BIOINFORMATICS

Vol. 18 no. 9 2002
Pages 1176–1183

Geometry of gene expression dynamics

S. A. Rifkin [1] and J. Kim [1, 2, 3,*]

[1]Department of Ecology and Evolutionary Biology, PO Box 208106, [2]Department of Molecular, Cellular, and Developmental Biology, PO Box 208103, and [3]Department of Statistics, Yale University, PO Box 208290, CT 06520, New Haven, USA

Received on August 30, 2001; revised on March 3, 2002; accepted on March 25, 2002

ABSTRACT

Motivation: A gene expression trajectory moves through a high dimensional space where each axis represents the mRNA abundance of a different gene. Genome wide gene expression has a dynamic structure, especially in studies of development and temporal response. Both visualization and analyses of such data require an explicit attention to the temporal structure.

Results: Using three cell cycle trajectories from *Saccharomyces cerevisiae* to illustrate, we present several techniques which reveal the geometry of the data. We import phase-delay time plots from chaotic systems theory as a dynamic data visualization device and show how these plots capture important aspects of the trajectories. We construct an objective function to find an optimal two-dimensional projection of the cell cycle, demonstrate that the system returns to this plane after three different initial perturbations, and explore the conditions under which this geometric approach outperforms standard approaches such as singular value decomposition and Fourier analysis. Finally, we show how a geometric analysis can isolate distinct parts of the trajectories, in this case the initial perturbation versus the cell cycle.

Contact: junhyong.kim@yale.edu

INTRODUCTION

During physiological and developmental processes of an organism, the molecular state of any given cell undergoes a cascade of changes in coordination with other cells and the environment. These molecular interactions have an inherent temporal structure—they obey dynamical rules and can be represented as trajectories through a state space. Given recent advances such as microarray technology (Schena *et al.*, 1995), a reasonable state space for these trajectories is a high-dimensional gene expression space—where each dimension represents the mRNA abundance of a different gene. However, microarray data is characterized by a relatively high degree of noise, an extremely large number of variables, and a small number of measurements, exactly the opposite of typical time-series data. Parametric approaches such as Fourier analysis may be too constraining, hiding crucial aspects of the data, while an analysis of the static structure of the data such as singular value decomposition (SVD) will not reveal most dynamical features. In this paper, we introduce two dynamical analysis techniques for gene expression data: dynamical structure visualization and geometric analysis of quasi-periodic dynamics. We apply our analyses to the well-analyzed Saccharomyces cerevisiae cell cycle data as an example and demonstrate the strength of our method using numerical simulations.

Trajectories through a state space can be (1) aperiodic or strictly divergent, where the trajectory does not return to the same state even in projections; (2) quasi-periodic, where the trajectory is characterized by a few dominant frequencies; (3) chaotic (where the trajectory does not repeat, but is confined); or (4) random, where there are lots of dominant frequencies. Biological dynamics are unlikely to be random so, while randomness is important as a null hypothesis, we do not focus on such trajectories. Understanding chaotic dynamics requires long time series and lots of data that for both biological and economic reasons are prohibitively difficult for microarray analysis. In this paper, we focus on quasi-periodic trajectories, and we demonstrate how to extract periodic components from such trajectories. It may seem that quasi-periodic trajectories will be rare except in cases like the cell-cycle. However, any gene that is regulated with at least one on-off cycle will show a periodic signature. Thus at the whole-genome level, we expect any sufficiently dense time series to display periodic components unless attention is restricted to a strictly divergent subset of genes.

SYSTEM

We used Mathematica 4.0 (Wolfram, 1999) to analyze the three yeast cell-cycle datasets—alpha-factor, cdc15, and cdc28—with adjustments for missing data as described in Rifkin *et al.* (2000) providing us with data for 5541 genes (see Cho *et al.*, 1998 and Spellman *et al.*, 1998, for details

*To whom correspondence should be addressed.

of the experiments). The cdc15 time series has samples every 10 minutes from 10 to 290, except for 5 timepoints. These we estimated by linear interpolation. The different datasets arise from three different ways in which the yeast cell cycle was arrested and synchronized prior to measurement of gene expression over time. Dynamically, they represent three different initial conditions for the system. Given the periodic nature of the cell cycle, it is evident that the activity of some subset of the genes or linear combinations of gene expression levels will show simple periodic dynamics with the frequency of the cell cycle. If the dominant expression dynamics of these genes is governed by a low-dimensional periodic component, it will lie in some unknown subspace of the high-dimensional state space of all the genes (6200 genes for yeast). We investigated the geometry of these dynamics in three stages: (a) locating and visualizing the possible periodic dynamics in the state space; (b) projecting it onto subspaces to estimate average trajectories with particular properties; and (c) statistically characterizing the extent of the periodic dynamics in the genome. We also demonstrate how previous results (Rifkin *et al.*, 2000) arise from the geometry of cell cycle dynamics.

