Redefining the Age of Clovis: Implications for the Peopling of the Americas
Michael R. Waters and Thomas W. Stafford Jr.
Science 315, 1122 (2007);
DOI: 10.1126/science.1137166

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of July 7, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/315/5815/1122.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2007/02/20/315.5815.1122.DC1.html

A list of selected additional articles on the Science Web sites related to this article can be found at:
http://www.sciencemag.org/content/315/5815/1122.full.html#related

This article cites 14 articles, 2 of which can be accessed free:
http://www.sciencemag.org/content/315/5815/1122.full.html#ref-list-1

This article has been cited by 62 article(s) on the ISI Web of Science
This article has been cited by 26 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/315/5815/1122.full.html#related-urls

This article appears in the following subject collections:
Anthropology
http://www.sciencemag.org/cgi/collection/anthro
ceiving array. A TRM made of eight commercial dipolar antennas is placed in the far field, 10 wavelengths away from the receiving array. (The electronic part of the setup is described in fig. S1.) When antenna 3 sends a short pulse (10 ns), the eight signals received at the TRM are much longer than the initial pulse because of strong reverberation in the chamber (typically 500 ns). An example of the signal received at one of the antennas of the TRM is shown in Fig. 2A. When antenna 4 is used as a source, the signal received at the same antenna in the TRM (shown in Fig. 2B) is considerably different, although sources 3 and 4 are \( \lambda/30 \) apart. When these signals are time-reversed and transmitted back, the resulting waves converge respectively to antennas 3 and 4, where they recreate pulses as short as the initial ones (Fig. 2, C and D). Measuring the signal received at the other antennas of the receiving array gives access to the spatial focusing around antennas 3 and 4 (Fig. 2E). The two antennas can now be addressed independently, because the focusing spots created around them are much smaller than the wavelength (typically \( \lambda/30 \)). The diffraction limit is overcome, although the focusing points are in the far field of the TRM.

The origin of the diffraction limit, and the way to overcome it, can be revisited by using the time-reversal concept and the Green’s function formalism, without the explicit use of the evanescent wave concept (20–22). The time-reversal wave, generated by a closed TRM, which converges to its source, is always followed by a spatially diverging wave due to energy flux conservation. Because the focal spot results from the interference of these two waves, the time-reversed field can always be expressed (for a monochromatic wave) as the imaginary part of the Green’s function (22). In a homogeneous medium, the imaginary part of the Green’s function oscillates typically on a wavelength scale. To create focal spots much smaller than the wavelength, one introduces subwavelength scatterers in the near field of the source. Therefore, the spatial dependence of the imaginary part of the Green’s function is modified to oscillate on scales much smaller than the wavelength.

A promising application of time-reversal subwavelength focusing is telecommunications. One way that has been proposed to increase the capacity, or nearly 50 years, it has been generally thought that small bands of humans carrying a generalized Upper Paleolithic tool kit entered the Americas around 11,500 radiocarbon years before the present (\(^{14}\)C yr B.P.). Adjusted \(^{14}\)C dates and a reevaluation of the existing Clovis date record revise the Clovis time range to 11,050 to 10,900 \(^{14}\)C yr B.P. In as few as 200 calendar years, Clovis technology originated and spread throughout North America. The revised age range for Clovis overlaps non-Clovis sites in North and South America. This and other evidence imply that humans already lived in the Americas before Clovis.

