BIOINFORMATICS

Vol. 17 no. 6 2001
Pages 495–508

# Aligning gene expression time series with time warping algorithms

John Aach and George M. Church*

*Department of Genetics and Lipper Center for Computational Genetics, Harvard Medical School, 200 Longwood Ave, Boston, MA 02115, USA*

Received on September 26, 2000; revised on February 23, 2001; accepted on February 28, 2001

## ABSTRACT

**Motivation:** Increasingly, biological processes are being studied through time series of RNA expression data collected for large numbers of genes. Because common processes may unfold at varying rates in different experiments or individuals, methods are needed that will allow corresponding expression states in different time series to be mapped to one another.

**Results:** We present implementations of time warping algorithms applicable to RNA and protein expression data and demonstrate their application to published yeast RNA expression time series. Programs executing two warping algorithms are described, a simple warping algorithm and an interpolative algorithm, along with programs that generate graphics that visually present alignment information. We show time warping to be superior to simple clustering at mapping corresponding time states. We document the impact of statistical measurement noise and sample size on the quality of time alignments, and present issues related to statistical assessment of alignment quality through alignment scores. We also discuss directions for algorithm improvement including development of multiple time series alignments and possible applications to causality searches and non-temporal processes ('concentration warping').

**Availability:** Academic implementations of alignment programs genewarp and genewarpi and the graphics generation programs grphwarp and grphwarpi are available as Win32 system DOS box executables on our web site along with documentation on their use. The publicly available data on which they were demonstrated may be found at http://genome-www.stanford.edu/cellcycle/. Postscript files generated by grphwarp and grphwarpi may be directly printed or viewed using GhostView software available at http://www.cs.wisc.edu/~ghost/.

**Contact:** church@arep.med.harvard.edu

**Supplementary information:** http://arep.med.harvard.edu/timewarp/supplement.htm.

*To whom correspondence should be addressed.

## INTRODUCTION

Recently developed high throughput assays for mRNA expression such as DNA microarrays, oligonucleotide arrays, microbeads, and Serial Analysis of Gene Expression (SAGE) (Brenner *et al.*, 2000; DeRisi *et al.*, 1997; Lockhart *et al.*, 1996; Velculescu *et al.*, 1995) have enabled researchers to study biological processes systematically at the level of gene activity. Clustering of expression data gathered by these means to functionally characterize genes (Eisen *et al.*, 1998; Tamayo *et al.*, 1999; Tavazoie *et al.*, 1999), and to classify samples and conditions (Aach *et al.*, 2000; Alizadeh *et al.*, 2000; Bittner *et al.*, 2000; Golub *et al.*, 1999) is now commonplace. An important area of application of these techniques is the study of biological processes that develop over time by collecting RNA expression data at selected time points and analyzing them to identify distinct cycles or waves of expression (see Table 1). Progress in the development of high throughput protein level assays (Gygi *et al.*, 1999, 2000) suggests that similar techniques will soon be used in the area of protein expression analysis. We will focus on RNA expression data.

Biological processes have the property that multiple instances of a single process may unfold at different and possibly non-uniform rates in different organisms, strains, individuals, or conditions. For instance, different individuals affected by a common disease may progress at different and varying rates. This presents an issue for analysis of biological processes using time series of RNA expression levels: to find the time point of one series that corresponds best to that of another, it is insufficient to simply pair off points taken at equal measurement times. Analysis of such time series may therefore benefit from the use of alignment procedures that map corresponding time points in different series to one another.

Dynamic time warping is a variety of time series alignment algorithm developed originally for speech recognition in the 1970s (Sakoe and Chiba, 1978; Velichko and Zagoruyko, 1970). Similar to algorithms used for sequence alignment, time warping aligns two time series against each other. Whereas sequence align-

**Table 1.** Examples of gene expression time series published in literature including unevenly sampled time series. Non-time series data points (e.g. mutants) published with the studies are not described

| Study | Published time points | References |
|---|---|---|
| Diauxic shift, yeast | 9, 11, 13, 15, 17, 19, 21 h | DeRisi *et al.* (1997) |
| Sporulation, yeast | 0, 0.5, 2, 5, 6, 7, 9, 11.5 h | Chu *et al.* (1998) |
| Cell cycle, yeast.<br>Cells synchronized by cdc28-ts | 17 time points from 0, every 10 min | Cho *et al.* (1998) |
| Cold shock, yeast<br>Heat shock, yeast<br>Reducing shock, yeast | 0, 20, 40, 160 min<br>0, 10, 20, 40, 80, 160 min<br>15, 30, 60, 120 min | Eisen *et al.* (1998) |
| Cell cycle, yeast[a]<br>Cells synchronized by | | Spellman *et al.* (1998) |
| • Alpha factor | 18 time points from 0, every 7 min | |
| • cdc15-ts | 10, 30, 50, 70, then every 10 min to 250, then 270, 290 (24 time points) | |
| • Elutriation | 14 time points from 0, every 30 min | |
| Embryo development, *Drosophila* | $\geqslant$ 18 h BPF[b], 4 h BPF[b], 0 h PF[b], 3, 6, 9, 12 APF[b] | White *et al.* (1999) |

[a] Collection of time series analyzed in this article.
[b] PF = puparium formation, BPF = before PF, APF = after PF.

ment algorithms consider the similarity of pairs of single bases or residues taken one from each sequence, time warping considers the similarity of pairs of vectors taken from a common k-dimensional space (feature space) taken one from each time series. Here the feature space comprises vectors of RNA expression levels from a common set of k genes. The time warping algorithms developed here are global alignment algorithms; therefore Needleman–Wunsch presents the most analogous sequence algorithm (Needleman and Wunsch, 1970). While in its most general form time warping makes no assumptions about the evenness or density of data sampling in the time series it aligns, simplifications and efficiencies are often possible when sampling rates are constant and of high density. These conditions are easily met when sampling speech data through appropriate electronics and data processing, but not for RNA expression level data where collection of data at a time point involves laborious and costly steps. Examples of unevenly and sparsely sampled RNA expression time series are common in the literature (see Table 1), and this will surely be true of protein time series as well. As a result, time warping algorithms developed for speech recognition cannot generally be directly applied to typical expression level time series. To demonstrate time warping on these data, we therefore implemented time warping algorithms for expression data from first principles as described in Kruskal and Liberman (1999), including an interpolative algorithm that to our knowledge has never been previously implemented.

Data quality, completeness, and normalization present additional issues when applying time warping to gene expression data. The different high throughput mRNA expression level assays are each affected by different sources and sensitivities to error and generate RNA expression levels using different kinds of normalizations, data quality indicators, and degrees of data completeness. For instance, expression levels derived from microarray and microbead studies are generally presented as ratios of the RNA expression levels of genes in an experimental condition compared to their levels in a control condition, while in oligonucleotide array and SAGE studies they are normalized for experimental conditions alone without reference to control conditions but in very different forms (normalized hybridization intensities for arrays and tag counts for SAGE). Such methodological differences, as well as differences in strains and cell culture conditions across studies, affect the comparability of data derived from different studies (Aach *et al.*, 2000). While these factors will affect time warping, we minimize them during this initial demonstration by confining attention to a collection of RNA expression level time series of the *Saccharomyces cerevisiae* cell cycle published in a single

study (Spellman *et al.*, 1998) where each series used cells specially prepared by a different technique to move synchronously through the cell cycle (see Table 2).