IMPLEMENTATION

Visualizing low-dimensional dynamical structure

Visualizing the data structure is a useful starting point for analyzing time-series data, however, the high dimensionality of expression array data causes a severe problem. With thousands of variables, individual time plots fail to convey the global dynamical structure. Standard ordination techniques such as Principal Component Analysis yield a low-dimensional summary of the static variation rather than the dynamical variation. We import a technique from nonlinear dynamical systems analysis to obtain a low dimensional picture of gene expression data that captures its dynamical features, although the short time-series prevents us from taking full advantage of the technique.

Attractor reconstruction (Packard *et al.*, 1980; Takens, 1981) characterizes the dynamics of a large number of hidden variables (i.e. the full phase space) from a measured single variable if the variables are sufficiently coupled. Under such conditions, measurements of the single variable taken at two or more quasi-independent time points provide qualitative information about the dynamics of all the variables. Microarrays make measurements obtainable for most of the variables, i.e. most of the gene expression levels. (Of course, this does not take into account the effect of the proteome or cellular conditions on these levels.) Since the variables are too many to analyze efficiently, we projected the trajectory onto a single vector direction in the full gene expression space and then used the attractor reconstruction technique to visualize the dynamics.

Given some vector direction $\boldsymbol{v}$, let $v_p(t) = \boldsymbol{v} \cdot \boldsymbol{x}(t)$ be the projection of the vector-valued time-series $\boldsymbol{x}(t)$ onto $\boldsymbol{v}$. Then $v_p(t)$ is a one-dimensional time series that will capture the dynamics of the low-dimensional periodic signal if it intersects with the subspace of this component. As mentioned, in a coupled system the qualitative dynamics of an attractor in the full phase space can be reconstructed from a time-delayed phase plot of a single variable. These time-delayed variables are of the form $\boldsymbol{v}_p(t + nd)$, $(n = 0, 1, \ldots, k - 1)$ where d is the time and k is the embedding dimension of the attractor (in all of our analyses $d = 1$ and $k = 3$). The time-series captured by $v_p(t)$ can therefore be used to characterize the dynamics of genomic expression and to geometrically search for the subspace of periodic dynamics. The vector, v, serves as a probe into the state space which, when immersed into the low-dimensional subspace of the periodic dynamics, displays a characteristic signal. We used SVD to find orthonormal vector directions of increasing static variation as candidates for the probe vector v (Rifkin *et al.*, 2000; Holter *et al.*, 2000). The number of candidate probe directions was sample size minus one for each data set. We then selected the candidate direction that yielded the smoothest non-intersecting trajectories. Standard attractor reconstruction techniques attempt to find the best embedding dimension of the data through iterative analysis (Abarbanel *et al.*, 1993). The yeast cell cycle time series is much too short for such an analysis. Consequently, for these data, the time-delayed phase plot is best considered a data visualization device. The data are projected onto a low-dimensional picture that best summarizes the global temporal structure of the high-dimensional time series.

Figure 1 shows the periodic dynamics of the yeast cell cycle in these time-delayed phase plots. In Figure 1(a), the cell begins in a perturbed state from exposure to alpha-factor pheromone and relaxes back to an oscillatory trajectory with a period coincident with the cell cycle. Figure 1(b) and (c) show similar visualizations for the cdc15 and cdc28 experiments. In contrast, Figure 1(d) depicts the reconstruction for a trajectory from a randomized time series. Although the short time series does not permit exacting assessment of the embedding dimension, under all three experimental conditions, the periodic dynamics are apparent in the three dimensional reconstruction and appear to settle to a plane after an initial perturbation. Within the resolution of these datasets, the dynamics can be characterized by a two-dimensional plane of oscillation with the periodicity of a single cell cycle (but there may be other significant components to the dynamics, see below).

