Seismic networks record millions of earthquake waveforms every year. The timing, amplitude, polarization, and frequency of their constituent arrivals contain precious information about seismic sources and deep Earth structures. For example, unexpected arrivals can reveal waves scattered or multipath by heterogeneities within Earth, down to the core-mantle boundary (CMB). Mapping heterogeneities in the CMB region is important for understanding the fate of subducted slabs, the origin of hot spot volcanism, and the nature of primitive geochemical reservoirs (1). This mapping can be accomplished by identifying and interpreting scattered arrivals, which is challenging because it requires distinguishing seismic fluctuations from noise and contextualizing arrival amplitudes and timing in a regime where models are not currently available.

Traditionally, these challenges are overcome by focusing on a specific target area. This allows leveraging of geometric arrangements to identify robust signals and aid in their interpretation. A standard procedure is to arrange waveforms by epicentral distance or azimuth to reveal trends in arrivals not predicted by models of the interior. In Earth’s CMB region, robustly identifying scattered waves has led to various discoveries, such as the D" discontinuity (2), ultralow-velocity zones (ULVZs) (3), mega-ULVZs (4), and abrupt variations in wave speed across the boundaries of large low-shear-velocity provinces (LLSVPs) that imply compositional heterogeneity (5). In all cases, interpreting seismic waveforms would be difficult in isolation, without the context provided by seismic waves on nearby paths.

Therefore, previous approaches have limited utility in poorly sampled regions and do not make use of the full statistical power of waveform datasets that span geographically diverse paths.

We conducted a large-scale, systematic search for seismic waves scattered by heterogeneity near the CMB across the Pacific basin. We focused on shear waves diffracting along the CMB (Sdiff), because they sample large areas at the base of the mantle. Waves scattered by heterogeneity in this region arrive after the main Sdiff phase, so the timing and amplitude of these Sdiff "postcursors" can constrain the location and nature of structures producing the scattering (4, 6–9). Because energy reflected from the surface can complicate the identification of Sdiff postcursors, we restricted our attention to waveforms from deep earthquakes (10). This yielded a dataset of ~6000 transverse-component waveforms aligned on the main Sdiff arrival and deconvolved by synthetics computed for a one-dimensional (1D) preliminary reference Earth model (PREM) (11). When plotted by distance between the source and the receiver (Fig. 1A), postcursors cannot readily be identified. Ordering by azimuth only makes sense locally and is not useful when working with data spanning large geographic areas.

We made use of an unsupervised graph-based manifold learning algorithm, "the Sequencer" (10), that orders objects to minimize dissimilarities between neighbors as well as globally across the entire sequence. In our case, the objects are Sdiff waveforms and dissimilarity is given by the Wasserstein metric (also known as the earth mover’s distance) between them (10). This approach can be used to reveal the main trend present in a dataset without requiring any model at all; it has been used in astronomy and has already led to the discovery of a trend relating the mass of supermassive black holes to the properties of their host galaxies (12). We used the Sequencer to optimally order our collection of Sdiff waveforms across the Pacific basin (Fig. 1B). We identified postcursors in >40% of waveforms, which indicates that postcursors are far more common than previously thought. Radial gradients in velocities cannot on their own explain the postcursors, regardless of ordering (Figs. SI, B and D, and S2).

We then explored the geographic distribution of heterogeneities giving rise to the postcursors by binning the fraction of waveforms with postcursors in r° radius bins (Fig. 1C). We found that heterogeneities large enough to scatter shear waves with periods >15 s are pervasive across the Pacific basin. In most locations, a substantial fraction (>0.3) of waves show postcursors (Fig. 1C, cyan). Postcursors are typically absent (Fig. 1C, peach) on paths that do not cross LLSVP boundaries and instead are confined either within the boundaries or well outside of them. Waveforms with and without postcursors seem to coexist at subwavelength scales (~160 km) in most locations (Fig. 1C). Many of the postcursors in the western and northern Pacific (Fig. 1D, R2 and R3) have large delay times (fig. S3B), indicating that they originate from distant scatterers and may travel through complex structures. These areas are suspected to host a partial melt created by paleo-slab from northwestern Pacific subduction zones (13) and a group of small ULVZ patches, along with slab debris beneath the northeastern boundary of the mid-Pacific LLSVP (14). Although large-amplitude postcursors are expected to be associated with a few localities south of the Aleutians (Fig. 1D, R3), bootstrap error estimates show that these signals are not statistically significant (fig. S4).