Below we present results showing that time warping produces alignments that are expected given the series being compared. We show that the stability of alignments is improved by using large sets of genes with low measurement noise levels. We compare time warping with clustering as a way of mapping corresponding time points, and assess the use of alignment scores in judging the quality of alignments. In discussion we describe potential enhancements to and alternative uses of the algorithms, compare time warping with Singular Value Decomposition (SVD) and Fourier analysis, and call attention to the possibility of time series superpositions.

## SYSTEMS AND METHODS

### Yeast cell cycle time series data sets

Data from Spellman *et al.* (1998), downloaded originally from the reference web site, were extracted from the ExpressDB database (Aach *et al.*, 2000). Briefly, original data collection and normalization were as follows: data were collected using microarrays spotted with amplified genomic DNA for all *S.cerevisiae* ORFs. Three time series of RNA expression levels were obtained for cultures synchronized by different methods and then released at time 0: alpha mating factor pheromone, a temperature sensitive cdc15 mutant, and elutriation (see Table 2). Hereafter we refer to these as the alpha, cdc15, and elu series, respectively. Data were presented as normalized $\log_2$ ratios of each gene's RNA expression level in the experimental sample at the time point to its level in a control culture sample at the same time point, where the control culture comprised unsynchronized cells of the same strain grown under the same conditions as the experimental culture, and where $\log_2$ ratios were normalized by adjusting each gene's $\log_2$ ratio to 0 over its time series. To help evaluate time alignments we desired more precise numerical estimates of each series' cell cycle period than the approximations presented in the original reference, and so computed the period of each series from the gene expression data by a variant of the analysis used there to identify which genes were cell cycle regulated. Results (see Table 2) are in good accord with approximate period information given in the original reference, although computed periods tend to be somewhat longer. Details on these calculations are on our web site.

Application of time warping algorithms to these time series required using subsets of genes which were shared by the two series compared and for which expression levels were available in the data for each time point in the two series. Several subsets of genes used for testing aspects of the algorithms are described in Table 3. Demonstrations in this article will focus mainly on alignments between the alpha and cdc15 series as alignments of these series appeared to be of higher quality than either of them with elu, possibly because they both cover ∼2 periods of the cell cycle whereas elu covers only ∼1 (Table 2). elu was also excluded for some analyses in the original reference for apparently similar reasons.

### Time warping programs

Four time warping programs genewarp, genewarpi, grphwarp, and grphwarpi were developed in C$^{++}$. NT executables for all programs are available from our web site. The four programs work in pairs: genewarp and grphwarp, and genewarpi and grphwarpi. genewarp performs a simple time warping, while genewarpi performs an interpolative warping (see Section Algorithm and implementation). grphwarp and grphwarpi are graphics generation programs that take a file produced by genewarp and genewarpi, respectively, to generate graphics for the alignments. Generated graphics are in the form of PostScript files (Adobe Systems Incorporated, 1999) that may be directly printed on compatible printers or viewed using software such as GhostView (http://www.cs.wisc.edu/∼ghost/) or Adobe Photoshop (Adobe, San Jose, CA). Generated graphics contain four parts: graphs of the two input time series in real time, a graph displaying the optimal path through the dynamic programming matrix (see Section Algorithm and implementation), and a graph of the alignment of the input series in a time frame where time values of aligned time points are the averages of the aligned input series time point values (the trajectory average time frame in Kruskal and Liberman (1999)). grphwarp and grphwarpi can be instructed to display only specific sets of genes from the input time series and time alignment, a useful feature when the number of input genes is large. It takes ∼3 s to align time series using genewarp containing 18 and 24 time points for 495 genes on a 200 MHz Pentium II computer with 96 MB RAM. Additional information on the programs is given on our web site along with samples of time series and parameter files that may be used to run them.

## ALGORITHM AND IMPLEMENTATION

We implemented algorithms based on Kruskal and Liberman (1999). Details are provided on our web site. Briefly, if expression levels of k genes are tracked during the unfolding of a biological process, the process can be conceived as tracing out a trajectory in k-dimensional space (k-space) over time. A time series $a$ for the process then consists of a set of time points $i$ ($0 \leqslant i \leqslant n$) each corresponding to a particular measurement time $t_i$, whose expression level values $\mathbf{a}_i$ define points in k-space on

**Table 2.** Periods and related information for time series used for demonstration of time warping

| Series | Cell cycle period[a] | Periods in series[b] | Periods / time point[c] | Initial phase[d] |
|---|---|---|---|---|
| alpha | $67.5 \pm 6.5$ | 1.8 | 0.10 | G1 |
| cdc15 | $119.0 \pm 14.0$ | 2.4 | 0.08 | Late M |
| elu | $422.5 \pm 77.5$[e] | 0.9 | 0.07 | G1 |

[a] Cell cycle periods and errors computed from RNA expression data as described in web supplement. Values are in minutes.
[b] Periods represented in entire time series (see Table 1).
[c] Periods represented by interval between consecutive time points. For the cdc15 series the interval was taken to be 10 min and applies to 18 out of the 24 series time points (see Table 1).
[d] Cell cycle phase of series at time 0 (Spellman *et al.*, 1998).
[e] Error value presented for elu series is an underestimate (see our web site for additional details).

**Table 3.** Subsets of genes employed in time warping experiments. Non-null genes are genes for which expression levels were available in the original reference data for every time point in the time series under consideration

| Gene subsets | Description | Usage |
|---|---|---|
| pgt50 | Non-null genes in both the alpha and cdc15 series with variances >50th percentile in both series individually and together, sorted in descending order by combined series variance (990 genes). | • Alignment stability testing (Table 4)<br>• Time series used for most algorithm testing |
| pgt33 | Same as alpha/cdc15-pgt50 except for use of 33rd percentile threshold (1549 genes). | • Alignment stability testing (Table 4) |
| pgt33-990 | Last 990 genes of alpha/cdc15-pgt33, containing genes with combined series variance percentiles >33 and $\leqslant$80.1. | • Alignment stability testing (Table 4) |
| pgt50-odd, pgt50-even | Division of pgt50 where each subset contains every other gene. These sets have approximately the same variance distribution as pgt50 but half the sample size (495 genes). | • Alignment stability testing (Table 4) |
| elu-pgt50 | Non-null genes in the elu series alone with variances >50th percentile (2883 genes). | • elu half series alignment (see text) |
| MET, CLB2, CLN2 | Genes from the MET, CLB2, and CLN2 clusters identified in (Spellman *et al.*, 1998) that are non-null in each of the alpha, cdc15, and elu series (11, 27, and 22 genes, respectively). | • Alignment visualization<br>• Small cluster alignments (web site) |

the trajectory (sample points). Relative to the $a$ series, a second time series $b$ for a different instance of the process may contain a set of time points $j$ ($0 \leqslant j \leqslant m$) corresponding to different times $u_j$, and whose sample points $\mathbf{b}_j$ may come from a trajectory that traces through different regions of k-space or traces through the same regions at different rates (see Figure 1a). Simple time warping uses dynamic programming to find the (many-to-many) mapping $i \leftrightarrow j$ that minimizes a weighted sum of the k-space distances between the corresponding sample points, subject to constraints of order preservation ($i < i'$ implies $j \leqslant j'$ and $j < j'$ implies $i \leqslant i'$) and globality (every $i$ maps to some $j$ and *vice versa*, $0 \leftrightarrow 0$, and $n \leftrightarrow m$). The mapping identifies an optimal time alignment of the two series. The task of finding it is set up as a dynamic programming problem by placing the time points of each series along the axes of a grid, representing alignments as paths through the grid cells, and finding the path with minimum accumulated weighted distance score (Figure 1b).

