

RESEARCH ARTICLE

# Determination of Temporal Order among the Components of an Oscillatory System

Sandra Barragán<sup>1</sup>, Cristina Rueda<sup>1</sup>, Miguel A. Fernández<sup>1</sup>, Shyamal D. Peddada<sup>2\*</sup>

**1** Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, Valladolid, Spain, **2** Biostatistics Branch, NIEHS (NIH), Research Triangle Park, North Carolina, United States of America

\* [peddada@niehs.nih.gov](mailto:peddada@niehs.nih.gov)



**OPEN ACCESS**

**Citation:** Barragán S, Rueda C, Fernández MA, Peddada SD (2015) Determination of Temporal Order among the Components of an Oscillatory System. PLoS ONE 10(7): e0124842. doi:10.1371/journal.pone.0124842

**Editor:** Juan Mata, University of Cambridge, UNITED KINGDOM

**Received:** March 9, 2015

**Accepted:** March 17, 2015

**Published:** July 7, 2015

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** We analyzed publicly available data. References to the data are provided in the manuscript.

**Funding:** The research of SB, MAF and CR was funded by the Spanish Ministerio de Ciencia e Innovación grant (MTM2012-37129). SB was also funded by the Junta de Castilla y León, Consejería de Educación and the European Social Fund within the Programa Operativo Castilla y León 2007–2013. The research of SDP was supported by the Intramural Research Program of the National Institute of Environmental Health Sciences (Z01 ES101744-04).

## Abstract

Oscillatory systems in biology are tightly regulated process where the individual components (e.g. genes) express in an orderly manner by virtue of their functions. The temporal order among the components of an oscillatory system may potentially be disrupted for various reasons (e.g. environmental factors). As a result some components of the system may go out of order or even cease to participate in the oscillatory process. In this article, we develop a novel framework to evaluate whether the temporal order is unchanged in different populations (or experimental conditions). We also develop methodology to estimate the order among the components with a suitable notion of “confidence.” Using publicly available data on *S. pombe*, *S. cerevisiae* and *Homo sapiens* we discover that the temporal order among the genes *cdc18*; *mik1*; *hhf1*; *hta2*; *fkp2* and *kfp5* is evolutionarily conserved from yeast to humans.

## Introduction

Oscillatory systems arise naturally in biological sciences such as in, circadian biology [1–3], cell biology [4–9], endocrinology [10], metabolic cycle [11], evolutionary psychology [12, 13], motor behavior [14], and so on. An unperturbed oscillatory system is a tightly regulated temporal process with several components that execute their functions in an orderly manner like an orchestra. Thus a temporal order among the components is intrinsic to an oscillatory system. For example, it is well-known that our daily sleep and wake patterns lead to a sequence of biochemical events in the body in an orderly manner, such as breakdown of molecules to generate energy (catabolism) during the wake period and anabolism that takes place during the sleep period where growth of tissues occurs. Discussing the oscillations of individual neurons of the suprachiasmatic nuclei (SCN) in a 24 hour period, [15] describe the temporal order of circadian genes such as *Bmal1*, *Clock*, *Period*, *Cryptochrome*, *Rev-erb* [3]. The effect of sleep patterns on the temporal order of several circadian genes and consequently the effect on oxidative stress and metabolism was discussed in [16].

The common underlying question of scientific interest is to determine (relative) time to peak expression of genes participating in the oscillatory system [7, 12], i.e. to determine the underlying temporal order. A related question of interest is to understand the differences in the

**Competing Interests:** The authors have declared that no competing interests exist.

oscillatory systems of different populations or experimental groups such as; environmental conditions, species, organs within a species [17, 18], etc. Often raw expressions from time course experiments are used to make such inferences. For example studying circadian genes in various tissues in a whole animal and those in a cell line, [2] note that “relative phasing of core clock genes was estimated by visual inspection and plotted on a circular phase map.” Although such visual methods are easy to understand and implement, and widely used, they ignore uncertainty associated with the estimated values of angular parameters. Consequently it is not entirely surprising that there are disagreements in the literature regarding phases and phase order of various cell-cycle genes, even within the same species let alone across species [19].

Notice that, in this paper, we are not trying to establish which genes are periodic [20, 21] or to cluster genes according to their expression pattern [22, 23] but to see if the different phase angles assigned in different experiments to orthologs coming from several species are compatible with a common ordering of the phase angles of these genes across the species considered.

