Align DOCUMENTATION

1. DOWNLOADING & PREPARATION

First, download the ZIP archive named Align. Unpack the ZIP on your Desktop to create a new directory with the same name. *Note that Align will save files to this directory*. Also, check that you have installed R on your computer. Additional required R packages include shiny, dplyr, ggplot2, ggtext, readr, viridis, and align (an R package that we wrote that hosts the Hay et al. (2019) algorithm translated into R; available through Cran R Project: https://cran.r-project.org/web/packages/align/index.html). These packages should download automatically if you do not already have them installed. For those who are not already experienced with R, we strongly recommend using RStudio as your coding environment.

2. STARTING THE DTW GUI

Double-click on Align.r, which will open the entire script in RStudio. Click the small green triangle near the center-top of the RStudio window to run the GUI script. This will open a separate window (the GUI) that you will use exclusively in the remainder of this guide. We do not recommend modifying any portion of the code unless you have prior experience with R and Shiny.

3. PREPARING YOUR DATA

Before you can utilize *Align*, you must first format your data correctly. Three example datasets (synthetic data based on Trampush and Hajek, 2017; Jiang et al., 2012; Husson et al., 2015) are included in the correct data format, which we recommend you mimic in the preparation of your own data.

Your target dataset should be saved in a .csv file with three data columns: $\delta^{13}C_{carb}$ in A, stratigraphic height in B, and the name in C (include the exact header labels: d13c, height, and name). Your candidate dataset(s) should be saved in a .csv file with three data columns for each candidate record and a blank column between candidate datasets. For example, $\delta^{13}C_{carb}$ data from Candidate #1 in A, stratigraphic height from Candidate #1 in B, name for Candidate #1 in C, column D is left blank, $\delta^{13}C_{carb}$ data from Candidate #2 in E, stratigraphic height from Candidate #2 in F, name for Candidate #2 in G, and so on (include exact header labels for each data column: d13c_1, height_1, name_1, d13c_2, height_2, name_2, etc.). You can include up to 3 candidate datasets in your candidate data file.

4. RUNNING THE DTW ALGORITHM WITH THE GUI

Once you have downloaded the necessary programs and packages, and prepared your data, you are ready to use *Align*. For these steps, remain in the "Generate New Alignments" tab at the top of the GUI window.

4.1 Uploading your data

First, using the "Browse…" button, upload your target dataset file. Second, using the next "Browse…" button, upload your candidate dataset(s) file. *Align* will automatically detect the number of candidate datasets, and the names of the target and candidate datasets from the .csv files.

4.2 Viewing your data

Once you have uploaded, click the green "Plot" button to verify that your data look correct. You can use the "Candidate record" drop down menu to select which candidate record you would look to view plotted next to the target record. You must plot these data before moving on to the next step.

4.3 Running the algorithm

Once you have uploaded and plotted your data, and chosen your g and edge ranges and increments, click the green "Run DTW algorithm" button to run the DTW algorithm and generate the alignment libraries for these target–candidate pairings. Look for notifications in the lower right hand of the window indicating algorithm progress. Plots of each target–candidate alignment, as well as accompanying .csv files, are output in the "Output_Data" and "Output_Images" directories within the [*candidate name-target name*] directory, which resides in the "Output" directory.

5. ALIGNMENT LIBRARY VIEWER

The alignment libraries for each target–candidate pairing can be viewed using the GUI viewer tool. Navigate to the "Alignment library viewer" tab at the top of the GUI window to use this tool.

5.1 Narrowing alignment libraries

First, use the "Candidate record" drop down menu in the upper left corner of the window to select the candidate record of interest. Next, use the two slider bars to *narrow* the alignment library as desired. The xc cutoff sets the minimum correlation coefficient threshold for inclusion in the library (default is set to 0.80). The overlap cutoff sets the minimum percent overlap threshold for inclusion in the library, where a 10% overlap threshold indicates that the aligned candidate record overlaps with at least 10% of the target record (default set to 10%). Lastly, type a nickname for these criteria: this will be used to save your narrowed alignment library data so you can experiment with different narrowing criteria without 'losing' previous libraries. Once you have selected your criteria and provided a nickname, click the green "Narrow alignment library" button to generate the narrowed library. Note: you ***MUST*** click the "Narrow alignment library" button ***prior*** to clicking the "Plot alignment" button below. Failure to do so may result in the app crashing, requiring you to start over.

5.2 Plotting different candidate record alignments

Once you have clicked the green "Narrow alignment library" button, the alignments that fit the indicated criteria will populated the lower "Candidate record" drop down menu. To display these alignments, select the alignment of interest from the drop-down menu and click the green "Plot alignment" button. Each plot is output to the "Output_Images" directory in a new directory named according to the criteria nickname you provided above.