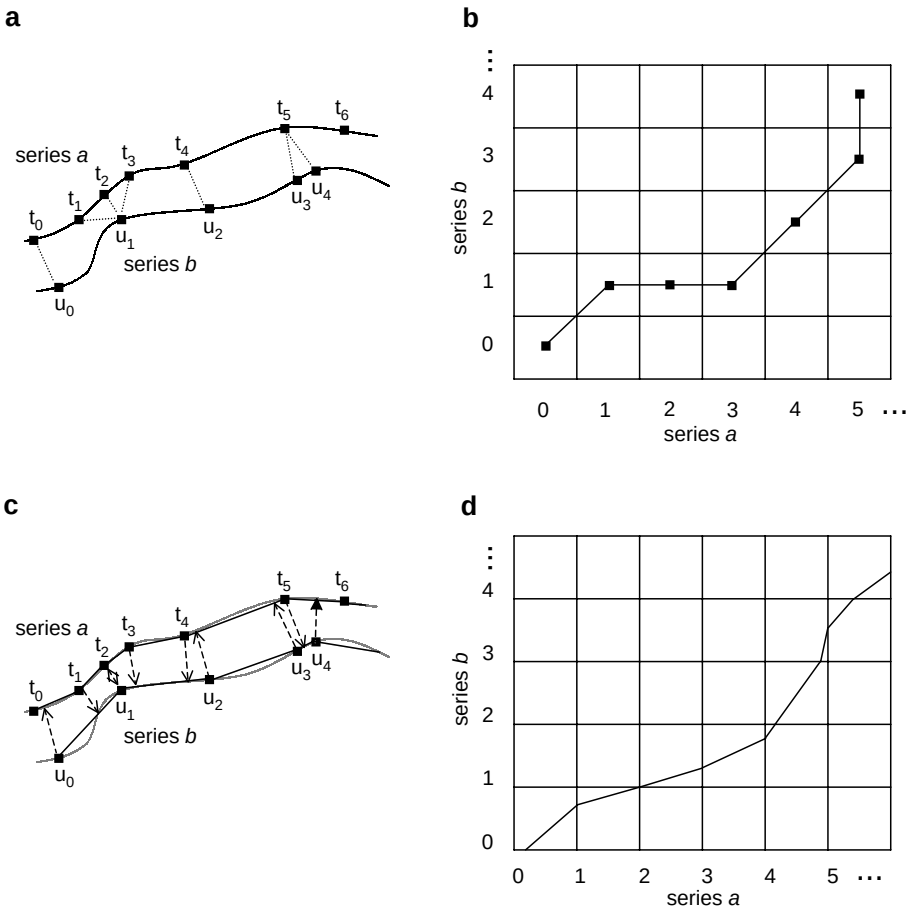


**Fig. 1.** Illustration of simple and interpolative time warping algorithms (see text and web site for details): (a) Two time series *a* and *b* in a two-dimensional feature space containing sample points from a continuous process, with sample points of each series mapped to each other by simple time warping. (b) Dynamic programming matrix for simple warping and the optimal path corresponding to (a). (c) Series *a* and *b* from (a) with sample points on each series mapped to interpolative points on the other by interpolative time warping algorithm. (d) Dynamic programming grid for interpolative warping and the optimal path corresponding to (c).

Horizontal or vertical segments of the optimal path identify places where multiple time points of one series correspond to a single time point of the other. Where measurement time intervals are comparable between the series, these may represent situations in which the instance of the biological process measured by one series moves quickly through a phase of the process relative to the instance measured by the other series. We call such situations compexps (compression/expansions) and they are analogous to the indels (insertion/deletions) considered in sequence alignment algorithms. Where the underlying processes sampled by time series are continuous or nearly so, compexps artificially represent process segments that span continuous intervals of time as jumping instantaneously at a point in time. Compexps may also result in artificially inflated alignment scores (Kruskal and Liberman, 1999). Use of interpolations between time points can address these issues. One option is to apply simple warping to time series which have been supplemented with interpolated values, an approach that has been used successfully in fitting time series to mathematical models (D'Haeseleer *et al.*, 1999). The genewarp and grphwarp programs can be used directly on such interpolated time series (see Figure 4c). In this method, compexps still appear in warps of interpolated time series but may represent smaller time intervals, and computed alignment scores will be based on comparisons of interpolated time points that do not represent actually measured values. For situations where these characteristics are undesirable, Kruskal and Liberman describe an interpolative algorithm that helps minimize them.

The key difference between the interpolative algorithm and simple time warping is that instead of finding a mapping between time points of one series and time points of

the other that minimizes the accumulated weighted distance between the corresponding k-space sample points, it finds a mapping between time points of the two series and linear interpolations between adjacent time points in the other series that minimizes the accumulated weighted distance between the corresponding k-space sample points and interpolated sample points of the other (see Figure 1c). We graph these alignments using a slightly modified version of the scheme used for simple time warping: Here time points are represented by the edges of the grid cells instead of their centers, segments of the alignment path by lines going from the lower or left edge of a grid cell to the upper or right edge, and interpolation fractions by the distances between path intersections with the lower or left edges of a grid cell and the lower left grid cell corners (see Figure 1d). Details are available on the web site along with a discussion of issues raised by this algorithm including time weight and interpolation definitions, reduced optimal paths, and interpolation time point order errors.

For both the simple and interpolative algorithms, execution time is $O(mnk)$ and memory requirement is $O(c_1mn + c_2k(m + n))$ for appropriate constants $c_1$ and $c_2$.

## RESULTS

In the following, please refer to Tables 1 and 2 for information about the time series (alpha, cdc15, and elu) and to Table 3 for information about sets of genes used in the alignment (MET, CLB2, pgt50, etc.). When referring to time series used as input to an alignment run, we use the notation *gene set:series*, e.g. pgt50:alpha refers to the time series defined by using only the genes from the pgt50 group from the alpha time series.