Geometrical decomposition of dominant periodic structure

Given dynamical trajectories with such periodic components, an especially important statistic is the phase of each

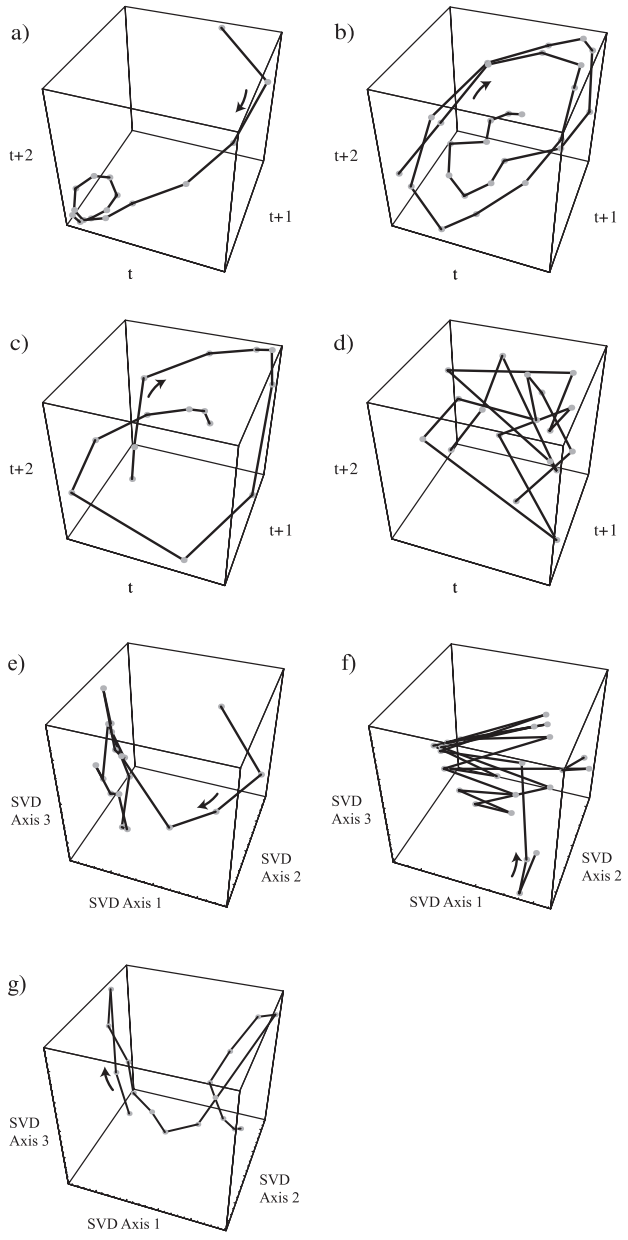

Fig. 1. Reconstructed trajectories from projections onto single SVD axes: (a) Alpha factor, SVD axis 1; (b) Cdc15, SVD axis 4; (c) Cdc28, SVD axis 2; (d) Random 20 timepoint trajectory. For comparison, the trajectories are plotted in the subspaces spanned by the first three SVD axes for each: (e) Alpha factor; (f) Cdc15; (g) Cdc28.

gene with respect to the dominant frequencies. An accurate assignment of a gene's phase will reveal groups of genes that turn on at the same time and possibly regulatory dependencies between genes of different phase groups. For data from two or more different genomes or different environmental conditions, the phase relationship of the genes can be used to assess whether individual genes or some subset of the genome is acting in a temporally concordant manner. Fourier analysis is a standard tool for analyzing periodic dynamics; estimation of the Fourier coefficients yields the dominant amplitude frequencies and corresponding phase of the individual genes. However, Fourier analysis can be problematic for short and noisy time series, especially if the dynamics contains a non-periodic component (see below). To overcome its limitations, we developed a non-parametric technique based solely on the geometry of the data.

We can think of sinusoidal periodic dynamics at a particular frequency along a vector direction as a projection (i.e. weighted sum) from a two-dimensional circular trajectory. This is the geometrical interpretation of Fourier decomposition of periodic trajectories. Conversely, if we have a periodic dynamic in high dimensions at a particular frequency, it will 'inscribe' a circular trajectory when projected onto some two dimensional plane. Each frequency can be associated with a different two-dimensional plane. The angular positions of the plane with respect to the original basis vectors—here representing mRNA abundances of the different genes—define the phase groups of the genes with respect to the oscillation at that particular frequency. Once such plane is obtained, we can assign phases to individual genes as well as use the relationship of planar subspaces to determine relatedness of global periodic gene expression patterns between different experimental conditions. Looking for subspaces (in this case planes) in which projections of multidimensional clouds of points or trajectories have particular features (in this case circularity) is an important technique of exploratory data analysis (Friedman and Tukey, 1974). The key lies in good characterizations of these interesting features.