It is important to note that phase or time to peak expression of an oscillatory gene is a parameter on a unit circle and not on the real line. Consequently standard methods of analysis, such as the t-test or ANOVA, designed for real line data, cannot be used. Toy example in [S1 File](#) amplifies the problem of using such methods for angular data. Yet, they are commonly used in the literature [16], which may potentially result in incorrect or meaningless interpretations of the data.

Analysis of angular data has a long history with well-developed theory and methodology documented in several books [24, 25]. Until recently much of the literature was developed for drawing inferences on individual parameters, but not for studying order among a set of angular parameters (e.g. phases of a system of oscillatory genes), which is the focus of this article. More precisely, suppose an oscillatory system consists of genes,  $g_1, g_2, g_3, \dots, g_8$ , with phase angles  $\phi_1, \phi_2, \dots, \phi_8$ , respectively. Then a researcher is typically interested in determining the circular order (temporal order) among these phase angles. For example, determine whether  $g_1$  peaks before  $g_2$  which peaks before  $g_3$ , etc.  $g_7$  peaks before  $g_8$  and  $g_8$  before  $g_1$ . Mathematically, determine if  $\phi_1$  precedes  $\phi_2$  which precedes  $\phi_3$  and so on  $\phi_7$  precedes  $\phi_8$  which in turn precedes  $\phi_1$  around the unit circle (e.g. [Fig 1](#)). We shall denote the order by  $\phi_1 \preceq \phi_2 \preceq \dots \preceq \phi_7 \preceq \phi_8 \preceq \phi_1$ .

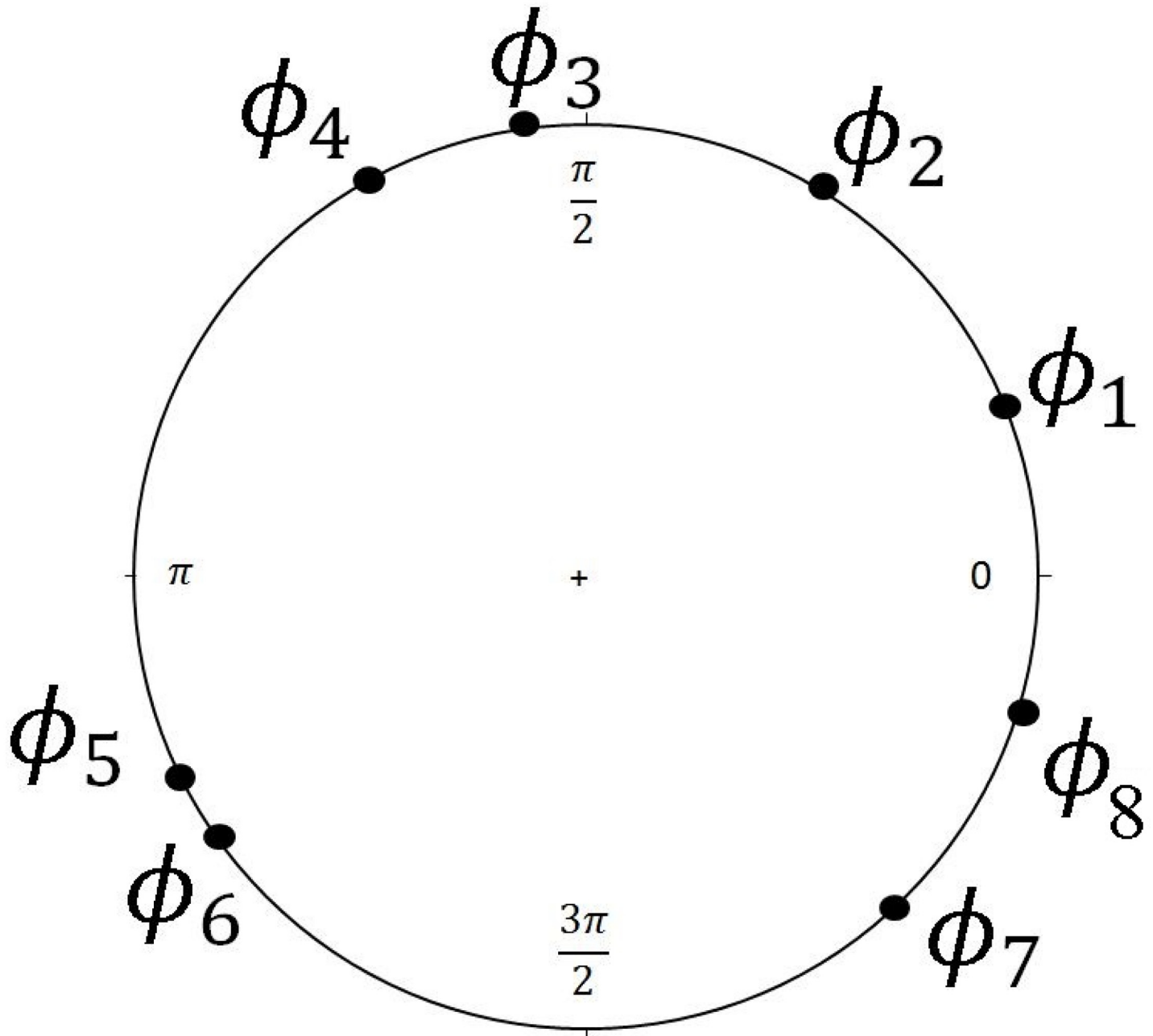
For two or more study groups (e.g. organs or species, etc.), researchers are typically interested in testing whether the temporal order of a set of oscillatory genes is conserved. If so, they are interested in discovering the common temporal order with an estimate of confidence. In this article we introduce a statistical framework to address such problems. We illustrate the methodology by discovering a temporal order among a core set of cell cycle genes that is conserved from yeast to humans. Although the methodology described in this paper is suitable for any oscillatory system, for convenience of exposition we use cell-cycle terminology.