### Whole series alignments and alignment stability

While genewarp can be used on any set of genes regardless of whether their individual time course expression profiles are similar, we first applied it to small clusters of genes identified as having similar profiles (Spellman *et al.*, 1998) so that the operation of the algorithm could be visualized easily. Several examples are presented on our web site. While we noted that by eye the rises and falls of the clusters in each pair of time series were successfully aligned, we also noted that the path graphs of the alignments of different clusters and time series were quite variable. Focusing on the alpha and cdc15 time series, we note that these series began in different cell cycle phases and were sampled in such a way that each time point represents approximately the same fraction of the cell cycle for most of the two series (Table 2). If time warping were working perfectly we would thus expect to see an optimal path graph that starts with a vertical or horizontal segment representing a short compexp that compensates for the initial phase difference followed by a long segment with slope close to 1. Such a pattern is not consistently seen in the small cluster alignments. We suspected that statistical variation due to the small size of the gene clusters and the influence of measurement noise might be masking the expected diagonal path graphs. To explore the impact of measurement noise on the stability of alignments we took the pgt50 set of 990 genes (see Table 3) and performed the following procedure 250 times: (a) we randomly partitioned the 990 genes into two sets of 495, pgt50-rp1 and pgt50-rp2; (b) ran genewarp alignments on pgt50-rp1:alpha versus pgt50-rp1:cdc15, and on pgt50-rp2:alpha versus pgt50-rp2:cdc15; (c) computed the area between the optimal path graphs of these two alignments when overlaid on the same grid ($\Delta A$). pgt50 genes have low relative measurement noise because they exhibit differential expression by dint of having high variance across the time series, hence have variances that reflect actual signal as well as noise, and thus exhibit higher signal to noise ratios. By contrast, genes that are not differentially expressed will have $\log_2$ ratio expression levels consistently close to 0 and their variance will reflect mostly measurement noise. We followed this same procedure for two other sets of genes that differed from pgt50 in their sample size and variance characteristics and compared their average $\Delta A$ values ($\overline{\Delta A}$) to the pgt50 $\overline{\Delta A}$ (Table 4). The pgt50 $\overline{\Delta A}$ values are significantly lower than all the alternatives, indicating that larger sample size and lower measurement noise improve alignment stability. We repeated these results with interpolative alignment using genewarpi and obtained results supporting this same conclusion (Table 4). Because of this stability, we focus mainly on the pgt50:alpha versus pgt50:cdc15 alignment in further discussion below.

The genewarp alignment for pgt50:alpha versus pgt50:cdc15 is shown in Figure 2 (see below for a genewarpi alignment). Though the optimal warping path graph (Figure 2c) is based on the full 990 genes in pgt50, only the MET, CLB2, and CLN2 clusters are shown in the time series graphs (Figures 2a, b, and d). This path graph appears to be close to the expectations described above for a perfect alignment by presenting a short mainly horizontal segment at the beginning of the graph followed by a long segment with an overall slope of $\sim$1. A feature that does not match this expectation is the nearly vertical segment in the upper right of the path graph. This feature is common to many of the alignments we have generated from the three series and even appears in some of the small cluster alignments (see our web site). One possible explanation is simply that the cdc15 series contains $\sim$0.5 cell cycle periods more than the alpha series (Table 2) while the horizontal segment at the start of the path graph suggests that the alpha series has to run through some early time points to synchronize with the cdc15 series start. The long vertical segment at the end of the path

**Table 4.** Dependence of stability of time alignments on statistical characteristics of the input time series data sets. Identified sets of genes were randomly partitioned 250 times and, for each partitioning, the alpha and cdc15 series data for the partition were aligned using simple (genewarp) and interpolative (genewarpi) time warping. The area between the optimal warping paths for the two partition alignments was computed for each trial and the average used as a measure of alignment stability. Data sets with different statistical characteristics exhibit different levels of alignment stability, with larger sets of genes with low relative measurement noise yielding the most stable alignments

| Gene set[a] | Comparison with pgt50[b] | | $\overline{\Delta A}$[c] | $s_{\Delta A}$[d] | $P$-value[e] |
|---|---|---|---|---|---|
| | Size | Noise | | | |
| Simple warping (genewarp) | | | | | |
| pgt50 | = | = | 7.22 | 3.21 | |
| pgt33-990 | = | more | 9.42 | 5.88 | 3.03e−07 |
| pgt50-even | less | = | 9.83 | 2.36 | 6.71e−23 |
| pgt50-odd | less | = | 12.71 | 3.43 | 1.93e−58 |
| Interpolative warping (genewarpi) | | | | | |
| pgt50 | = | = | 3.46 | 1.85 | |
| pgt33-990 | = | more | 8.06 | 3.00 | 8.48e−69 |
| pgt50-even | less | = | 8.46 | 3.19 | 9.31e−73 |
| pgt50-odd | less | = | 6.50 | 3.32 | 6.65e−32 |

[a] Gene sets as described in Table 3.
[b] Comparison of gene set with pgt50. See Table 3 and text for details.
[c] Area between optimal warping paths resulting from genewarp or genewarpi alignment of each partition, averaged over the 250 random partitionings.
[d] Sample standard deviation of $\Delta A$.
[e] $P$-value of two-tailed $t$-test of equality of pgt50 and data set $\overline{\Delta A}$ values. The difference between pgt50 $\overline{\Delta A}$ scores and those of all other data sets is statistically significant for both simple and interpolative alignments. We have no explanation for the fact that the difference between pgt50-odd and pgt50-even $\overline{\Delta A}$ values is also statistically significant for both simple and interpolative alignments ($P = 4.23\text{e}{-25}$ for genewarp and $4.38\text{e}{-11}$ for genewarpi).

graph may simply represent excess time points at the end of the cdc15 series that do not correspond with alpha points and have no choice but to compexp with points at the end of the alpha series. Dephasing of the synchronized cell cultures towards the end of the time series might also contribute to the generation of the vertical segment. Dephasing would result in a flattening out of the time course of expression levels for each cell-cycle regulated gene (possibly seen in the MET and CLN2 clusters of Figure 2b), and the feature space distances between these sample points and any one sample point in the other series would be about the same. Thus, once time warping found a point $p$ in the pgt50:alpha series with minimal distance to the first dephased point in pgt50:cdc15, distances of the remaining dephased pgt50:cdc15 points to $p$ would also be small, encouraging a compexp with $p$.

## Half series time warps

The alpha series contains nearly two full periods of the cell cycle and the cdc15 series contains slightly more than two (Table 2). An additional test of time warping would be to see if it could map the first and second cycles of each of these two series. Again, if warping were perfect we should expect to see an optimal warping path be close to a line of slope 1. Meanwhile, the elu series contains less than a full cell cycle period (Table 2) so an alignment of its two halves should not yield a graph featuring a slope 1 segment. We performed genewarp alignments on all three pairs of half series and found that they all met their respective expectations. Details may be found on our web site.

## Statistics of optimal alignment scores

Our use of warping path slope as an indicator of the correctness of an alignment has only been possible because we are dealing with time courses that are designed to track the same processes, are for the most part evenly sampled, and whose time points cover approximately the same fraction of the process in each series. We cannot expect these criteria to be met as more and different kinds of time series are collected and compared; moreover, our assessments based on slope have been qualitative and a quantitative measure of alignment quality is highly desirable. In other uses of dynamic programming such as sequence and structure comparison, the alignment score has proved a powerful means of evaluating alignment quality when the probability of obtaining a given score or better can be estimated. A frequently employed method is to estimate the distribution of scores expected by random queries and to locate the score for a query of interest in this distribution (Karlin and Altschul, 1990; Levitt and Gerstein, 1998; Pearson, 1998). However, this cannot yet be done for time warping since it requires either a theory-based estimate of the distribution of possible scores for random queries or a large sample of scores from alignments within a database of comparable time series, neither of which is available.