For each time-series, we first selected an initial two-dimensional plane with a pair of orthonormal vectors. Next, we projected the original time-series, $x(t)$, onto this plane. The projected trajectory was normalized against a regular polygonal trajectory constructed by taking the mean distance from the centroid to the datapoints and the mean angle between the centroid and two successive datapoints (see Figure 2(a) for an example). An ideal single-period trajectory with a sinusoidal form would circumscribe a circle on a two-dimensional plane. These normalizing trajectories are approximations to this ideal. We used the following objective function to indicate departures from these polygonal trajectories:

$$f = \frac{\sum_{i=0}^{n=1} |v_{\text{proj}}(t+1) - v_{\text{proj}}(t)|}{\sum_{t=0}^{n=1} A_{\text{proj}}(t+1,t)} \Bigg/ \frac{\sum_{t=0}^{n=1} |v_{\text{poly}}(t+1) - v_{\text{poly}}(t)|}{\sum_{t=0}^{n=1} A_{\text{poly}}(t+1,t)} \quad (1)$$

where v_{proj} is the normalized time-series projected onto the candidate two-dimensional plane, $A_{\text{proj}}(t+1, t)$ is the signed area of the triangle formed by $v_{\text{proj}}(t+1)$, $v_{\text{proj}}(t)$, and the centroid, and v_{poly} and A_{poly} are the equivalent quantities for the regular polygonal trajectory. The f statistic measures the ratio between the length of the projected trajectory and the area swept by the trajectory relative to the polygonal standard. We used numerical optimization routines to search through the fifteen dimensional space of significant static variation identified by the SVD analysis (Rifkin *et al.*, 2000). We call the resulting two-dimensional planes the dominant frequency planes. The trajectories projected onto the dominant frequency planes for each of the three experimental conditions are shown in Figures 2(a)–(c).

Since we are searching for the most circular two-dimensional characterization of the time-series, our procedure is a non-parametric decomposition for the dominant frequency. The trajectories on the dominant frequency planes can be projected back to individual gene axes effectively extracting the signal component given by the dominant frequency sinusoidal wave. As transcription proceeds, different sets of genes will turn on, shifting phases of gene expression (Spellman *et al.*, 1998). Figures 2(d)–(f) show the trajectories for pairs of genes with different phases. The top plot in each figure shows the raw trajectories for both genes, while the bottom depicts only the dominant frequency components.

Projecting the gene basis axes onto the dominant frequency plane complements the decomposition of a particular gene's trajectory into dominant frequency and other components. In Figure 3, the positive gene basis axes divide the dominant frequency trajectory into phases of gene expression. Alternatively these plots are the view of the positive gene basis axes as seen through a transparent dominant frequency plane. In Figure 3(a) the cdc genes make a cell-cycle clock on the cdc15 dominant frequency plane. In each of the panes in Figure 3 the cdc clock for the particular experiment surrounds the trajectory. Figure 3(b) portrays the view of all the gene basis axes, while Figures 3(c)–(e) focus on particular groups of genes, demonstrating gross consistency between the experiments with some variation in the details (whether biological or experimental). If the dominant frequency trajectories were perfectly circular (and that component of each gene's trajectory perfectly sinusoidal) the gene basis axis projection would cross the trajectory at its maximum amplitude. Because these trajectories are discrete and only approximately circular, the directions of the gene basis axes approximate this ideal.

The three yeast experiments use different initial synchronization techniques on different strains of yeast. If the dominant frequency planes are significantly related to each other, it would suggest that, despite the discrepancies

