorphism was found at a locus in natural isolates of *Drosophila teisseri* (*22*), and two intron-gain alleles segregate with an intron-free version at a locus in the microcrustacean *Daphnia pulex* (*23*). The latter study, in particular, inspired us to look more deeply for evidence of recent intron gain or loss in *D. pulex*.

By artificially removing intron sequences from all predicted gene sequences of the annotated *D. pulex* genome [clone TCO (*24*)] and querying the exon-exon boundaries (n = 110,021) against another *D. pulex* genome sequence (TRO), we detected putative intron-free alleles. After filtering for paralogy and false positives, such as processed pseudogenes, we sequenced the genomic regions surrounding 24 intron presence/absence positions across 84 natural isolates of North American *D. pulex* species as well as in eight *Daphnia* outgroup species. Gene trees constructed from flanking-exon sequence for each presence/absence polymorphism revealed the phylogenetic relationships of the polymorphic alleles, and from these data we inferred that 87.5% (21/24) of the intron polymorphisms reflect recent intron gains, with three reflecting intron losses (figs. S1 to S24). Most of the gains (15/28) were exclusive to Oregon populations, a genetically isolated subclade of North American *D. pulex* (*25*, *26*) with a historically low effective population size (*27*). Active splicing of all polymorphic introns was confirmed with reverse transcription polymerase chain reaction sequencing.

The features of newly arisen introns in *D. pulex* are inconsistent with most hypothesized mechanisms of intron origin (*7*). We found no support for intron gains resulting from tandem duplications of fragments of coding DNA or insertions of transposable elements. Furthermore, the polymorphic intron sequences identified seem to be evolutionary novelties absent from well-characterized eukaryotic and prokaryotic genomes. Except for gains at one locus (Dappu-42116_2, fig. S20), Blastn searches using recently gained intron sequences against the *D. pulex* genome assembly (http://wfleabase.org/blast), *D. pulex* genome trace files, and the full GenBank repository did not retrieve any homologous sequence hits.

We observed that short direct repeats, ranging in size from 5 to 12 base pairs, flank many

[1]Biology Department, Indiana University, Bloomington, IN 47405, USA. [2]Hubbard Center for Genome Studies, University of New Hampshire, Durham, NH 03824, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: milynch@indiana.edu