As an alternative, we explored whether the quality of an alignment might be assessed statistically by randomly shuffling the vectors of gene expression levels associated with time points in each of the pgt50:alpha and pgt50:cdc15 time series individually and obtaining the genewarp alignment score for the shuffled series, repeating this procedure for 500 iterations. This is a variant of procedures originally used to evaluate Needleman–Wunsch and FASTA sequence alignments (Needleman and Wunsch, 1970; Pearson and Lipman, 1988). An advantage of this method is that it is not affected by many factors that might bias alignment scores in a general database context, including varying numbers of genes in common between query and database time series, numbers of time points in each series, and data normalization differences. We hoped that the actual score for the pgt50:alpha versus pgt50:cdc15 alignment (3850.62, see Figure 2c) would

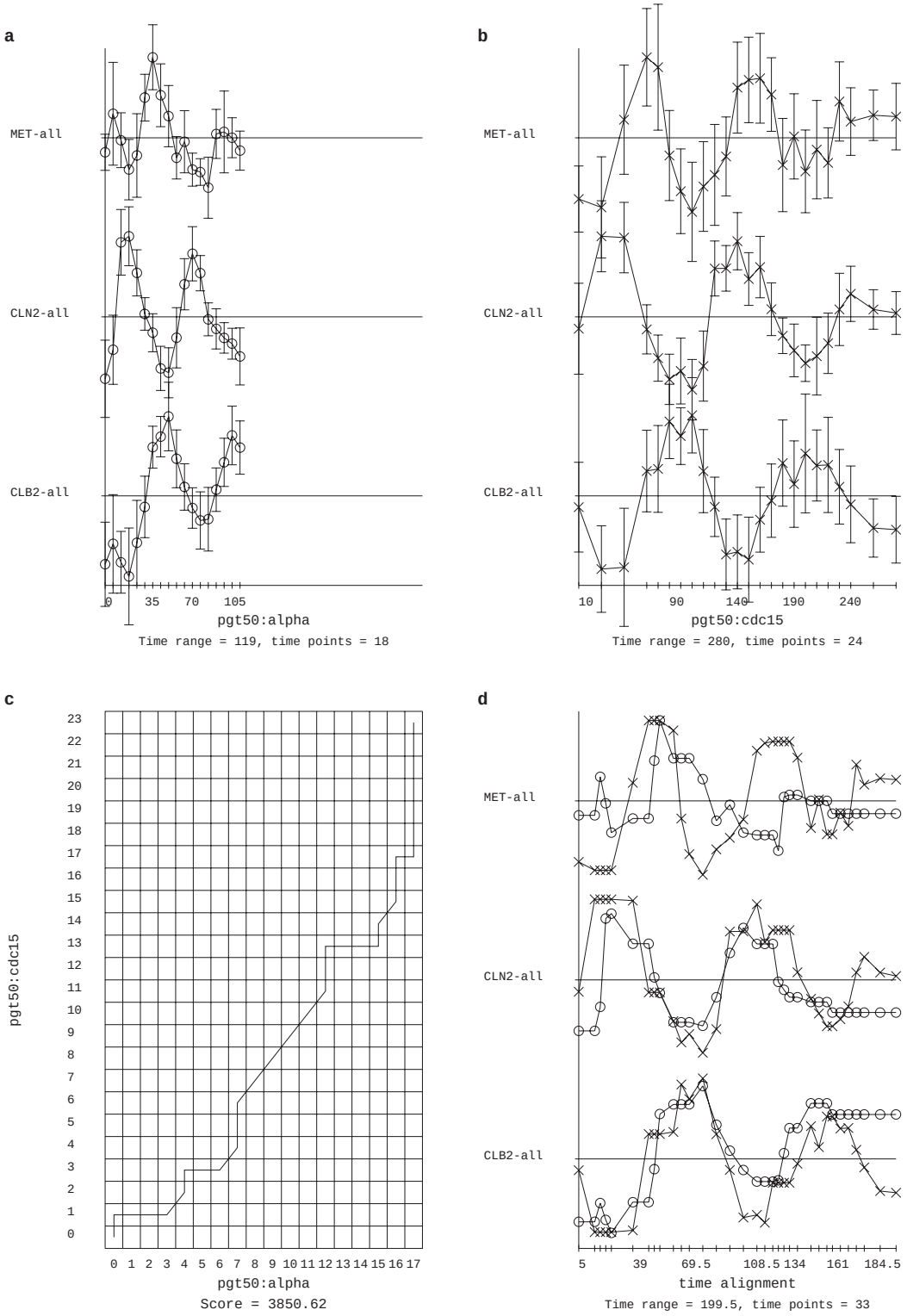


**Fig. 2.** Overall genewarp alignment of alpha and cdc15 series using large set pgt50 of differentially expressed genes. This alignment is more stable than those of smaller sets of genes or genes that are less differentially expressed (see text). (a) Average trajectories of MET, CLB2, and CLN2 clusters of genes in input pgt50:alpha data series. Error bars represent standard deviations at each time point. (b) Average trajectories with error bars of the same clusters in pgt50:cdc15 series. (c) Path graph and optimal path based on entire set of 990 genes in the input data. (d) Time alignment of the MET, CLB2, and CLN2 clusters based on (c).

be found at a low percentile of the shuffled series score distribution. However, this value proved to be at percentile 25.5 of the shuffled series score distribution. We conclude that use of shuffled series does not provide a good basis for evaluating scores statistically. The fact that the alpha and cdc15 are periodic so that more shuffles might score as well as the original than might arise in non-periodic series does not affect this conclusion, as such shuffles would be extremely improbable (calculations not shown). It is possible that the short lengths of time series (here 18 and 24 time points) compared to sequence lengths in typical sequence alignment situations make it harder to make use of score statistics in time warping.

## Comparison with clustering

To compare time warping with clustering, we clustered the combined set of alpha and cdc15 series time points over the expression levels of the pgt50 gene set with Ward's algorithm (Everitt, 1980), a hierarchical clustering algorithm that minimizes variance when combining subclusters, using SPLUS 2000 (MathSoft, Seattle, WA). The results are in Figure 3. As clustering groups time points together regardless of what series they come from and does not use the order information inherent in the time values themselves, Figure 3 contains both groupings of time points from individual series (e.g. figure grouping a) and groupings that contain mixtures from both series (grouping b), as well as groupings of nearby (c) and remote (d) time points. The many examples of leaf-level clusters containing adjacent time points of a series (e.g. grouping c) are likely due to the fact that RNA expression levels change slowly from one time point to the next; hence, the feature vectors associated with each time point are likely to be closer in feature space to those of adjacent time points than any others. Feature vectors of time points a period apart should also be close in feature space and thus tend to group together, and this may be behind grouping (d) which contains two alpha series time points 56 min apart, not far from the 67.5 min computed period (Table 2). From the point of view of time point mapping, time warping's ability to keep the two series apart and focus on the correspondences between them is a clear advantage over clustering. Both clustering and time warping have a tendency to treat adjacent time points similarly, resulting in compexps in time warping and the clusters of adjacent time points just noted. But clustering's ability to map temporal correspondences may be blurred by its inability to distinguish feature vectors in a series that are close due to similar cell states from those close due to time point adjacency, compared to time warping which can distinguish them.

