Supplementary information to:
Global patterns of Earth's dynamic topography since the Jurassic

S1 Cluster analysis methodology

The k-means algorithm partitions a given data set S of N observation vectors into k ($\leq N$) clusters. The inputs for such a calculation are the N d-dimensional observation vectors, e.g. measurements of d quantities at N different locations (vectors $\boldsymbol{x}_i \in S$, $i = 1,...,N$), and the number of expected clusters, k. The algorithm then computes the sum

$$s = \sum_{j=1}^{k} \sum_{x_i \in S_j} \left\| x_i - \mu_j \right\|^2$$

where S_j ($j = 1,...,k$) is the j-th partition or cluster ($S = \mathrm{U}_{j=1} S_j$, S_j pairwise disjoint) and the $\boldsymbol{\mu}_j$ are its so-called centroids, or averages. The generalised distance (norm) of each point in a cluster to the cluster centroid, $||\boldsymbol{x}_i - \mu_j||$, constitutes the so-called similarity or affinity matrix. In an iterative process the algorithm is aiming to minimise that sum by refining the choice of the centroids. The final result is obtained when either the maximum number of iterations has been completed or the change of the calculated sum has fallen below a given threshold. For the current case we take the time series of each sample point out of the N sample locations on the chosen grid as a d-dimensional point (d - number of time steps observed) that serves as input into the clustering algorithm, thus $\boldsymbol{x}_i = \boldsymbol{h}_i$, where $\boldsymbol{h}_i$ is one d-dimensional vector of elevations for all time steps from some cluster S_j. It is important to note, that, although sample vectors represent time series, due to the completely symmetric nature of the norm with regards to coordinates (i.e. the individual times) the algorithm is agnostic of histories of samples. If certain time steps would be swapped likewise for all sample locations the distance (norm) between them would still yield the same result.

As the appropriate number of clusters is a subjective choice and difficult to determine *a priori*, we test a range of possible values for model M1, trying to find a minimal k that represents our classification. In other words, the choice of k itself need not necessarily match our number of categories. It should, however, reflect the geodynamic classification in some way (by combining several cluster into one). Thus, we can corroborate in an independent and unbiased fashion through an automated analysis that the assumptions of our classification scheme are correct.

We perform the analysis on time-series of dynamic topography for onshore regions only, with topography reconstructed from the mantle frame of reference into the frame of reference for each plate using the plate tectonic reconstruction software GPlates (Boyden et al., 2011). We evaluate time series of dynamic topography on a grid of nearly 2 x 10^5 equidistantly spaced nodes, providing a resolution of ~ 28 km on the surface, sampled at 5 Myr intervals in time. The cluster analysis was applied for the dynamic topography time-series from 150 Ma to present – this excludes the first 50 Myr of the model run where the topography is most uncertain.

S2 Additional figures

This section lists supplementary files with additional figures. The file “02_SUPP_DynaTopo_10Myr.pdf” shows a to Fig. 4A corresponding sequence of global dynamic topography in 10 Myr intervals; “03_SUPP_DynaTopo_plateframe_10Myr.pdf” comprises an analogous sequence in the plate frame of reference. For Australia “04_SUPP_DynaTopo.AUS_10Myr.pdf” has a to Fig. 6 corresponding sequence of dynamic topography in 10 Myr steps. Finally, “05_SUPP_Cluster_analysis.pdf” holds results of Sect. 3.3.4 for all number of clusters $k \neq 2$.