HOW DYNAMIC TIME WARPING PRODUCES A LIBRARY OF STRATIGRAPHIC ALIGNMENTS

We provide an abbreviated explanation of how the dynamic time warping algorithm produces a library of stratigraphic alignments. We highly recommend that first-time users review the more detailed explanation provided in Hagen et al. (2023, *GSA Today*). First, target and candidate matrices are constructed whose number of rows and columns equal the length of the

target and candidate $\delta^{13}C_{carb}$ sequences, respectively. The target $\delta^{13}C_{carb}$ values fill target matrix column 1 and are replicated to fill all remaining columns. The candidate $\delta^{13}C_{carb}$ values are transposed to fill candidate matrix row 1 and replicated to fill the remaining rows. The next step is to construct an n by m matrix of all of the possible $\delta^{13}C_{carb}$ pairings from the target and candidate sequences. Each matrix element is computed as the difference between an index in the target (t_n) and the candidate (c_m) sequences: $C(n,m) = (t_n - c_m)$ and squared to give a squared-difference matrix.

An alignment takes the form of a 'warping path' that assigns each candidate index a target index by minimizing the sum of the squared differences (or 'cost'). This path is achieved through successive diagonal, horizontal, and vertical steps across the squared difference matrix, each of which implies a bed-to-bed alignment. The DTW algorithm objectively finds an optimal pathway in terms of a sequence of diagonal, vertical, and horizontal steps that minimize the associated sum of squared residuals. A diagonal step implies an equivalent rate of *relative* sediment accumulation between the candidate and target time series. A vertical or horizontal step instead inserts a hiatus in deposition at the candidate or target sections, respectively.

When aligning $\delta^{13}C_{carb}$ sequences without independent temporal constraints (e.g., biostratigraphic and/or geochronologic) stratigraphers have little or no *a priori* information about the total temporal overlap with the target section, nor the relative rates of sediment accumulation between target and candidate sections. To address these uncertainties, the algorithm explores various optimal warping paths across the squared difference matrix, conditional on the systematic application of the *edge* and *g* penalty functions (see below) that alter the values of the squared difference matrix and thereby favor specific stratal pairings.

The *edge* penalty function explores whether the two sequences span the same total interval of time and is so named because the right and bottom squared difference matrix edges align the stratigraphically highest (youngest) target and candidate $\delta^{13}C_{carb}$ values whereas the left and top edges align the lowest (oldest) $\delta^{13}C_{carb}$ values. *Edge* values > 1 increase the value of the squared difference for a specific stratal pairing, discouraging their alignment, whereas when $0 < edge < 1$, stratal pairings are encouraged.

The *g* penalty function is useful for enforcing various levels of similarity of sediment accumulation rate(s) at the two stratigraphic sections throughout their shared deposition history using a range of *g* values. Values of $g > 1$ penalize stretching or squeezing by increasing the augmented cost of all off-diagonal matrix cells, and the opposite is true for $g < 1$; a *g*-value equal to 1 does not augment the cost matrix.

Next, the algorithm calculates the accumulation of cost, assembling a cumulative difference matrix (CDM; see accompanying publication for more details). Every possible pairing of *g* and *edge* values from the input ranges produces a CDM, and the warping path across the CDM begins at the lower-right corner and progressively steps horizontally, diagonally, or vertically to the minimum value of the eight adjacent cells, always looking two steps ahead. For each CDM, the corresponding $\delta^{13}C_{carb}$ alignment begins with the stratigraphically lowest cell of the starting edge (the right column or bottom row and terminates upon meeting an end edge (the left column/top row). Note when the algorithm encounters equivalent values the diagonal is adopted to maximize temporal correspondence by minimizing the insertion of hiatuses. For the adopted *edge* and *g* parameter values, the warping path specifies the *globally* optimal alignment of each $\delta^{13}C_{carb}$ value of the candidate sequence with the target sequence, with empty rows representing target $\delta^{13}C_{carb}$ values with no time-equivalent at the candidate section (an imposed hiatus). By repeating this process for a range of *edge* and *g* values, the algorithm systematically generates

alignments that encapsulate a spectrum of assumptions about the shared temporal history (via *edge*) and relative rates of sediment accumulation (via *g*) at the target and candidate stratigraphic sections. Different pairings of *edge* and *g* values can produce visually distinct alignments. Together we present the objective alignments arising from all *edge* and *g* pairings as an alignment library for further parsing by statistical analyses and geological insight.

ADDITIONAL RESOURCES

This documentation provides a simple walk-through for using *Align*, as detailed in Hagen et al., 2023, *GSA Today*. For additional details, please refer to the publication.