To see how the clustering may reflect these influences, we labeled the alpha time points in Figure 3 according to how well the hierarchically closest cdc15 time point(s)

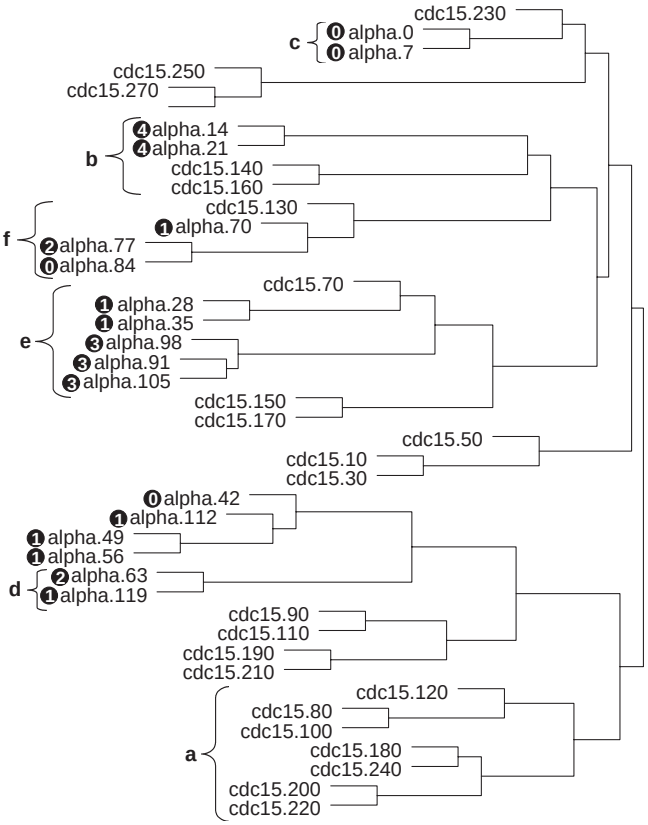


**Fig. 3.** Results of clustering of alpha and cdc15 time points based on the pgt50 set of genes and comparison of time point clusters with time warping alignments of pgt50:alpha versus pgt50:cdc15 (see Figure 2) and the alignments of pgt50:alpha and pgt50:cdc15 half series (see text and web supplement). Groupings a–f and time point labels 0–4 are described in the text.

to an alpha point (cluster matches) corresponds with the cdc15 time point(s) mapped to it by the time warping of Figure 2 (warp matches). Alpha points labeled with 1 have a cluster match that exactly matches a warp match. Alpha points labeled with 2 have a cluster match that is adjacent in the cdc15 time series to a warp match. Alpha points labeled with 3 have a cluster match that is a cdc15 period away from a warp match, where cdc15 points that correspond across a period are identified by the cdc15 half series time warping described above and on our web site. Alpha points labeled with a 4 have a cluster match adjacent to a cdc15 point a period apart from a warp match. Alpha points that do not meet any of these conditions are labeled with 0. From these assignments, possible origins of larger groupings may be inferred. For instance, grouping e appears to have begun from clusterings of the alpha 91, 98, and 105 min points, and then the alpha 28 and 35 min points, by adjacency. The alpha 28 and 35 pair then grouped with a cdc15 70 min point that represents the

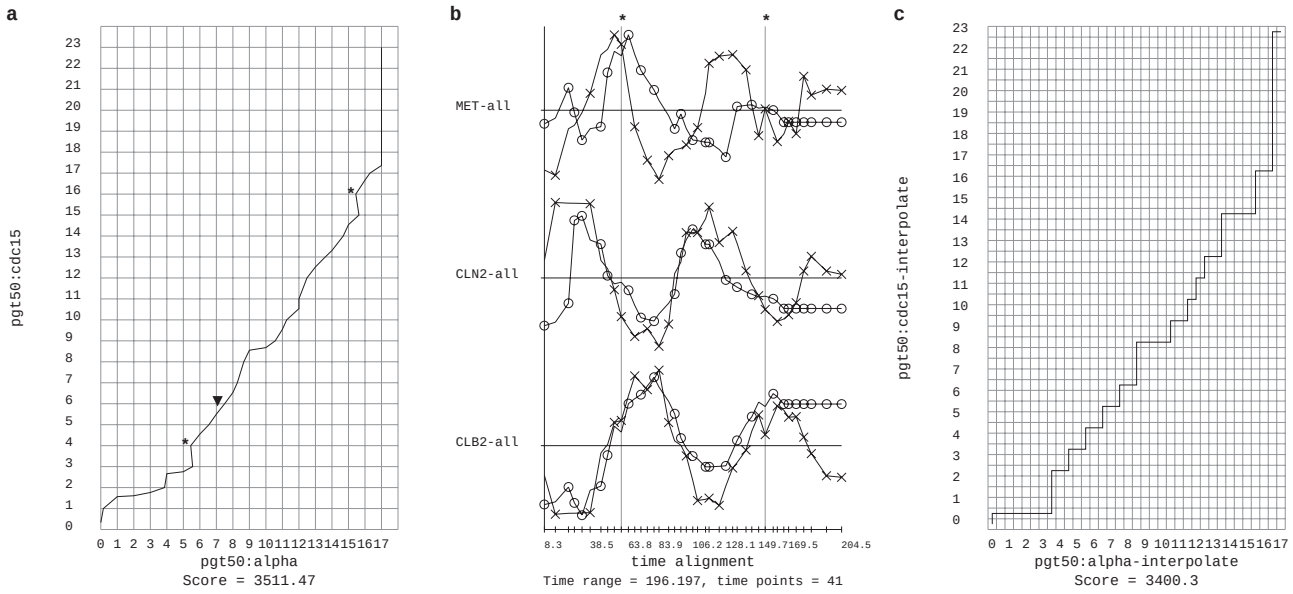


**Fig. 4.** Interpolative alignments of pgt50:alpha versus pgt50:cdc15. (a) Interpolative alignment of pgt50:alpha versus pgt50:cdc15 time series as depicted by grphwarpi-generated path graph. Input time series graphics from grphwarpi are identical to Figures 2a and b. Two interpolation time point order errors (see text) are indicated by *. Arrowhead: time point of alpha series which has been repeated with Gaussian noise to mimic arrest (see text and Figure 5). (b) grphwarpi-generated time alignment graphic of MET, CLB2, and CLN2 clusters as determined by (a). The effect of the order errors in (a) is seen in the sawtooth deviations of the alpha time series (circular ticmarks) at the vertical lines marked by *. (c) Non-interpolative genewarp alignment of pgt50:alpha versus pgt50:cdc15 where these time series were first modified by interpolation of additional time points halfway between each pair of consecutive originally provided time points, as depicted by grphwarp-generated path graph. Time point numbers in the graphic have been modified to be consistent with the path graph of the original uninterpolated pgt50:alpha versus pgt50:cdc15 series shown in Figure 2c.

same cell cycle state, and the alpha 91, 98, and 105 min grouping then joined in by virtue of being a period away from the cdc15 (and also the alpha series) perspective. Similarly, grouping f apparently consists of three alpha points grouped by adjacency, which are then joined to a cdc15 point that exactly corresponds to one of them.