**References and Notes**

22. Further details are available as supporting material on Science Online.

**Redefining the Age of Clovis: Implications for the Peopling of the Americas**

Michael R. Waters* and Thomas W. Stafford Jr.2

The Clovis complex is considered to be the oldest unequivocal evidence of humans in the Americas, dating between 11,500 and 10,900 radiocarbon years before the present (\(^{14}\)C yr B.P.). Adjusted \(^{14}\)C dates and a reevaluation of the existing Clovis date record revise the Clovis time range to 11,050 to 10,800 \(^{14}\)C yr B.P. In as few as 200 calendar years, Clovis technology originated and spread throughout North America. The revised age range for Clovis overlaps non-Clovis sites in North and South America. This and other evidence imply that humans already lived in the Americas before Clovis.
Identifying when the Clovis complex first appeared and knowing the complex’s duration is critical to explaining the origin of Clovis, evaluating the Clovis-first model of colonization of the Americas, determining the role of humans in the extinction of late Pleistocene megafauna, and assessing whether people inhabited the Americas before Clovis. We determined a more accurate time span for Clovis by analyzing the revised existing Clovis $^{14}$C date record and reporting high-precision accelerator mass spectrometry (AMS) $^{14}$C ages from previously dated Clovis sites. Our AMS $^{14}$C dates are on culturally specific organic matter—bone, ivory, and seeds—that accelerator mass spectrometers can date accurately ($4, 5$) to precisions of $\pm 30$ years at 11,000 $^{14}$C yr B.P.

Clovis technology has strong Old World antecedents, but Clovis-specific traits (e.g., fluted lanceolate projectile points) probably originated in the New World, south of the continental ice sheets ($3$). Clovis tools and debitage identify and unify archaeological sites over a broad geographic range. Clovis sites and artifacts cluster in North America, especially in the contiguous United States ($1$). A small number of Clovis artifacts have been recovered from Mexico and possibly as far south as Venezuela ($6$). Even though Clovis covers a broad geographic range, only 22 Clovis sites in North America have been directly $^{14}$C-dated ($1, Table 1, and table S1$). The $^{14}$C dates from these sites traditionally place Clovis between 11,500 and 10,900 $^{14}$C yr B.P. ($1, 7, 8$). However, the $^{14}$C dates from 11 of these sites are problematic and do not provide accurate or precise chronological information to determine the age of Clovis ($5$).

Three sites (East Wenatchee, Washington; Blackwater Draw, New Mexico; and Cactus Hill, Virginia) have Clovis diagnostic artifacts but lack precise ages ($5$). Three sites (Lubbock Lake, Texas; Kanorado, Kansas; and Indian Creek, Montana) fall within the Clovis age range but lack diagnostic Clovis artifacts ($5$). The site of Sheridan Cave, Ohio, provides only bracketing ages for Clovis artifacts ($5$). Questions exist about the accuracy of the $^{14}$C dates from Aubrey, Texas ($5$), where diagnostic Clovis artifacts were found. We obtained three dates from the Sheaman site, Wyoming, that averaged 10,305 $\pm 15$ $^{14}$C yr B.P. These dates indicate that the Clovis context at Sheaman is mixed with younger cultural materials ($5$). Finally, associations between Clovis artifacts and $^{14}$C-dated faunal remains at two sites (Wally’s Beach, Canada; and Union Pacific, Colorado) are unresolved ($5$). Because of these problems, we excluded the dates from these sites in assessing the age of Clovis.

This leaves 11 sites with a total of 43 $^{14}$C dates (Table 1 and table S1) ($5$). These sites have assemblages of Clovis artifacts in secure geological contexts. Existing ages from five sites (Anzick, Montana; Paleo Crossing, Ohio; Lehner, Arizona; Murray Springs, Arizona; and Jake Bluff, Oklahoma) already have high-precision $^{14}$C dates on credible materials. We obtained nine new ages from seeds and highly purified bone and ivory collagen for five imprecisely dated sites (Lange-Ferguson, South Dakota; Dent, Colorado; Domebo, Oklahoma; Shawnee-Minisink, Pennsylvania; and Colby, Wyoming) ($4, 5$). In addition, we obtained five ages on human remains from the Anzick site, Montana ($5$). We attempted to date samples from Sloth Hole, Florida, but the samples contained no collagen.

These 43 $^{14}$C dates place the beginning of Clovis at $\sim 11,050$ $^{14}$C yr B.P. (reducing former estimates by 450 $^{14}$C yr) and its end at $\sim 10,800$ $^{14}$C yr B.P. (younger than previous estimates by 100 $^{14}$C yr). Accurate calendar correlation of $^{14}$C ages from the Clovis time period is not currently possible because of correlation uncertainties ($9$). The Clovis-period segment of the INTCAL04 calibration is based on $^{14}$C-dated marine foraminifera and is not accurate for the Clovis time period ($10$). The most accurate calibration for this time period is provided by a floating European tree-ring chronology that is provisionally anchored to INTCAL04 ($11$). Using this tentative calibration ($11$), we estimated that Clovis has a maximum possible date range of 13,250 to 12,800 calendar yr B.P.—a span of 450 calendar years (Fig. 2). By taking the youngest possible calibrated age for the oldest Clovis site and the oldest possible calibrated age for the youngest Clovis site, a minimum range for Clovis is calculated as 13,125 to 12,925 calendar yr B.P.—a span of

---

*To whom correspondence should be addressed. E-mail: mwaters@tamu.edu

---

1Departments of Anthropology and Geography, Center for the Study of the First Americans, Texas A&M University, 4352 TAMU, College Station, TX 77843-4352, USA.