The temporal order derived by the proposed methodology could potentially help biologists to discover or explore novel regulatory relationships among the genes in the oscillatory system. Thus our methodology can potentially lead to new hypotheses for biologists to study.

## Materials and Methods

### Estimation of temporal order

Before describing the methodology to test hypothesis regarding the circular order among a set of oscillatory genes, we discuss the problem of estimating their common unknown circular order (assuming it exists). Using this estimator we then develop a statistical procedure to test the null hypothesis that a given set of oscillatory genes in two or more study groups (or populations) share the same temporal order.



**Fig 1. An illustration of the temporal order among genes  $g_1, g_2, \dots, g_8$  whose phase angles are in order along a circle (in counterclockwise direction).**

doi:10.1371/journal.pone.0124842.g001

In addition to estimating the unknown phase angles  $\phi_1, \phi_2, \dots, \phi_n$  the goal is also to estimate the true relative order among them, denoted by  $O = (o_1, o_2, \dots, o_n)$ , where  $\phi_{o_1} \preceq \phi_{o_2} \preceq \dots \preceq \phi_{o_n} \preceq \phi_{o_1}$ . Note that  $O$  is rotation invariant. Thus by moving the pole around the circle between each consecutive pair of angular parameters, we obtain  $n$  possible equivalent orders to  $O$ . The goal is to estimate  $O$  using data obtained from  $p$  experiments. We will denote the estimator of  $O$  as  $\tilde{O}$  and is obtained by the procedure explained below.

Typically, researchers conduct time course gene expression studies to obtain the phases of each cell-cycle gene. For the  $i^{th}$  gene in the  $j^{th}$  experiment, let  $\theta_{ij}$  denote the estimate of phase angle  $\phi_i$  obtained by using the Random Periods Model, RPM [26]. Since the estimates obtained from RPM are not constrained by any order among the phase angles, they are called the unconstrained estimators. Accordingly, let

$$\Theta_j = (\theta_{1j}, \dots, \theta_{ij}, \dots, \theta_{nj})' \quad \forall j = 1, \dots, p,$$

denote the vector of RPM estimators of  $(\phi_1, \phi_2, \dots, \phi_n)'$  obtained from the  $j^{th}$  experiment. Stacking all such estimators for the  $p$  experiments together, we have  $\Theta = [\Theta_1, \dots, \Theta_p]$ .

We estimate  $O$  using the minimum distance principle as follows. Let  $\mathfrak{D}$  denote the set of all possible orders among  $\phi_1, \phi_2, \dots, \phi_n$ . Using the data from the  $j^{th}$  experiment, under a given order  $O \in \mathfrak{D}$ , let  $\tilde{\Theta}_j^{(O)} = (\tilde{\theta}_{1j}^{(O)}, \tilde{\theta}_{2j}^{(O)}, \dots, \tilde{\theta}_{nj}^{(O)})'$  denote the circular isotonic regression estimator (CIRE) of  $\phi_1, \phi_2, \dots, \phi_n$  under the circular order constraint  $O$  [8].