## Interpolative alignment

We used genewarpi to align the pgt50:alpha and pgt50:cdc15 series and the grphwarpi-generated path graph and time alignment is shown in Figures 4a and b. The input time series graphs are identical with those in Figures 2a and b. The interpolative path graph remains close to the expectations for a perfect alignment in having a long segment of slope near 1. Deviations from the slope 1 line appear by eye less pronounced than the non-interpolative alignment graph in Figure 2c. Interpolative time warping can sometimes generate alignments where interpolated time points are out of proper time sequence (interpolation time point order errors). Two interpolation time point order errors are marked in Figure 4a with asterisks. They appear as path graph segments with negative slopes where the second of two consecutive interpolations in the alpha series has an earlier time value than the first. These order errors also appear at the asterisked vertical lines as small sawtooths in the alpha series graph lines (circular tic marks) in the overall time alignment in Figure 4b. Because of the order errors, the alignment score of 3511.47 must be smaller than the score of the truly optimal path. (See our web site for explanations of this and the related issues below.) The alignment was rerun with the enforceorder option, which suppresses order errors at the cost of returning a possibly suboptimal alignment. When running with enforceorder the appearance of a similar graph to the original cannot be guaranteed; however, in this case, the enforceorder graph is identical except for correction of the order errors. Because the score with enforceorder is 3512.49 (see web site), the score of the truly optimal path must be in the small range between 3511.47 and 3512.49. As expected, the optimal alignment score is less than the score of 3850.62 returned by simple non-interpolative alignment of the same series (Figure 2c). For purposes of comparison, we also performed a simple genewarp alignment on the same pgt50:alpha and pgt50:cdc15 time series but where these time series were first modified by addition

of interpolated time points halfway between each pair of consecutive time points in the original series (Figure 4c). This, too, results in an alignment score lower than the original simple alignment of the uninterpolated time series (3400.3 versus 3850.62) and presents an optimal path that tracks closely to a diagonal of slope 1. However, as noted above (see Section Algorithm and implementation), compexps are not eliminated in this alignment. Indeed, in this particular interpolative alignment, the path graph is composed entirely of compexps and presents no diagonal segments at all.

### Generality and robustness of alignments

We wished to confirm that the algorithms produce expected results in situations where we would not expect the time series to align with a nearly diagonal path graph slope, and where the time points of the two series do not represent nearly equal fractions of a biological process. We therefore created 100 instances of a time series pgt50:alpha.tp7+4rand that mimicked what pgt50:alpha would be like if the cells arrested at time point 7 (49 min) for 28 min, by repeating four copies of time point 7 after the original at the usual 7 min alpha series time intervals with admixtures of Gaussian noise. We chose standard deviations for the noise for each gene by taking the time point 7 expression level value and the expression levels for the four other time points whose values were the closest to the time point 7 value, reasoning that this choice mimics an arrest at time point 7 better than considering the time point 5, 6, 8 and 9 values (which could reflect possibly large motions through the time point 7 value for some genes). We chose time point 7 (which is in S/G2 according to Alter *et al.* (2000)) because it is in the midst of a series of time points that appear in a strongly diagonal segment of the genewarpi path graph with pgt50:cdc15 (Figure 4a, arrowhead). If time warping is working properly, alignments of these modified series against pgt50:cdc15 should generate path graphs in which the original diagonal is interrupted by a horizontal compexp for four time points. Figure 5 presents a superimposition of all 100 path graphs. The 100 path graphs are all so similar that even as a group they are visually indistinguishable from a single graph containing a perfectly horizontal line at the repeated time points. We also computed the areas between each of these path graphs against one containing a perfectly horizontal line at the repeated points, and found $\Delta A = 0.0349 \pm 0.0116$ (mean $\pm$ SD). These results suggest that time warping will work properly in situations in which time points in two time series represent unequal portions of a biological process, and where the process unfolds at varying rates. They also suggest that time warping can map series that would be difficult to map by simpler means such as linear fits between time axes.

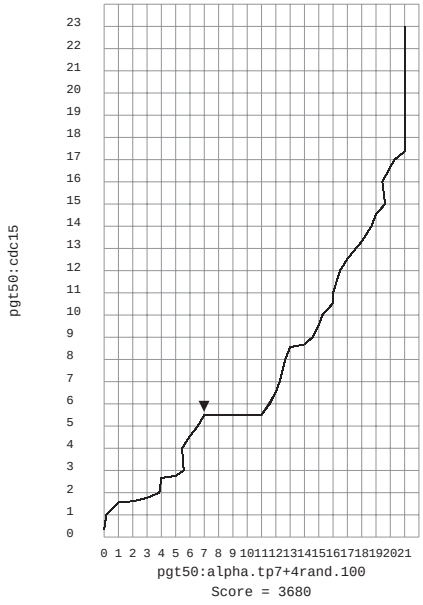


**Fig. 5.** Superimposition of 100 path graphs generated from genewarpi alignment of modified version of pgt50:alpha against pgt50:cdc15 where four copies of pgt50:alpha time point 7 are inserted into the pgt50:alpha series with admixtures of Gaussian noise, mimicking a 28 min arrest in the alpha series at the state achieved at time point 7. Time point 7 in the alpha series is in the midst of a strongly diagonal section of the path graph of the genewarpi alignment with pgt50:cdc15 (Figure 4a, arrowhead). The result of all 100 alignments of pgt50:cdc15 with noise modified insertions into pgt50:alpha is insertion of a four time point segment visually indistinguishable from a horizontal line (arrowhead, above) that correctly represents the mapping of pgt50:cdc15 against the arrest simulated in modified pgt50:alpha series.

## DISCUSSION

The programs described here provide a basic capability to align RNA or protein expression time series but can be used in other applications. One possibility is to use them to align composite time series that contain phenotype parameters in addition to expression levels that are measured on the same time course. Depending on the domain of application, these might include cell-specific parameters such as average cell size or physiological parameters such as blood pressure or temperature. The relative contributions of such parameters to alignment score calculations can be adjusted using feature weight parameters already supported by the programs. The alignment programs can also be used not only to align RNA and protein expression series individually, but series that combine both RNA and protein data. Finally, the programs can also be applied to aligning non-temporal series such as expression profiles for cells over a range of concentrations of compounds (concentration warping). These would be of interest whenever cells have a common dose response trajectory at the

RNA or protein expression level but move through it at different rates relative to changes in compound concentration when the cells are from different strains or culture conditions.

In addition, many modifications and enhancements to the algorithms are possible, including support for (a) time series containing genes with incomplete sets of measurement values (see equation (5) in the Algorithm and implementation section of the web supplement) (b) gaps (Kruskal and Liberman, 1999), (c) local time series alignments (cf. Smith and Waterman, 1981), (d) multiple time series alignments, and (e) incorporation of constraints that would bar alignments between time series segments covering widely unequal time intervals (Sakoe and Chiba, 1978) (however, this cannot be done using path slope constraints which implicitly presume even sampling across the time series). Other directions for development would include (f) development of hidden Markov models (Durbin *et al.*, 1998) for time alignment of expression data, and (g) support for what we call causality searches: Here the time series expression profile of a gene or gene set is warped against that of another gene or gene set in the same time series, and the object is to find cases where the second genes follow similar trajectories to the first but where they are generally delayed in the series relative to the first, identifying the first genes as candidate regulators of the second.

Development of statistics for time alignment scores is an area of opportunity for improving time warping. A possible direction would be to base alignment scores on functions of Euclidean distance rather than directly on these distances. In an analogous situation, structure alignment score distributions have been successfully fit to extreme value distributions when algorithms maximize similarity scores based on $S_{ij} = M/(1 - (d_{ij}/d_0)^2)$ where $d_{ij}$ are Euclidean distances and $M$ and $d_0$ are suitable constants, rather than minimizing weighted sums of $d_{ij}$ directly as done here (Levitt and Gerstein, 1998). A more general issue will involve development of appropriately normalized forms of scores that take into account variable numbers of common genes and different time series lengths that will enable comparisons of scores of a query time series against a database of other time series of similarly normalized expression data that differ in these properties. As with clustering (Aach *et al.*, 2000), time alignments of time series of expression level data gathered by different methodologies or subjected to different normalizations will likely be affected by biases and artifacts and should be interpreted with caution.