Notably, nearly all Sdiff wave sampling near Hawaii and the Marquesas Islands shows postcursors (Fig. 1C, pink). The amplitude of postcursors (with respect to the main Sdiff) in the Hawaiian region (Fig. 1D, R1) appears to be three times larger than the typical amplitude found throughout the Pacific basin (Fig. 1D, R2, R3, and R4).

When contextualized across all available data using the Sequencer, the region to the northwest of Hawaii stands out in terms of the prevalence of postcursors, their amplitude, and the spatial extent of the area associated with high-amplitude postcursors (Fig. 1C and D). Therefore, we zoomed in on this region and performed a similar analysis of waveforms with turning points within 20° of Hawaii (Fig. 2A). The nature of such high-amplitude signals is discussed in the methods and materials section of the supplementary materials (10). By plotting the waveforms in the optimal order determined by the Sequencer, we readily identified a subpopulation featuring strong postcursors (Fig. 2B). Because of the limited geographic area, this population can...
also be identified by sorting the waveforms according to azimuth, with delay times increasing up to 40 s (fig. S5A). However, when visualized in the order determined by the Sequencer, the moveout of the postcursors is clarified, and coherent geographic patterns can be mapped.

By plotting each waveform’s position in the sequence at the midpoint of its diffracting path, we found that waveforms with postcursors appear to the northwest of Hawaii, whereas those without appear predominantly to the southeast (Fig. 1D, R4). Moreover, the order identified by the Sequencer reveals a very distinctive spatial pattern (Fig. 2E), with waveforms appearing late in the sequence flanking a region in which midsequence waveforms cluster. This coherent pattern is reflected in both delay time and amplitude of the postcursors observed northwest of Hawaii (Fig. 2, C to E) and does not follow a simple azimuthal trend. Postcursor delay times gradually increase toward the center of the cluster, whereas large-amplitude postcursors (above ~0.5 with respect to the main Sdiff phase) are found mostly southward, near the northern edge of the Pacific LLSVPs (Fig. 2E). The small number of weak postcursors detected to the southeast (Fig. 1D, R4) does not show coherent geographic trends, suggesting that they were not produced by a structure of the type located beneath Hawaii.

We also observed an anticorrelation between postcursor delay time and amplitude in the region (Fig. 2, C and D), consistent with expectations for a localized wave-speed anomaly. This observation motivated us to use the presence of such a correlation as a detector of localized structures. Thus, at each location across the Pacific basin, we estimated the slope and amplitude of the linear fit to the delay time and log amplitude of postcursors with diffracting paths located within 5° of the location. As expected, we found steep negative slopes in the vicinity of Hawaii, across a region that is substantially larger than anywhere else in the Pacific basin (Fig. 3A). In addition, we also detected a similar signature, with a slightly gentler slope, close to the Marquesas (Fig. 3A), indicating a previously unidentified localized wave-speed anomaly. Near both hot spots, the negative slopes are significant at a 99% level of confidence (Fig. 3B). Other locations with possible detections of localized wave-speed anomalies are in the vicinity of Alaska, Kamchatka, and along the northern edge of the Pacific LLSVP, but none are as clear as those near the Marquesas.