2Stafford Research Laboratories, 200 Acadia Avenue, Lafayette, CO 80026, USA.

---

Fig. 1. Map showing the location of Clovis and other early sites. The numbers correspond to those found in Table 1. Other sites are 31, Monte Verde, Chile; 32, Nenana Complex sites, Alaska; and 33, Broken Mammoth, Alaska.
200 calendar years. The ages for all Clovis sites overlap within this 200-year period, and this time span probably represents the true range of Clovis. However, the absolute calendar placement of the floating tree-ring record is disputed (12). By an alternative calibration (12), the maximum time range for Clovis is 13,110 to 12,660 calendar yr B.P., and the minimum time range is 12,920 to 12,760 calendar yr B.P. (Fig. 3). Regardless of the exact calendar dates, the 200-year duration for Clovis remains secure because the floating dendrochronological sequence provides calendar-year separations between two 14C-dated sites.

The oldest Clovis sites (n = 3 sites) are located in Montana, South Dakota, and Florida; younger Clovis sites are located in the interior (n = 5) of the United States and in the Southwest (n = 2) and East (n = 1). The distribution of dated sites shows no clear indication of north-south or east-west age differences that would indicate movement of people in one direction or another. Instead, Clovis technology seems to have appeared synchronously across the United States at ~11,050 14C yr B.P. This pattern of 14C dates is compatible with two contrasting hypotheses.

First, this pattern could support the idea that there was a rapid spread of Clovis people across an empty continent. Demographic models suggest that people exiting the ice-free corridor could have occupied the contiguous United States within 100 years or less (13). Although there is much speculation about a coastal migration of the first Americans from both Asia and Europe (14, 15), the revised date range for Clovis reopens the possibility of a Late Glacial migration through the ice-free corridor that separated the Laurentide and Cordilleran Ice Sheets. People could have easily traveled through the ice-free corridor after ~11,500 14C yr B.P. (1)—if at least 200 calendar years before the oldest known Clovis date. The biface and blade industry of Nenana (16) was well established at the Broken Mammoth site, Alaska, to 11,770 ± 210 14C yr B.P. (WSU-4351)—at least 300 calendar years before our oldest recalibrated Clovis date. The Nenana lithic assemblage shows strong similarities to the Clovis lithic assemblage (17). It is possible that either Nenana people or others with a biface and blade industry traveled through the corridor, and once south of the ice sheets, they developed the technological hallmarks characteristic of Clovis and spread rapidly across the continent.

An alternative interpretation is that the instantaneous appearance of Clovis across North America represents the rapid spread of Clovis technology through a preexisting but culturally and genetically undefined human population in North America (18). In this case, Clovis technology could have been introduced to this population through a Late Glacial migration of Clovis or Clovis progenitors or developed in situ from a pre-Clovis technology already in the Americas. Regardless of which hypothesis is correct, our revised chronology indicates that Clovis technology spread rapidly.

Faunal remains associated with dated Clovis sites constrain the timing of the extinction of Proboscideans at the end of the Pleistocene. Mammoths and mastodons were an important source of food and raw materials used to manufacture bone and ivory tools (3), as well as perishable items from soft tissues. Proboscidean remains are associated with seven of the well-dated Clovis sites (Lange-Ferguson, Sloth Hole, Dent, Donebo, Lehner, Murray Springs, and Colby), and the last occurrence of mammoth in the United States is dated at ~10,900 14C yr B.P. After this time, Clovis and sites of other complexes (e.g., Goshen and Folsom) contained only bison and other extant species. The extinction of mammoth and mastodon coincides with the main florescence of Clovis.

