doi: 10.1038/nature05666

Supplementary Methods:

We search continuous waveform records during non-volcanic tremor activity for events similar to previously recorded and located LFE "template events" using a matched-filter technique. A matched filter is recognized as an effective means of detecting a known signal in the presence of noise and is extremely powerful when applied simultaneously across multiple stations[8]. For this study, we use waveforms from Japan's Hi-net high sensitivity borehole seismic network operated by the National Research Institute for Earth Science and Disaster Prevention (NIED). As template events, we select 677 low-frequency earthquakes (LFEs) occurring between June 2002 and June 2005; these events represent the best recorded of those located by Shelly et al.[6], with at least 6 three-component Hi-net stations (18 channels) recording each event. This criterion helps ensure that each template event is well located and maximizes our ability to detect any similar events within the continuous data in the presence of noise.

We search tremor episodes systematically for events resembling these template events by cross-correlating corresponding stations and components between the template events' *S*-waves and continuous data. For each template event at each station, we use 4 seconds of the waveform, starting 1 second before the *S*-wave arrival time estimated by the Japan Meteorological Agency (JMA) at that station. We selected this waveform length and positioning in order to be sure to capture the main *S*-wave arrival (the strongest arrival) of the template LFE, without including too much of the waveform that may be dominated by noise or other sources. All waveforms are bandpass filtered between 1 and 8 Hz. We calculate the correlation coefficient as a function of time, shifting the window in increments of 0.05 seconds through the continuous waveforms (see Figs. 3 and S1). At each point we compute the sum of the correlation coefficients across all channels. This provides a measure of waveform similarity that is insensitive to the absolute amplitude of the input signal. The same procedure is repeated for each template event.

When the correlation sum exceeds the threshold, we record a "strong" detection. Because all stations are considered simultaneously, the detected event must originate from nearly the same position as a given template event, with nearly the same source mechanism. At times, multiple nearby template events may detect the same event in the data; in this case we assign the location of the detected event to the location of the template event that registers the most robust detection. The detection robustness is calculated as the ratio between the correlation sum and the detection threshold.

We also attempt to detect events slightly offset from our template events by allowing a slight shift (0.4 s) in waveforms between stations compared to the pattern of the template event. In this case, we take the maximum correlation coefficient from the 0.4-second window at each station before summing. This scheme compromises noise tolerance in order to detect events over a slightly wider area. We maintain this "weak" detection separately from the "strong" detection (with no allowed shift) described above. For the weak detection, the threshold is set at 9 times the median absolute deviation of the distribution above the median.

Using these methods, we observe peaks in the correlation sum when the continuous tremor data exhibits similar waveforms across stations/components. We use the absolute value of the correlation sum as our detection statistic. To define a detection threshold we use the median absolute deviation (MAD), which is defined as the median of the absolute values of the deviations about the median[17]. MAD is an estimator of the variability in a distribution that is robust with respect to outliers, which in this case would correspond to positive detections. For a normally distributed random variable the standard deviation is 1.4826×MAD. We set a conservative threshold for a positive detection at 8 times MAD, which corresponds to approximately 5.4 sigma above the mean for the normal case. This level was chosen to suppress spurious detections while retaining as many legitimate detections as possible. The probability of exceeding 5.4 sigma for a normally distributed random variable is approximately 3.3×10^{-8}. In a 1-hour period, we sample 72,000 time steps for each of 677 template events, giving an expectation of about 1 false detection of a "strong" event per hour. By contrast, we record 1288 "strong" detections during August 29, 2005 17:00-18:00 and 767 "strong" detections during September 2, 2005 19:00-20:00.

The correlation sums, which we use as the detection statistic, are not independent. Moreover, there may be other correlations in the data that lead to non-Gaussian behavior. For these reasons we use a synthetic test to examine the performance of the LFE detection algorithm under controlled conditions. This test works identically to our normal detection procedure, using the actual data but with stations and components randomized. Instead of correlating the template event waveforms among corresponding stations and components, waveforms are correlated with those from a different station. The stations are randomized among those recording the event, with the randomization different for each template event but kept the same during the time-period of interest. Using this procedure for ten complete runs of all 677 template events over a 1-hour period of tremor, we recorded an average of 0.4 "strong" detections and 1.2 "weak" detections per hour, thus confirming an extremely low level of spurious detections.