As in [4] and [8] the sum of circular errors (SCE), which serves as the distance between  $\Theta_j$  and the order  $O$ , is defined as follows.

**Definition 1** The Sum of Circular Errors (SCE) corresponding to circular order  $O$  for data in the  $j^{th}$  experiment,  $\Theta_j = (\theta_{1j}, \theta_{2j}, \dots, \theta_{nj})'$ , is given by:

$$SCE(\Theta_j, \tilde{\Theta}_j^{(O)}) = \sum_{i=1}^n \{1 - \cos(\theta_{ij} - \tilde{\theta}_{ij}^{(O)})\}.$$

For a given order  $O$ , its mean sum of circular errors (MSCE) over all  $p$  experiments is given by:

$$d(\Theta, O) = MSCE(\Theta, \tilde{\Theta}^{(O)}) = \sum_{j=1}^p \omega_j \frac{1}{n} SCE(\Theta_j, \tilde{\Theta}_j^{(O)}), \tag{1}$$

where  $\omega_j$  is the weight associated with  $j^{th}$  experiment. Suppose  $\theta_{ij} \sim M(\phi_i, \kappa_j)$  where  $M$  denotes the von-Mises distribution with angular mean  $\phi_i$  and concentration parameter  $\kappa_j$  (known), then we define  $\omega_j = \frac{\kappa_j}{\sum_{j=1}^p \kappa_j}$ .

The optimum circular order can be obtained by solving the following minimization problem:

$$\min_{O \in \mathfrak{D}} d(\Theta, O) = \min_{O \in \mathfrak{D}} \sum_{j=1}^p \omega_j \frac{1}{n} SCE(\Theta_j, \tilde{\Theta}_j^{(O)}). \tag{2}$$

The above problem resembles the classical problem of determining the “true” order or ranks among  $n$  objects using the scores assigned by  $p$  independent “judges”. For example, suppose there are  $n$  gymnasts competing in an event and there are  $p$  judges assigning scores to each of the contestants. The goal is to estimate the true rank among the  $n$  contestants using the scores assigned by the  $p$  judges. Although this NP-hard problem [27] is well-studied in the Euclidean space [28–31], it has not been discussed for other geometries such as the circle. Due to the underlying geometry, the Euclidean space based methods cannot be directly applied here.

Since the above formulation is NP hard even for real line data, we obtain an approximate solution by reformulating Eq (2) as a traveling salesman problem (TSP) which is known to be NP-complete [32, 33].

The TSP is well-studied in the graph theory literature [34–36] and is formulated using a weighted graph which is a triple consisting a set of nodes, a set of edges and a cost associated

with each edge. The purpose of TSP is to determine the tour with minimum total cost, where a tour is the path traveled by a salesman such that all nodes are visited and each node is visited exactly once. In our application genes are the nodes, edge is the path between two genes and a tour is a circular order among the genes. For the simulations we have performed with a moderate number of elements to be ordered (notice that, as usual in these problems, the optimum value cannot be computed in a reasonable time when the number of elements increases), this TSP approach performed very well so that we expect the tour with minimum total cost to be a good approximation to our original problem [Eq \(2\)](#).

To determine the tour with minimum total cost we first define the total cost of traveling between nodes  $h$  and  $k$  in the  $p$  experiments ( $E_{hk}$ ) as the weighted sum  $E_{hk} = \sum_{j=1}^p \omega_j E_{hk}^j$ , where  $E_{hk}^j$  is the cost in the  $j^{th}$  experiment. For each  $j$ , the cost  $E_{hk}^j$  is defined through a measure of distance between the nodes  $h$  and  $k$ . A common measure of distance between a pair of points on a unit circle is  $1 - \cos(\theta_{kj} - \theta_{hj})$  [25]. This measure is symmetric but cell-cycle is a biological process where the functional relations between genes are not symmetric. Without loss of generality the sequential order of events (or phases) of cell-cycle may be represented in the counter-clockwise direction around the unit circle. For this reason we define distances asymmetrically, depending upon whether the salesman is traveling counter-clockwise ( $d_1$ ) or clockwise ( $d_2$ ) as follows:

$$d_1^j(h, k) = \begin{cases} 1 - \cos(\theta_{kj} - \theta_{hj}) & \text{if } 0 \leq \theta_{kj} - \theta_{hj} \leq \pi \\ 3 - \cos(\theta_{kj} - \theta_{hj} - \pi) & \text{if } \pi < \theta_{kj} - \theta_{hj} \leq 2\pi, \end{cases}$$

$$d_2^j(h, k) = \begin{cases} 3 - \cos(\theta_{kj} - \theta_{hj} - \pi) & \text{if } 0 \leq \theta_{kj} - \theta_{hj} \leq \pi \\ 1 - \cos(\theta_{kj} - \theta_{hj}) & \text{if } \pi < \theta_{kj} - \theta_{hj} \leq 2\pi. \end{cases}$$

Asymmetric distances are common in the application of TSP and are widely studied [37]. Using the above distances, we define the cost of traveling between the nodes  $h$  and  $k$  in the experiment  $j$  as follows:

$$E_{hk}^j = \min(d_1^j, \alpha d_2^j),$$

where  $\alpha$  represents the penalty for traveling in the clockwise direction. Based on extensive simulation studies using different values of  $\alpha$ , we found  $\alpha = 3$  provided the best results and hence we use this value throughout the paper.

Let  $X$  denote an  $n \times n$  matrix where  $x_{hk} = 1$  if the salesman travels directly from node  $h$  to node  $k$ , otherwise let  $x_{hk} = 0$ . No sub-tours are allowed. Let  $\mathcal{X}$  denote the collection of all such matrices which represent a tour. Then, TSP reduces to solving the following minimization problem

$$\min_{X \in \mathcal{X}} \sum_{hk} x_{hk} E_{hk} = \min_{X \in \mathcal{X}} \sum_{j=1}^p \omega_j \left( \sum_{hk} x_{hk} E_{hk}^j \right). \tag{3}$$

We denote as  $\tilde{X}^0$  the solution of [Eq \(3\)](#). The resulting order among the nodes denoted as  $\tilde{O}^0$  is taken to be an approximate solution to [Eq \(2\)](#). To improve this approximation, we refine it by eliminating any local bumps (i.e misalignment of order). The chances of misalignment of order can occur locally as the number of nodes (genes) increases or as some nodes get closer to each other. We accomplish this by modifying the *Local Kemenization* algorithm that was originally developed by [38] for the Euclidean data to the present context of circular data. We call the resulting algorithm the *Circular Local Minimization* algorithm. It consists of checking each

consecutive triple  $(h, k, l)$  of adjacent elements in  $\tilde{O}^0$  (while preserving the estimated circular order among rest of the elements) to see if a permutation of  $o_h, o_k, o_l$  improves the result. Namely, we calculate the MSCE as defined in Eq (1) between the possible new circular order, with the permutation, and the data. If the new MSCE is smaller then the circular order is appropriately changed. The resulting refined estimate is  $\tilde{O}$ .

### Comparison of temporal orders

Suppose there are  $S$  experimental groups and  $n$  genes in each group that oscillate. Let  $O_s, s = 1, 2, \dots, S$ , denote the order among the phase angles of the  $n$  genes in the  $s^{th}$  group. Then the problem of interest is to test:

$$H_0 : O_1 = O_2 = \dots = O_S = O_*$$

$$H_1 : H_0 \text{ not true.}$$

The equality sign “=” in the null hypothesis denotes “identical circular order” which would be represented by  $O_*$ . Corresponding to the  $s^{th}$  group,  $s = 1, 2, \dots, S$ , suppose there are  $p_s$  experiments. Let  $P = \sum_{s=1}^S p_s$  denote the total number of experiments. Then the above hypothesis can be tested along the lines of classical analysis of variance (ANOVA). Let  $\tilde{O}_s$  denote the estimated order obtained with the experiments from the  $s^{th}$  group and  $\tilde{O}_*$  denotes the estimated common order under the above null hypothesis obtained by using the data from  $P$  experiments combining data from all  $S$  groups.