Our revised ages for Clovis overlap dates from a number of North American sites that are technologically or culturally not Clovis. The earliest dated sites of the Goshen complex (Mill Iron, Montana; and Hell Gap, Wyoming) (19) overlap the age range of Clovis (Figs. 2 and 3, and Table 1, and table S1). This indicates that Goshen is either coeval with the entire range of Clovis or briefly overlaps the later stages of the Clovis time period. Clovis also overlaps the date for the Arlington Springs human skeleton from Santa Rosa Island, California (Figs. 2 and 3 and Table 1) (20). No artifacts were found with the Arlington Springs human remains, and its cultural affiliation is unknown. The presence of human remains on Santa Rosa

---

**Table 1.** Summary of 14C dates from Clovis and Clovis-age sites. Single 14C dates, date ranges, and averaged dates are reported. If multiple 14C dates were available from a single-component site, the dates were averaged with the method in (28). All dates are given at 1σ SD. n, number of dates.

<table>
<thead>
<tr>
<th>Site</th>
<th>Date (14C yr B.P.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clovis sites (credible ages and Clovis diagnostics)</strong></td>
<td></td>
</tr>
<tr>
<td>1. Lange-Ferguson, SD (n = 3)</td>
<td>11,080 ± 40</td>
</tr>
<tr>
<td>2. Sloth Hole, FL (n = 1)</td>
<td>11,050 ± 50</td>
</tr>
<tr>
<td>3. Anzick, MT (foreshaft ages) (n = 2)</td>
<td>11,040 ± 35</td>
</tr>
<tr>
<td>4. Dent, CO (n = 3)</td>
<td>10,990 ± 25</td>
</tr>
<tr>
<td>5. Paleo Crossing, OH (n = 3)</td>
<td>10,980 ± 75</td>
</tr>
<tr>
<td>6. Donebo, OK (n = 1)</td>
<td>10,960 ± 30</td>
</tr>
<tr>
<td>7. Lehner, AZ (n = 12)</td>
<td>10,950 ± 40</td>
</tr>
<tr>
<td>8. Shawnee-Minisink, PA (n = 5)</td>
<td>10,935 ± 15</td>
</tr>
<tr>
<td>9. Murray Springs, AZ (n = 8)</td>
<td>10,885 ± 50</td>
</tr>
<tr>
<td>10. Colby, WY (n = 2)</td>
<td>10,870 ± 20</td>
</tr>
<tr>
<td>11. Jake Bluff, OK (n = 3)</td>
<td>10,765 ± 25</td>
</tr>
<tr>
<td><strong>Clovis sites (indirectly dated and Clovis diagnostics)</strong></td>
<td></td>
</tr>
<tr>
<td>12. East Wenatchee, WA (n = 1)</td>
<td>&lt;11,125 ± 130</td>
</tr>
<tr>
<td><strong>Clovis-age sites (credible ages but no Clovis diagnostics)</strong></td>
<td></td>
</tr>
<tr>
<td>13. Indian Creek, MT (n = 1)</td>
<td>10,980 ± 110</td>
</tr>
<tr>
<td>14. Lubbock Lake, TX (n = 2)</td>
<td>11,100 ± 60</td>
</tr>
<tr>
<td>15. Bonneville Estates, NV (n = 1)</td>
<td>11,010 ± 40</td>
</tr>
<tr>
<td>16. Kanorado, KS (n = 2)</td>
<td>10,980 ± 40</td>
</tr>
<tr>
<td>17. Arlington Springs, CA (n = 1)</td>
<td>10,960 ± 80</td>
</tr>
<tr>
<td><strong>Problematic Clovis and Clovis-age sites</strong></td>
<td></td>
</tr>
<tr>
<td>18. Sheridan Cave, OH (above artifacts, n = 5)</td>
<td>10,600 ± 30</td>
</tr>
<tr>
<td>19. Blackwater Draw, NM (n = 3)</td>
<td>10,920 ± 50</td>
</tr>
<tr>
<td>20. Cactus Hill, VA (n = 2)</td>
<td>11,300 ± 235</td>
</tr>
<tr>
<td>21. Wally’s Beach, Canada (n = 4)</td>
<td>10,920 ± 250</td>
</tr>
<tr>
<td>22. Union Pacific, WY (n = 1)</td>
<td>11,350 ± 80 to 10,980 ± 80</td>
</tr>
<tr>
<td>23. Aubrey, TX (n = 2)</td>
<td>11,280 ± 350</td>
</tr>
<tr>
<td>24. Sheaman, WY (n = 3)</td>
<td>11,570 ± 70</td>
</tr>
<tr>
<td><strong>Ages from other early sites</strong></td>
<td></td>
</tr>
<tr>
<td>25. Mill Iron, MT (Goshen) (n = 4)</td>
<td>10,840 ± 60</td>
</tr>
<tr>
<td>26. Hell Gap, WY (Goshen) (n = 1)</td>
<td>10,955 ± 135</td>
</tr>
<tr>
<td>27. Cerro Tres Tetas, Argentina (pre-Fishtail, n = 5)</td>
<td>10,935 ± 35</td>
</tr>
<tr>
<td>28. Cueva del Minero, Argentina (pre-Fishtail, n = 2)</td>
<td>10,985 ± 40</td>
</tr>
<tr>
<td>29. Piedra Museo, Argentina (pre-Fishtail, n = 2)</td>
<td>10,960 ± 45</td>
</tr>
<tr>
<td>30. Fell’s Cave, Chile (Fishtail, n = 1)</td>
<td>11,000 ± 170</td>
</tr>
</tbody>
</table>
Island is unequivocal evidence that water crafts were used during Clovis time and that a Pacific maritime-coastal adaptation was probably contemporaneous with Clovis. At Bonneville Estates Rockshelter, Nevada (21), the earliest date from a series of hearths is coeval with the Clovis time period and is associated with stone artifacts. Diagnostic artifacts have yet to be found at this level, and it is unknown whether this early horizon is associated with Clovis or stemmed points.