Applying these procedures, we detect events within continuous tremor many times weaker than our (already weak) template events and can tolerate some degree of interference from multiple, concurrent sources. To examine our detection capability, we perform another synthetic test. In this test, we take a template event not active during a particular hour of tremor and add a scaled-down version of it repeatedly over the 1-hour period. An important benefit of this technique is that it includes the noise of the real data and also tests the ability of the detection algorithm to register a detection when other events are present in the data. An example of such a test is shown in Figure S2, where template event #83, located close to but not in the active region in this example, is scaled by 1/10 and then added into the real tremor sequence every 100 seconds. We then perform our normal detection procedure on the "modified" tremor. This test confirms a remarkable detection capability, with 34 of 35 of the events successfully detected – a type II error rate of less than 5%. Occasional "leakage" of detections to very nearby template events provides an estimate for our location uncertainty of ~3 km.

Supplementary Notes:

17. Rousseeuw, P., and Leroy, A., *Robust Regression and Outlier Detection,* 360 pp., Wiley, (2003).

doi: 10.1038/nature05666

Supplementary Figures and Legends

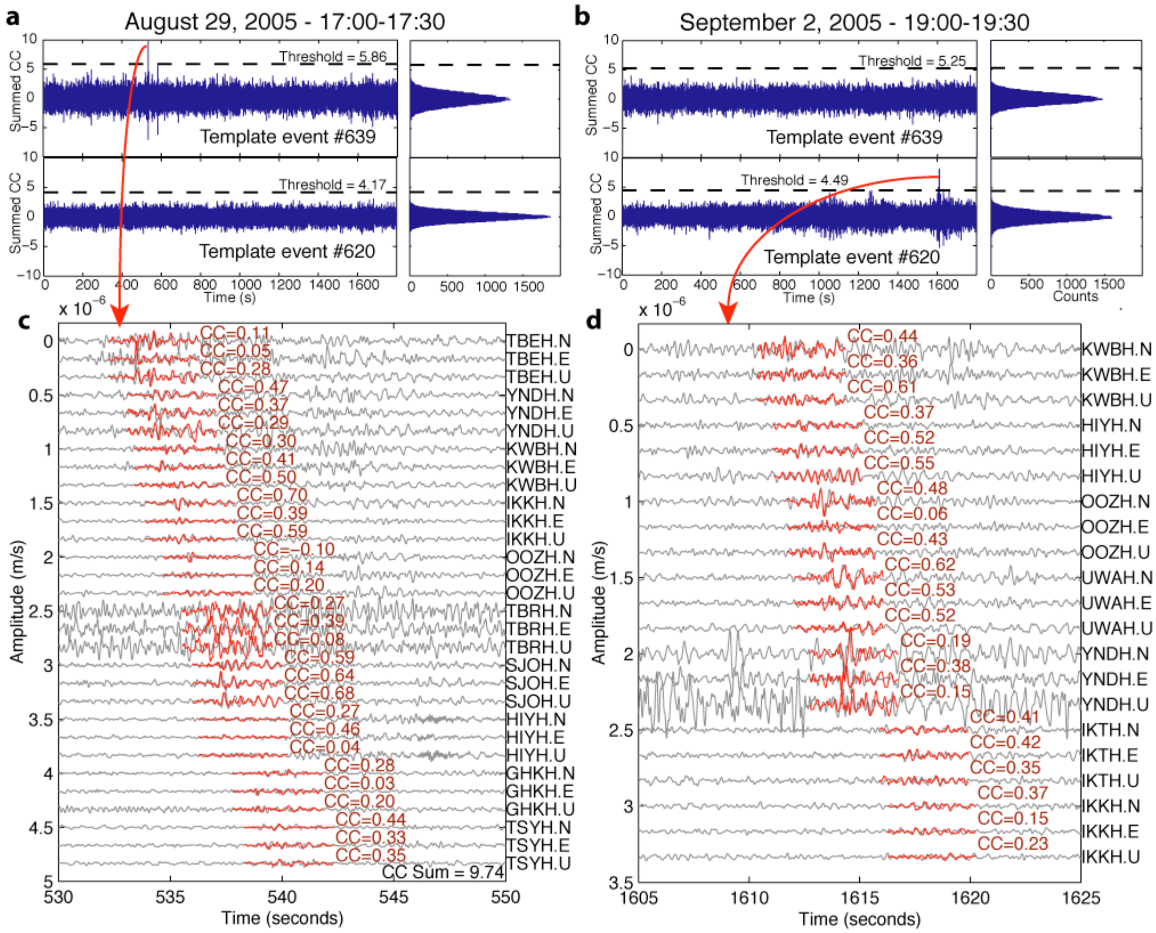

Figure S1 | Correlation sum functions and detection examples for two template LFEs during two tremor episodes. a, Correlation sum functions for template events 639 (upper) and 620 (lower) for the tremor episode August 29, 2005, 17:00-17:30, with histogram of correlation sum values shown to the right. Event #639 has a clear detection while event 620 shows no activity. **b,** Same as **a**, but for the episode September 2, 2005, 19:00-19:30. Notice that in this case, it is event #620 (lower) that is active, while event 639 (upper) is quiet. **c**, Waveforms at the time detection for template event #639 on August 29, 2005 episode. Continuous tremor waveforms are shown in gray and template event waveforms in red for each component of 10 Hi-net stations. The correlation coefficient (CC) for each trace is shown next to the template event waveforms. The station names and components are given to the right of each trace. Additional detections (not shown) are also present during this time window. Waveforms are bandpass filtered between 1 and 8 Hz and template event amplitudes are scaled to match the continuous data. **d,** Same as **c**, but for the detection with template event #620 on September 2, 2005.