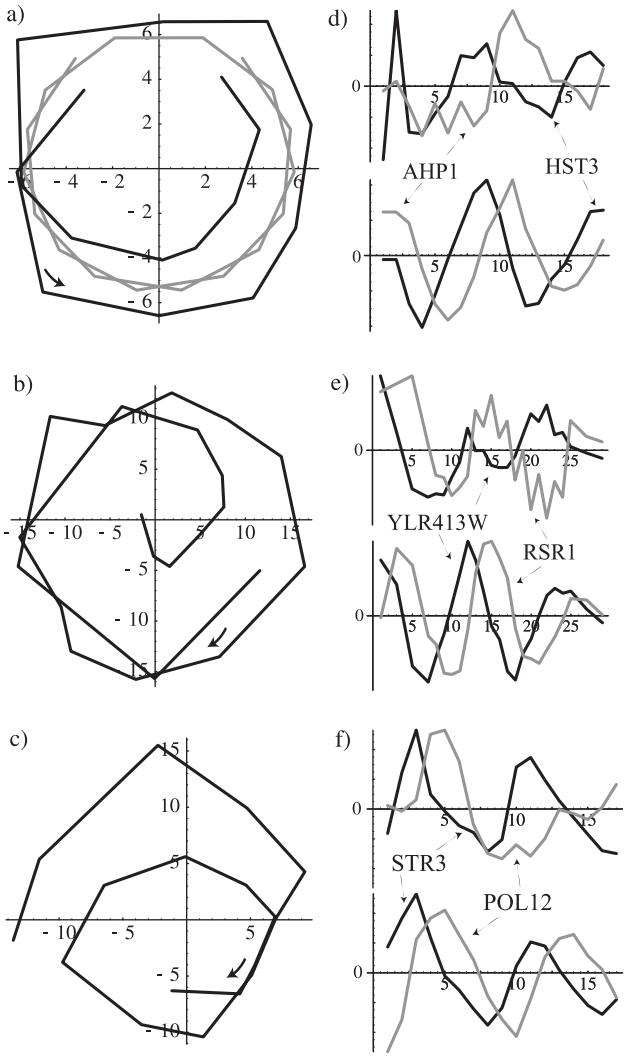

Fig. 2. (a)–(c) Projected trajectories onto the dominant frequency plane for each experiment. Axes are in log-relative expression units for the basis vectors of the plane, linear combinations of actual gene expression levels. The trajectories have been oriented so that the approximate peak of HTA1 expression is at noon. (a) Alpha factor; (b) Cdc15; (c) Cdc28. (d)–(f) The dominant frequency signals of pairs of genes which form basis waveforms for the projected trajectories. The members of a pair of waveforms are orthogonal to each other and the two for each experiment lie at 90° to each other. The amplitudes have been scaled. (d) Alpha factor; (e) Cdc15; (f) Cdc28.

in starting conditions and genetic background, all three datasets share a similar dominant temporal interaction pattern of gene expression levels. To this end, we computed the pairwise canonical correlations between the three planes. Any two planes will have vectors (one in each plane) that are maximally correlated. This correlation and the correlation between the orthogonal vectors to these

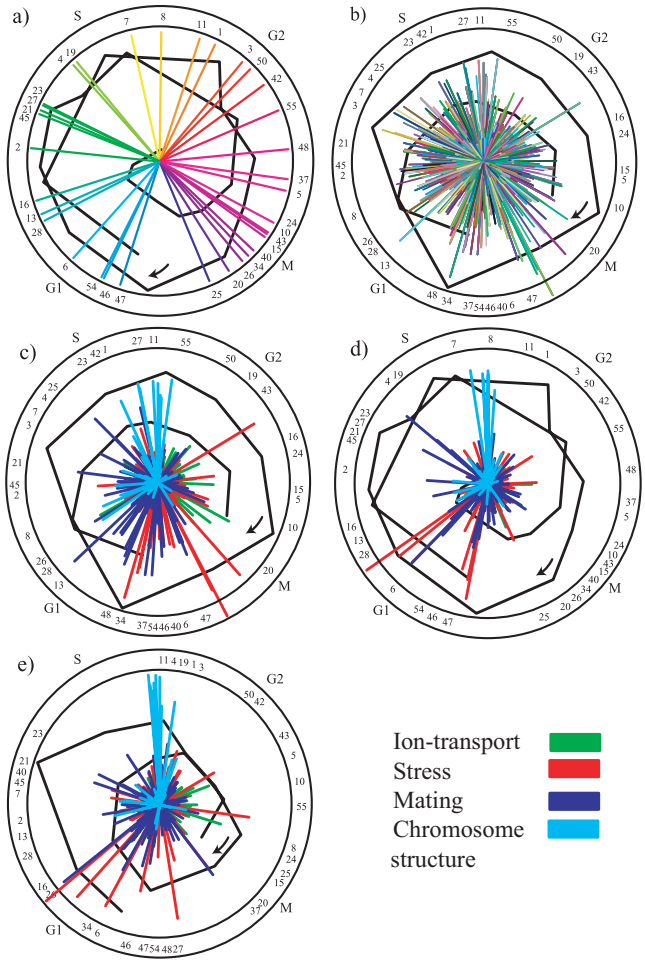

Fig. 3. Cell cycle clocks: phases of gene expression illustrated by projections of gene basis axes onto the dominant frequency planes oriented as in figure 2 with the projection of the HTA1 positive axis at noon: (a) Cdc genes (labeled by number) on the cdc15 plane making a cell-cycle clock. Length is irrelevant in this plot; (b) All genes on the alpha factor plane with the alpha factor cdc clock marking the cell-cycle progression on the outside. Lengths are proportional to the percentage of the variance in a gene's expression accounted for by the dominant frequency, or alternatively to the angle between the dominant frequency plane and and the gene basis axis; (c)–(e) Selected categories of genes projected onto the dominant frequency claims for the various experiments with cdc clocks around the outside. Lengths are as in (b): (c) Alpha factor; (d) Cdc15; (e) Cdc28.