Several sites in South America have yielded 14C dates that are coeval with Clovis (Figs. 2 and 3). These include the early archaeological horizons at Cerro Tres Tetas, Cueva Casa del Minero, and Piedra Museo, Argentina, and the earliest Fishtail point horizon at Fell’s Cave, Chile (Fig. 1, Table 1, and table S1) (8, 22). The actual calendar dates of these South American sites may be slightly more recent because Southern Hemisphere samples have a lower initial 14C content than contemporaneous samples in the Northern Hemisphere. This latitudinal difference causes Southern Hemisphere terrestrial materials to be 5 to 80 14C years older than contemporaneous samples in North America (23). The extent of this interhemisphere offset in 14C years for the Clovis time period is unknown, but it is probably less than 80 years. Even with an 80-year correction to the 14C dates from these four sites, Clovis, Fishtail, and other early complexes in the Southern Cone of South America are still contemporaneous.

The presence of non-Clovis sites that are contemporaneous with Clovis in both North and South America implies that Clovis does not represent the earliest occupation of the Americas. It would probably have taken a minimum of 600 to 1000 years for the first Paleoamericans and their descendents to travel by land from the southern limit of the ice-free corridor to Tierra del Fuego (13, 24)—a distance of over 14,000 km (Fig. 1). However, at most 300 to 350 calendar years separate the oldest possible date for Clovis and the youngest possible ages from the well-dated unequivocal sites in southernmost South America (Figs. 2 and 3). The difference is reduced to 200 calendar years, when the minimum date for the beginning of Clovis in North America and the youngest dates for the sites in South America are considered. It is highly improbable that within 200 to 350 calendar years, people entered North America; adapted to biomes ranging from artic tundra to grasslands, deserts, and rainforests; increased in population; and reached the southern tip of South America within the span of 10 to 18 human generations. This suggests that human populations already existed in the New World before Clovis.

There is an emerging archaeological record that supports a pre-Clovis human occupation of the Americas. Stone tools and butchered mammoth remains dating to ~12,500 14C yr B.P. have been found at the Schaefer and Heibor sites in Wisconsin (25, 26). Older butchered mammoth remains dating to ~13,500 14C yr B.P. have been recovered from the Mud Lake site, Wisconsin (25, 26). In South America, humans appear to have been present at 12,500 14C yr B.P. at Monte Verde, Chile (27). The archaeological data now show that Clovis does not represent the earliest inhabitants of the Americas and that a new model is needed to explain the peopling of the Americas.