As an alignment algorithm, time warping is a different kind of tool than other kinds of analysis recently applied to expression time series. Though it has points of contact with Fourier analysis (e.g. as used in Spellman *et al.*, 1998) and SVD (Alter *et al.*, 2000; Holter *et al.*, 2000; Raychaudhuri *et al.*, 2000), it is not a method of decomposing signals into a weighted sum of basis functions or vectors and simplifying them by focusing on principal terms. Though these techniques hold much promise, they are not principally alignment tools and we foresee the need for research into the techniques and conditions of applying them to alignment. For instance, we would be interested in seeing how Fourier analysis might be used to align series that contain varying relative rate differences or arrests (cf. Figure 5), and in understanding how densely and evenly sampled time series must be to apply Fourier analysis effectively in this way. Like clustering, SVD does not make use of the information provided by the measurement times of the time points it analyzes and thus might require special techniques to avoid confusing temporal proximity with non-temporal similarity of cell state in aligning series. Though we do not doubt that these tools will eventually offer powerful methods of time series alignment in some applications, time warping is available now as a simple and general alignment tool.

While time warping maps two time series in a way that compensates for varying relative rate differences in gene expression levels moving along similar expression trajectories, it does this by mapping the time points themselves and therefore operates as if all the genes in a cell move relatively in lock step. But we suspect that biology will present cases where some pathways or gene groups in a cell move quickly through an expression trajectory while, at the same time, others are moving slowly, relative to another time series of the same process. We call such situations time course superpositions. Their detection raises an interesting scientific problem because pathways and gene groups whose time courses are superposed may contain only small numbers of genes and their apparently distinct time course relative to other genes might be indistinguishable from the statistical instability we found characteristic of small gene group alignments (see Results). For instance, the MET gene cluster appears to align less well than the CLB2 and CLN2 series in the middle segment of the time alignment of Figure 2d, but we cannot conclude that it is on a separate time course because this cluster contains only 11 genes, too small to assure a stable alignment. If only a few large sets of genes with superposed time courses exist, the sets might be found by clustering the individual time series and their variant time courses characterized by time warping the clusters. Otherwise, repeated experiments or demonstration of independent and unsynchronized regulation would be required. Superpositions may be interesting to consider in connection with cell checkpoints, for the latter may represent situations where different pathways operate under independent regulation and need to be synchronized by a supervening process.

Gene expression time series analysis will become increasingly important as researchers apply high throughput assays for gene expression to the understanding of biological processes as they unfold over time. We have adapted time warping algorithms to gene expression data and demonstrated them on published yeast gene expression time series, showing that they are capable of mapping corresponding cell states across these series. By using time series information available to but not used by clustering, we have also shown that time warping is superior to clustering in this task. We believe that time warping, with the suitable enhancements indicated above, will soon join clustering and signal analysis as a key bioinformatics tool.

## ACKNOWLEDGEMENTS

We very much thank the following people for help: Paul Spellman for information concerning his cell cycle experiments and results, Yizong Cheng for helpful discussions about algorithm options and issues, and Mark Liberman for helpful discussions on the application of signal processing techniques to gene expression data, and three anonymous referees for critical comments on this manuscript. We also thank the Lipper Foundation, DOE grant DE-FG02-87ER60565, and NIHLB PGA for their funding of this work.

## REFERENCES

Aach,J., Rindone,W. and Church,G.M. (2000) Systematic management and analysis of yeast gene expression data. *Genome Res.*, **10**, 431–445.

Adobe Systems Incorporated. (1999) *PostScript Language Reference*. Addison-Wesley, Reading, MA.

Alizadeh,A.A., Eisen,M.B. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–511.

Alter,O., Brown,P.O. and Botstein,D. (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl Acad. Sci. USA*, **97**, 10 101–10 106.

Bittner,M., Meltzer,P. *et al.* (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*, **406**, 536–540.

Brenner,S., Johnson,M. *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, **18**, 630–634.

Cho,R.J., Campbell,M.J. *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **2**, 65–73.

Chu,S., DeRisi,J. *et al.* (1998) The transcriptional program of sporulation in budding yeast. *Science*, **282**, 699–705.

DeRisi,J.L., Iyer,V.R. and Brown,P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.

D'Haeseleer,P., Wen,X., Fuhrman,S. and Somogyi,R. (1999) Linear modeling of mRNA expression levels during CNS development and injury. *Pac. Symp. Biocomput.*, 41–52.

Durbin,R., Eddy,S., Krogh,A. and Mitchison,G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge.

Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14 863–14 868.

Everitt,B. (1980) *Cluster Analysis*. Halsted Press, New York.

Golub,T.R., Slonim,D.K. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.

Gygi,S.P., Corthals,G.L., Zhang,Y., Rochon,Y. and Aebersold,R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl Acad. Sci. USA*, **97**, 9390–9395.

Gygi,S.P., Rist,B., Gerber,S.A., Turecek,F., Gelb,M.H. and Aebersold,R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.*, **17**, 994–999.

Holter,N.S., Mitra,M., Maritan,A., Cieplak,M., Banavar,J.R. and Fedoroff,N.V. (2000) Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc. Natl Acad. Sci. USA*, **97**, 8409–8414.

Karlin,S. and Altschul,S.F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl Acad. Sci. USA*, **87**, 2264–2268.

Kruskal,J.B. and Liberman,M. (1999) The symmetric time-warping problem: from continuous to discrete. In Sankoff,D. and Kruskal,J. (eds), *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison.* CSLI Publications, Stanford, pp. 125–161.

Levitt,M. and Gerstein,M. (1998) A unified statistical framework for sequence comparison and structure comparison. *Proc. Natl Acad. Sci. USA*, **95**, 5913–5920.

Lockhart,D.J., Dong,H. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.

Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.

Pearson,W.R. (1998) Empirical statistical estimates for sequence similarity searches. *J. Mol. Biol.*, **276**, 71–84.

Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.

Raychaudhuri,S., Stuart,J.M. and Altman,R.B. (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. In *Pacific Symposium for Biocomputing.* Oahu.

Sakoe,H. and Chiba,S. (1978) Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP **26**, 43–49.

Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.

Spellman,P.T., Sherlock,G. *et al.* (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.

Tamayo,P., Slonim,D. *et al.* (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application

to hematopoietic differentiation. *Proc. Natl Acad. Sci. USA*, **96**, 2907–2912.

Tavazoie,S., Hughes,J.D., Campbell,M.J., Cho,R.J. and Church,G.M. (1999) Systematic determination of genetic network architecture. *Nat. Genet.*, **22**, 281–285.

Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484–487.

Velichko,V.M. and Zagoruyko,N.G. (1970) Automatic recognition of 200 words. *Int. J. Man-Mach. Stud.*, **2**, 223.

White,K.P., Rifkin,S.A., Horban,P. and Hogness,D.S. (1999) Microarray analysis of *Drosophila* development during metamorphosis. *Science*, **286**, 2179–2184.