After using the Sequencer to reveal the leading trend in our waveform dataset and map the presence of heterogeneity in the Pacific region, we investigated the physical origin of the postcursor signals through waveform modeling. We carried out a systematic suite of wave-propagation simulations through candidate structures based on known features of the CMB region (10). Seismically imaged structures near the CMB span a wide range of sizes, from ~5000-km LLSVPs at one end (15, 16) to ~100-km ULVZs on the other (3, 17). So-called “mesoscale” structures have also been documented, including the Perm (16) and Kamchatka...
anomalies, as well as unusually large ULVZs (4, 6, 20). Therefore, we explored two types of candidate structures: (i) cylindrical regions of reduced or elevated shear-wave velocity \( (V_S) \) with dimensions reminiscent of ULVZs and (ii) undulating boundaries of large low-\( V_S \) regions reminiscent of LLSVPs (Fig. 3). We explored the effects of the lateral abruptness of velocity changes across the boundaries of both types of structures.

Our waveform simulations confirmed that postcursor log amplitude decays linearly with delay time up to 20° away from a low-velocity cylinder (Fig. 4A, inset). The slope of this decay (Fig. 3) is controlled by geometrics and seismic attenuation, while cylinder height and width and the magnitude of its velocity reduction determine the amplitude. A high-velocity cylinder produces postcursors that are four to five times weaker, with similar log amplitude decay (Fig. 4A). Differences in effects of fast and slow anomalies are discussed in the materials and methods (10). Trade-offs among these physical parameters make it impossible to uniquely map slope and zero-delay amplitude to anomaly size, shape, and \( V_S \) reduction (Fig. 3 and fig. S6). Nevertheless, we found that a published model for the Hawaii ULVZ (4, 7, 8) (fig. S7) with a 20% reduction in \( V_S \) [height \( (H) = 25 \) km] (Fig. 4A, dashed red lines) matched the amplitude and delay time of the Marquesas postcursors. This indicates the presence of a mega-ULVZ beneath the Marquesas, although our dataset does not constrain its precise location and characteristics.

However, this model fails to reproduce the large postcursor amplitudes we observe near Hawaii. Instead, an arrow buttaler (\( H = 600 \) km) low-velocity body representing a deeply rooted plume conduit can match the amplitude and delay time of Hawaii postcursors, as does a 50-km-tall cylinder with more gradual boundaries (\( H = 50 \) km, with smooth edges) (Fig. 4A). Waveforms computed in the plume root model for an earthquake which samples the Hawaii region agree well with observations (MB in figs. S8 and S9). Such a narrow (<500-km-wide) mantle plume is not resolvable by travel-time tomography (21) but should become visible in full waveform inversions (22).
incorporating the postcursor waveforms. Although more complicated models cannot be ruled out, such as a 100-km-tall extremely low $V_S$ (25% reduction) structure embedded in the northern edge of Pacific LLSVPs (7), we prefer the plume root structure to explain the amplitudes and waveforms of Hawaii postcursors.

Away from Hawaii and the Marquesas, postcursor amplitude (Fig. 4B, black) is constant with delay time and significantly greater (Fig. S10) than the largest post-Sdiff signal in waveforms that do not show a postcursor when sequenced (Fig. 4B, gray). Postcursor amplitudes cannot be attributed to deconvolution artifacts. To further support our model, we show waveforms from the PREM model (Fig. S10).