Let  $d(\Theta^{(s)}, \tilde{O}_s) = \sum_{j=1}^{p_s} \omega_j^s \frac{1}{n} SCE(\Theta_j^{(s)}, \tilde{\Theta}_j^{(\tilde{O}_s)})$  denote the corresponding value of the objective function Eq (2) for the experiments in the  $s^{th}$  group. Here  $\omega_j^s$  denotes the weight corresponding to the  $j^{th}$  experiment in the  $s^{th}$  experimental group. Adding over all  $S$  experimental groups we have the following which resembles the *within groups variability*,  $\sum_{s=1}^S d(\Theta^{(s)}, \tilde{O}_s)$ .

Let  $d(\Theta^{(\cdot)}, \tilde{O}_*) = \sum_{s=1}^S \sum_{j=1}^{p_s} \omega_j^s \frac{1}{n} SCE(\Theta_j^{(s)}, \tilde{\Theta}_j^{(\tilde{O}_*)})$  denote the corresponding value of the objective function Eq (2) using the data for all  $P$  experiments. This expression resembles the *global variability*. Hence, resembling the classical ANOVA, one may consider any monotonic function of the following test statistic for testing above null hypothesis:

$$T = \frac{d(\Theta^{(\cdot)}, \tilde{O}_*) - \sum_{s=1}^S d(\Theta^{(s)}, \tilde{O}_s)}{d(\Theta^{(\cdot)}, \tilde{O}_*)}$$

Since not all species (in this case the experimental groups) are represented by equal number of experiments and not all experiments are subject to same experimental error/noise, we use a “weighted” resampling method to derive the p-values based on  $T$  that takes into account all such features of the data. The goal is to create artificial species that resemble the original species in terms of the expected proportions of experiments within each species. We therefore select experiments randomly with replacement and equal probabilities per species and per experiment within species. Thus each experiment in the  $s^{th}$  species has a probability  $1/(Sp_s)$  of selection. Under this sampling scheme we select  $P$  random experiments with replacement from the  $P$  actual experiments and assign the first  $p_1$  to artificial species 1, the next  $p_2$  to artificial species 2 etc. The weights per experiment are suitably calculated with each resample. Extensive simulation experiments, under a variety of configurations of phase angles and the order among phase angles were conducted to evaluate the Type I error rate of the proposed resampling scheme. Based on our results, detailed in the S3 File, we discover that the proposed resampling procedure yields honest statistical test in the sense that the estimated Type I error never exceeds the

nominal rate of 0.05 by more than a standard error. Furthermore, the proposed methodology enjoys very high power even under minor departures from the null hypothesis.

For genes identified to satisfy a common global order, we use the above resampling procedure in combination with the estimation procedure described in the previous section to estimate the common global partial order with confidence as follows. We take the union of most frequent orders coherent with the common global order to deduce the global partial order. The sum of the frequencies of those orders relative to the total number of resamples provides the confidence coefficient. To illustrate the methodology, suppose  $g_1, g_2, \dots, g_5$  are determined to satisfy common global order among 3 species according to the above test. Suppose we obtain 1000 samples according to the above resampling scheme and for 600 of them the global order is  $g_1 \preceq g_3 \preceq g_4 \preceq g_5 \preceq g_2 \preceq g_1$  and for 300 of them the global order is  $g_1 \preceq g_3 \preceq g_4 \preceq g_2 \preceq g_5 \preceq g_1$ . For the remaining 100 resamples, suppose the global orders are arbitrarily distributed among the other possible orders. Note that in a large proportion of resampled data the order between  $g_2$  and  $g_5$  is not consistent. In 60% of the resamples  $g_5$  precedes  $g_2$  whereas in 30% of the resamples the order is reversed. In such cases we assign a “partial order” to indicate that the order between  $g_2$  and  $g_5$  is undetermined. Thus the global partial order in this toy example is given by  $g_1 \preceq g_3 \preceq g_4 \preceq \{g_5, g_2\} \preceq g_1$  with 90% confidence.