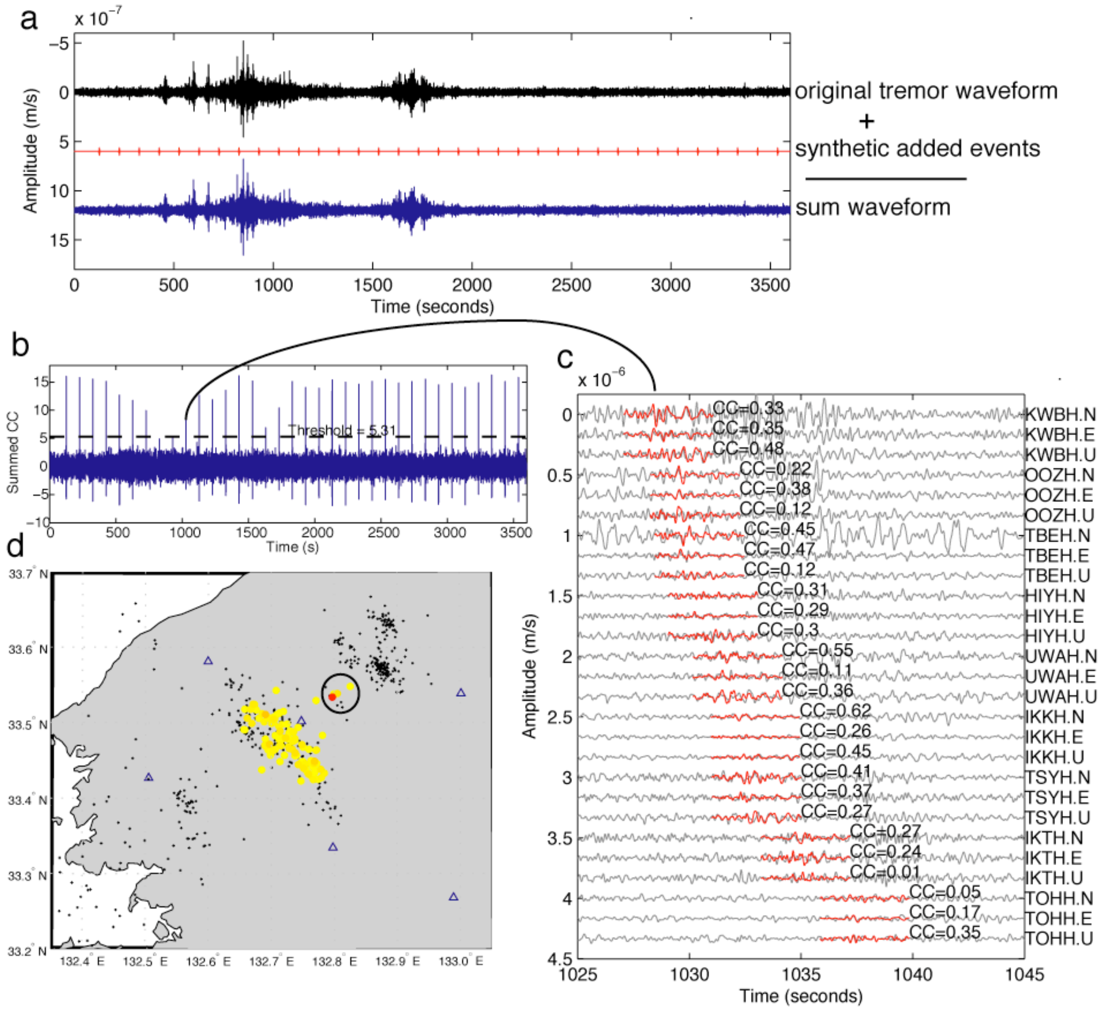

Figure S2 | Synthetic test of LFE detection capability in the presence of noise and other events. **a**, Original tremor waveform (top, black), added test waveform consisting of a single template event (#83) scaled by 1/10 repeated every 100 s (middle, red), and the modified tremor that is the sum of these two signals (bottom, blue). The waveforms shown are unfiltered, east-component from station KWBH, September 2, 2005 19:00-20:00. Note that the maximum amplitude of the scaled event is smaller than the noise level and much smaller than the amplitude of the visible tremor. **b**, Correlation sum function showing the detections for the template event that was inserted into the real data. In this example, 34/35 instances of the scaled template events are recovered. A correlation peak from the undetected event is visible, but falls below the detection threshold due to the strong competing tremor at this time. **c**, Waveforms at the time of detection for the event inserted near t=1030. Gray waveforms are the modified tremor, while red waveforms are from the template event. The correlation coefficient (CC) for each trace is shown next to the template event waveforms. The station names and components are given to the right of each trace. **d**, Map view showing "strong" detected events for this synthetic example, with the color indicating number of detections from 1 (yellow) to 34 (red). Circle in map view shows added event (bright red) and a small amount of "leakage" to 3 nearby LFEs, two of which are located within 0.5 km. The third event is located approximately 3 km away, but has only a single detection compared with 34 for the inserted event. Events outside the circle are detections present in the unaltered tremor data.