vectors are the canonical correlation between the two planes (Seber, 1984). This technique can be straightforwardly extended to linear subspaces of any dimension. Two random planes in a 5541 dimensional space will have maximal correlation near zero with high probability (e.g. the probability of a chance correlation as high as 0.08 is on the order of 10^{-6}). In this reduced 15 dimensional subspace, the mean maximal correlation between two random planes is around 0.5. The maximal canonical correlations between these planes (and their probabilities based on a simulation) were: between alpha-factor and cdc15, 0.83 ($p \cong 0.001$); cdc15 and cdc28, 0.86 ($p \cong 005$); cdc28 and alpha-factor, 0.74 ($p \cong 0.03$). The correlations between the vectors orthogonal to these maximally correlated vectors are all significant at $p < 0.001$. Furthermore, the various trajectories retained their qualitative appearances after direct projection onto the subspaces of the others, a result not to be expected if the subspaces did not overlap. The dynamical subspaces spanned by the three different experiments are identical.

Comparison to Fourier and static singular value decompositions

To evaluate the efficacy of our non-parametric geometric procedure, we simulated a dataset of expression levels of 5000 genes over 20 timepoints cycling with period 10 with varying phases and with amplitudes drawn from our estimate of the alpha-factor cycle. The periodic dynamics were assumed to follow a simple oscillation with each vector component given by $x_i = a_i \sin(wt) + b_i \cos(wt)$, where w determines the period and a_i and b_i determine the phase. To each vector component, we added random noise drawn from a normal distribution with mean zero. The variance of the added noise was determined by the actual variance of the yeast data with a varying multiplicative factor ranging in successive simulations from 0.15 to 1. For the Fourier analysis, we performed a Fourier decomposition of the data matrix, found the dominant frequency for each gene and took the mode of this distribution of frequencies, and then projected the data matrix onto the plane of that modal frequency. (Note that this majority rules method is quite different in spirit from our averaging method.) We also used SVD to estimate the two-dimensional largest variance plane. Each estimate was evaluated by computing a canonical correlation of the estimated plane to the true plane.

Over this range of noise and for this simple oscillatory dynamics, Fourier analysis, static SVD analysis (Rifkin *et al.*, 2000), and the method presented here all found planes with similar canonical correlations to the true plane (regressions of first canonical correlate on the noise factor: $1.06 - (0.52 \pm 0.0029)x$, $1.08 - (0.58 \pm 0.0039)x$, $1.07 - (0.57 \pm 0.0036)x$). However, it is unlikely that the entire gene expression dynamics of a group of cells will be confined to a single plane. Typical cells differentiate, respond to environmental cues and perturbations, move, age, and grow, not all of which will be repeated in every cell cycle. In the current yeast dataset case, the perturbations due to initial conditions, e.g. the alpha factor arrest, induce directional deviations from the normal cell-cycle as seen in Figure 1. In geometric terms, there may

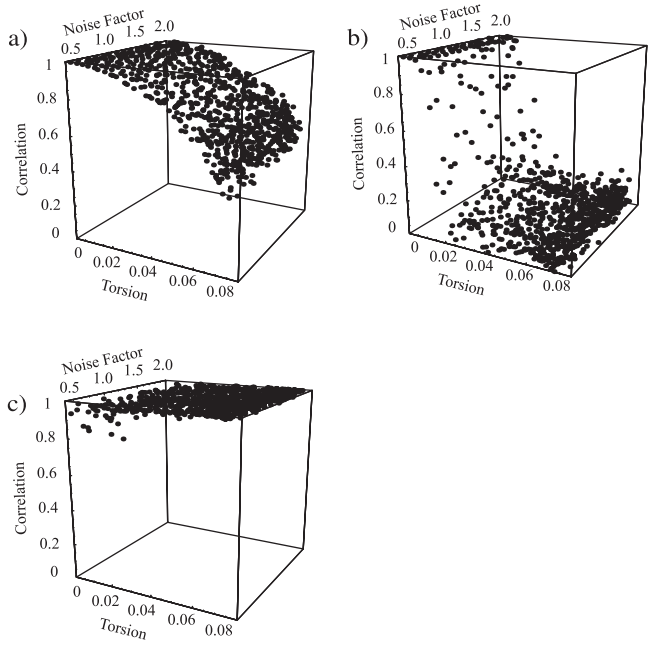

Fig. 4. Performance of cell-cycle estimate procedures against torsion and noise (see text for details): (a) Fourier analysis; (b) SVD; (c) Geometric decomposition The points are confined to a three dimensional volume, so the estimated plane will (practically) always intersect the true plane. Therefore the second canonical correlate, not the first, is of interest.