Fig. 3. Detection of localized structures. Map of (A) slope and (B) confidence range of the relationship between delay times and log amplitudes for postcursors identified in Fig. 1B and corresponding to raypaths that diffract within 5° of each location. Inset in (A) shows the average postcursor waveforms in Hawaii and the Marquesas (black), average of all nonpostcursor waveforms (green), and average of synthetic waveforms from PREM (magenta). Significance of these postcursor waveforms is discussed in the materials and methods (10) (fig. S10). Large negative value of the slope cluster around Hawaii and close to the Marquesas, indicating the presence of a localized source to the postcursors there. The geographic extent of the Pacific LLSVP (18) is shown (purple contour), as are the five largest hot spots by mass flux (black crosses) (35).
(10) (fig. S10). The spatial pattern of postcursor amplitude variations has a characteristic length scale of ~3000 km (fig. S11), which is similar to the Fresnel width for a scattered arrival with a delay time of 30 to 50 s. We ruled out that the structure beneath Hawaii alone produces the postcursors observed across the northern Pacific, because our waveform simulations indicate that regardless of height, this structure cannot produce postcursors of sufficiently large amplitude on paths >20° away from its center (fig. S8). Rather, two scenarios can plausibly explain postcursors identified by the Sequencer across the northern Pacific: (i) scattering from multiple smaller anomalies distributed across the region and (ii) scattering or multipathing across sufficiently laterally abrupt boundaries of the LLSVPs.

Multiple anomalies distributed across the Pacific could have complex geometries or comprise multiple ULVZs with various geometries and sizes (23). A conglomerate of individual ULVZs that are smaller than our wavelength shear-wave pulses (Fig. 1B) that propagate ing or multipathing across sufficiently laterally distributed across the region and (ii) scatter-

by the Sequencer across the northern Pacific: (i) scattering from multiple smaller anomalies distributed across the region and (ii) scattering or multipathing across sufficiently laterally abrupt boundaries of the LLSVPs.

Alternatively, sharp edges and complexities associated with the LLSVP (27, 28) can broaden shear-wave pulses (Fig. 1B) that propagate nearby (29) or result in multipathing (9) and scattering that can produce postcursors. As demonstrated by our waveform simulations, the postcursors resulting from waves approaching the edges of such a large-scale structure head-on (Fig. 4B, light blue) generally produce weaker-than-observed postcursors. On the other hand, postcursors generated from the waves that transmit obliquely across the same 5% impedance contrast (Fig. 4B, lavender) better reproduce observed amplitudes. The effects of the smoothness of those edges (i.e., sharpness of the velocity along the boundary) were negligible. To test postcursor generation with a more realistic model, we have sharpened the bottom 600 km of the SEMUCB-WMTQ tomographic model (30) to introduce sharper edges for the LLSVPs (M4 in fig. S8B). Synthetic waveforms produced by this sharpened model provide a moderately improved fit to the postcursors observed outside the Hawaii region (fig. S9) but could not generate the high-amplitude postcursors near Hawaii (Fig. 4A and fig. S8).

Exploring a large dataset with the Sequencer enables a data-driven analysis of seismic waveforms without any prior expectations. We anticipate this approach to be useful for many types of datasets beyond seismograms. Often, observed phenomena are driven by a leading effect or parameter. In such cases, there should exist a 1D manifold representing this under-
All authors helped edit the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Seismic data used in this manuscript are available through the IRIS Data Management Center (DMC). The source code of the Sequencer, together with implementation details, can be found at https://github.com/dalya/Sequencer/. An online version of the algorithm is available at http://sequencer.org.

**SUPPLEMENTARY MATERIALS**

science.sciencemag.org/content/368/6496/1223/suppl/DC1

Materials and Methods

Figs. S1 to S16

References (36–44)

14 January 2020; accepted 4 May 2020

10.1126/science.aba8972
Sequencing seismograms: A panoptic view of scattering in the core-mantle boundary region
D. Kim, V. Lekic, B. Ménard, D. Baron and M. Taghizadeh-Popp

Sequencing for seismic structures
Structures illuminated by seismic waves at the core-mantle boundary of the Earth are traditionally found by focusing on a specific target area. Kim et al. used an unsupervised manifold learning algorithm called "the Sequencer" to automatically detect anomalies in seismic data (see the Perspective by Miller). Using this technique, they uncovered structures at the core-mantle boundary across the entire Pacific region all at once. They found many structures previously identified, but also a new, ultra-low-velocity zone beneath the Marquesas Islands. Science, this issue p. 1223; see also p. 1183