Supplementary Movie Legends:

Supplementary Movie 1: Animation showing detected events with time during non-volcanic tremor for the hour beginning September 2, 2005 at 19:00. **Top panel**: Map view of western Shikoku region. Template events are plotted as small black crosses. Colored circles represent a detected event. Filled circles represent "strong" detection (no time shift allowed between stations relative to template event) while open circles represent a "weak" detection (time shift of up to 0.4 s allowed). The shade of the circle represents the robustness of the detection, with light orange a detection just above the threshold level and bright red a detection at 2 or more times the threshold. Each frame represents 2 seconds, with only the strongest detection per frame plotted. The symbols are plotted in reducing size and shading toward black for 3 frames beyond the detection time in order to guide the eye. Blue triangles show station locations; the filled triangle indicates the station with waveforms plotted in the bottom panel. The time listed at the top corresponds to the approximate time of the first S-wave arrival at any station. **Bottom panel:** A sample velocity waveform, one hour in duration, corresponding to the time-period of the animation. Waveform is bandpass filtered between 1 and 8 Hz. Portions plotted in red indicate times with a detected event similar to a template event. Note that nearly all high-amplitude segments of the waveform are matched by LFEs. The vertical blue bar indicates the point in time represented in the map view.

Supplementary Movie 2: Same as Supplementary Movie 1 but for the hour beginning August 29, 2005 at 17:00. In this case, tremor is active in a region distinct from that active in Supplementary Movie 1.