be a directional component to the trajectory separate from the cycle. Ideally, an analysis technique should be able to discriminate between the dynamical contributions of these cellular activities. To test this ability, in addition to the random noise factor (ranging now from 0.5 to 2), we added a linear directional component to the data such that the simulated cells corkscrewed through the gene expression space. As the torsion of the trajectory increased, Fourier analysis and SVD lost their ability to estimate the true plane—even when Fourier analysis only used the information from the actual cycling frequency—while our geometric decomposition continued to isolate the cyclic dynamics (Figure 4).

Geometry of the perturbations

The ability to separate cyclical from directional components of the trajectory means that this geometric analysis can also be used to partial out the fundamental cell-cycle related expression activity and used to examine gene expression patterns related to other physiological functions, such as the perturbations due to initial conditions. We begin by assuming that our dominant frequency plane geometrically characterizes the cell-cycle related gene expression. We then looked for vector directions away from

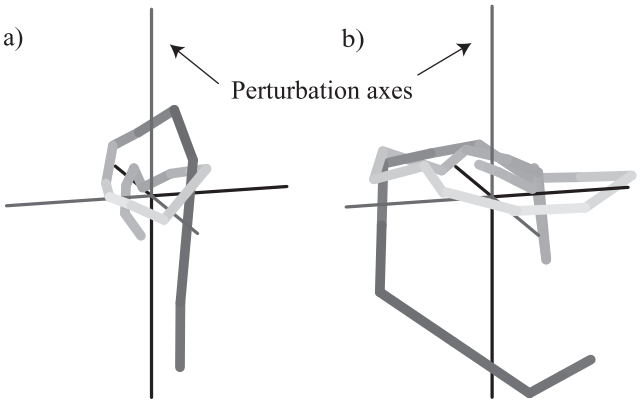

Fig. 5. The cell state relaxing back to the dominant frequency plane along a perturbation axis for (a) alpha factor and (b) cdc15. The trajectories have been smoothed for illustration by averaging successive timepoints. The flow proceeds from bottom to top.

this plane with biological significance. Figure 1(a) shows that in the first timepoints of alpha-factor experiment, the cell state approaches the dominant frequency plane from a distant initial condition. Given this apparent relaxation into the cell cycle, the first five timepoints can be used to estimate the direction of the perturbation. We searched for a vector direction perpendicular to the common frequency planes along which the trajectory varied the most for the first five timepoints while varying the least for the latter timepoints—the cycling timepoints. That is, we found a vector direction which best explains the variation for the first five time points while being least correlated with the later time points. As Figure 5(a) shows, the perturbed cell in the alpha factor experiment approaches the cell cycle plane from an angle. We used canonical correlations and a set of random vectors to assess whether certain functional groups contributed to this perturbation axis more and less than one would expect by chance ($p < 0.05$). Genes whose proteins are involved in cell growth, division, and DNA synthesis, chromosome structure, signal transaction, mating and the cell cycle genes identified in Spellman *et al.* (1998) contributed significantly more than random; genes with proteins involved in protein destination, protein synthesis, and, more specifically, ribosomal protein genes contributed significantly less. A particular gene may be both perturbed and cycling because its trajectory is a superposition of several components, and a gene basis axis or subspace may lie significantly close to two orthogonal subspaces.

After an initial displacement that affects some of the cycling genes, the alpha factor trajectory streaks toward the dominant frequency plane before beginning to cycle. In comparison, cdc15 the trajectory spirals towards the cell cycle plane beginning its cycle from the outset (Fig-

ure 5(b)). This spiral may indicate a division between the dynamics of perturbation response and the dynamics of normal cell cycle activity, a division which is not as crisp in the alpha-factor experiment. In the cdc15 experiment, the measurements were made on a heat-shock synchronized temperature-sensitive cdc15-2 strain followed by a shift back to a permissive temperature (Spellman *et al.*, 1998). The perturbation axis falls significantly in the subspaces spanned by genes involved in metabolism, energy, metabolism, stress response, and the previously identified cell cycle genes, and is significantly orthogonal to the subspace spanned by genes involved in protein synthesis, protein destination, intracellular transport, and transcription.