References and Notes

1. C. V. Haynes Jr., in *Paleoamerican Origins: Beyond Clovis*, R. Bonnichsen, B. T. Lepper, D. Stanford,
Reports

1126
23 FEBRUARY 2007 VOL 315 SCIENCE www.sciencemag.org

Quantitative Phylogenetic Assessment of Microbial Communities in Diverse Environments

C. von Mering,1* P. Hugenholtz,2 J. Raes,1 S. G. Tringe,2 T. Doerks,1 L. J. Jensen,1 N. Ward,3 P. Bork1†

The taxonomic composition of environmental communities is an important indicator of their ecology and function. We used a set of protein-coding marker genes, extracted from large-scale environmental shotgun sequencing data, to provide a more direct, quantitative, and accurate picture of community composition than that provided by traditional ribosomal RNA–based approaches depending on the polymerase chain reaction. Mapping marker genes from four diverse environmental data sets onto a reference species phylogeny shows that certain communities evolve faster than others. The method also enables determination of preferred habitats for entire microbial clades and provides evidence that such habitat preferences are often remarkably stable over time.

Microorganisms are estimated to make up more than one-third of Earth’s biomass (1). They play essential roles in the cycling of nutrients, interact intimately with animals and plants, and directly influence Earth’s climate. Yet our molecular and physiological knowledge of microbes remains surprisingly fragmentary, largely because most naturally occurring microbes cannot be cultivated in the laboratory (2).

For characterizing this “unseen majority” of cellular life, the first step is to provide a taxonomic census of microbes in their environments (3–6). This is usually achieved by cloning and sequencing their ribosomal RNA (rRNA) genes (most notably the 16S/18S small subunit rRNA). This approach has been extremely successful in revealing the overwhelming diversity of microbial life (7), but it also has some limitations due to quantitative errors: The polymerase chain reaction (PCR) step introduces amplification bias, and it generates chimeric and otherwise erroneous molecules that hamper phylogenetic analysis (8, 9).

Shotgun sequencing of community DNA (“metagenomics”) provides a more direct and unbiased access to uncultured organisms (10–13): No PCR amplification step is involved, and because no specific primers or sequence anchors are needed, even very unusual organisms can be captured by this technique. Although current metagenomics data are still not entirely free of quantitative distortions (mostly due to sample preparation), remaining biases are bound to diminish further with the optimization of yield and reproducibility of DNA extraction protocols (14–16).

To make use of metagenomics data for taxonomic profiling, we analyzed 31 protein-coding marker genes previously shown to provide sufficient information for phylogenetic analysis [they are universal, occur only once per genome, and are rarely transferred horizontally (17)]. We extracted these marker genes from metagenomics sequence data (9), aligned them to a set of hand-curated reference proteins, and used maximum likelihood to map each sequence to an externally provided phylogeny of completely sequenced organisms [tree of life; we used the tree from (17), although any reference tree can be used as long as the marker genes have been sequenced for all its taxa]. Our procedure provides branch length information and confidence ranges for each placement (18) (Fig. 1), allowing us to make use of these data for taxonomic profiling (19) and ecological analysis (20). This unknown sequence (accession no. 1BAA4525), aligned them to a set of hand-curated reference proteins, and used maximum likelihood to map each sequence to an externally provided phylogeny of completely sequenced organisms [tree of life; we used the tree from (17), although any reference tree can be used as long as the marker genes have been sequenced for all its taxa]. Our procedure provides branch length information and confidence ranges for each placement (18) (Fig. 1), allowing us to make use of these data for taxonomic profiling (19) and ecological analysis (20).

References

www.sciencemag.org/cgi/content/full/315/5815/1122/DC1

Materials and Methods

Supporting Online Material

www.sciencemag.org/cgi/content/full/315/5815/1122/DC1

SOM Text

Table S1

1 November 2006; accepted 12 January 2007
10.1126/science.1137166

1 European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. 2 DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. 3 The Institute for Genomic Research, Rockville, MD 20850, USA.

*Present address: University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
†To whom correspondence should be addressed. E-mail: peer.bork@embl.de