DISCUSSION

The tools we present here are first steps towards analyzing genome-wide gene expression data from a structural dynamical systems view. But even these preliminary techniques yield a biologically useful and consistent picture. Within the resolution of the yeast data, we have demonstrated that a low-dimensional periodic trajectory characterizes the dynamics of the yeast cell cycle. The dominant dynamical interactions of the genome for different initial conditions are significantly related to one another and involve the majority of the genome. However, the data has coarse resolution and is noisy and short. Simulation experiments suggest that the noise is not problematic because the large number of genes readily reveals the global dynamical structure (data not shown). The time resolution at approximately 10 minute intervals is somewhat coarse, but we did not detect higher frequency dynamics within the resolution limit of the data (but see Klevecz and Dowse, 2000). The shortness of the time-series is more problematic. This prevents us from making any quantitative assessment of the embedding dimension and other standard dynamical analyses. Nevertheless, the results confirm the suggestions of Klevecz and Dowse (2000) and Holter *et al.* (2000, 2001) that the temporal dynamics of the gene expression in the cell cycle are rather simple and show how this simplicity becomes apparent in their geometrical context. The notion that genomic dynamics must be complicated arises, in part, from the combinatorial possibilities for its very many components (Boguski, 1999). However, in the yeast cell cycle, the dynamics of the large number of genes are restricted to a moderate number of possibilities, greatly limiting the possible global complexity. From a systems viewpoint, the complexity is bounded by a set of constraints, both temporal and physiological.

As early as the 1940's, C.H.Waddington and M.Delbrück speculated on the dynamical nature of molecular interactions, especially about whether the entire genome is strongly coupled or compartmentalized and whether the cascade of molecular dynamics follows a sequence of transitions between quasi-stable internal states (Waddington, 1939; Delbrück, 1949; Thom, 1983). These yet unanswered questions have important implications on how we understand the cell at the whole genome level and how we understand the molecular transitions involved in differentiation and pathological responses. Unfortunately, the number of measurements required to characterize the biochemistry of a cell made any empirical explorations of these ideas unfeasible. Large-scale gene expression analyses, while not characterizing all of the relevant variables in a cell, enable us to make reasonable approximations to the states of cells under certain conditions. Genome-wide time-series measurements of physiological (DeRisi *et al.*, 1997) or developmental (White *et al.*, 1999) processes may soon enable us to revive these long dormant biological ideas with adequate data and forge them into a useful tool in the shift from studies of a few genes to systemic analyses of entire genomes.

ACKNOWLEDGEMENTS

We thank members of the Kim Lab and Guy Sella for comments on the material presented here. This work is supported by a National Science Foundation Graduate Fellowship to S.A.R. and a Merck Genome Research Institute Grant and NIH pre-NPEBC to J.K.

REFERENCES

Abarbanel,H.D.I. *et al.* (1993) The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.*, **65**, 1331–1392.

Boguski,M.S. (1999) Biosequence exegesis. *Science*, **286**, 453–455.

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

Delbrück,M. (1949) *Unités Biologiques Douées de Continuité Génétique*. CNRS, Paris.

DeRisi,J.L.*et al.* (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.

Friedman,J.H. and Tukey,J.W. (1974) A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, **c23**, 881–890.

Holter,N.S. *et al.* (2000) Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc. Natl Acad. Sci. USA*, **97**, 8409–8414.

Holter,N.S. *et al.* (2001) Dynamic modeling of gene expression data. *Proc. Natl Acad. Sci. USA*, **98**, 1693–1698.

Klevecz,R.R. and Dowse,H.B. (2000) Tuning in the trascriptome: basins of attraction in the yeast cell cycle. *Cell Prolif.*, **33**, 209–218.

Packard,N.H. *et al.* (1980) Geometry from a time series. *Phys. Rev. Lett.*, **45**, 712–714.

Rifkin,S.A. *et al.* (2000) Constraint structure analysis of gene expression. *Func. Integr. Genomics*, **1**, 174–185.

Schena,M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.

Seber,G.A.F. (1984) *Multivariate Observations*. Wiley, New York.

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

Takens,F. (1981) Detecting strange attractors in turbulence. *Lecture Notes Math.*, **898**, 366.

Thom,R. (1983) *Mathematical Models of Morphogenesis*, Translated by Brookes,W.M. & Rand,D., Wiley, New York.

Waddington,C.H. (1939) *An Introduction to Modern Genetics*. Macmillan, New York.

White,K.P. *et al.* (1999) Microarray analysis of *Drosophila* development during metamorphosis. *Science*, **286**, 2179–2184.

Wolfram,S. (1999) *The Mathematica Book*, 4th edition, Cambridge University Press, Cambridge.