## Results

### Motivation and background

Since cell division cycle is an essential process for growth and development of all living organisms, there has been considerable interest among cell biologists to identify cell-cycle genes that are evolutionarily conserved in their functions across multiple species [5–7, 9, 19, 39]. Cell-cycle is a well-coordinated process where events must take place in an orderly fashion for a successful cell division. Hence genes participating in the cell division cycle express in an order according to their function. Throughout this section we focus on only those cell-cycle genes that have a periodic or oscillatory expression (i.e. dynamic) and not those genes that participate in cell division cycle but are static in their expression. Thus a question of interest is to determine, among periodically expressed genes, whether the order of peak expression is evolutionarily conserved. Such questions were extensively discussed and debated during the past decade using gene expression data obtained from budding yeast (*S. cerevisiae*), fission yeast (*S. pombe*) and human Hela cell [5–7, 9]. There are several biological complexities associated with such questions. Firstly, there is considerable disagreement in the literature on the number of genes that are periodic in multiple species [5–7, 9]. As noted in [19], there is considerable disagreement among studies even within the same species. They observed that the three recent studies on the fission yeast [6, 7, 9], together identified about 1400 genes to be periodic, yet only about 10% of these genes were common to all three studies and only about 30% were common to any pair of studies. Given that there is such a large disagreement among studies even within the same species, it is not surprising that there are diverse opinions regarding the number of genes that are periodic in the two species of yeast, namely, the budding yeast (*S. cerevisiae*) and the fission yeast (*S. pombe*). Conservative estimates of the number of genes that are periodic in both species of yeast is about 35 and the number that are periodic in the two yeasts and humans is about 11, see [4]. Furthermore, among genes that were identified to be periodic within the same species by different studies, there are disagreements regarding the phase of peak expression of some genes. For example, [40] assigned *E2F5*, an important transcription factor, to G2/M phase whereas according [41, 42] it peaks during G1/S phase. In the case of fission yeast, [7] assigned *cdc18*, a gene whose protein is essential for the initiation of DNA replication, to G1/S phase whereas [6] as well as cyclebase ([www.cyclebase.org](http://www.cyclebase.org)) [43] assigned the gene to peak in

the M phase. It has been a challenging problem to determine if the phase of a cell-cycle gene is conserved evolutionarily. This is partly because, in addition to the above mentioned issues, the amount of time a cell spends in a given phase is not evolutionarily conserved. For example, a fission yeast cell spends more than 70% of its time in the G2 phase while a budding yeast cell spends about equal time in all phases.

Secondly, a gene needs to be converted into protein before it performs its function. So, even if a cell-cycle gene's function is conserved evolutionarily, its phase may not necessarily be. Thirdly, for a given gene in a particular species it may have multiple orthologs in other species, hence it is a many to many mapping and not a one to one mapping. Since not all orthologs are equally periodic (using the periodicity measure provided in cyclebase), it is a challenging problem to discuss conservation of phase across all orthologs of a gene. Thus it is not surprising for [5] to state that these analysis reveal that periodic expression is poorly conserved at the level of individual genes: conserved periodic expression across the organisms considered is observed in only five cases and for only two of these is the timing conserved as well, namely histones *H2A* and *H4*.

Although, for the above reasons, it may be difficult to ascertain if the phase of a cell-cycle gene is evolutionarily conserved, it may be plausible that the relative order among a collection of cell cycle genes may be evolutionarily conserved. An attempt was made in [4] to answer this question by testing the null hypothesis that the relative order of a subset of cell-cycle genes is conserved between fission yeast and budding yeast. They also performed a similar test between fission yeast and human Hela cells. A drawback with their methodology is that they assume the relative order of cell-cycle genes is known with certainty in one of the two species that are being compared. This is analogous to the "one sample test". Furthermore their methodology is not suitable for testing for the order in more than two species. The present methodology, however, overcomes those deficiencies. In this section we illustrate the methodology by analyzing the phase angle data on 11 cell-cycle genes that are known to be periodic in the 3 organisms. In addition to testing whether the relative order is conserved among the 3 species, we discover the order along with an estimate of confidence in the estimated order. Before proceeding further, we like to remark that [4] do not draw distinctions between orthologs and paralogs since their goal was to determine conservation of order among periodic genes. Again, as noted earlier, not all orthologs of a gene across species are equally periodic -some may not be periodic at all. In such cases, rather asking the question if the relative order of a gene is conserved across all species for all orthologs of a gene, we limit only to the most periodic ortholog (as determined by databases pombase and cyclebase). As in [4] we use the periodicity rank provided in cyclebase. The only exception is human ortholog of *ace2*, which we took to be ZNF367.

Remark: For illustration purposes, in this section we are only considering the case where one is interested in testing the order  $g_1 \preceq g_2 \preceq \dots \preceq g_n \preceq g_1$  among a set of singleton genes  $g_1, g_2, \dots, g_n$ . However, as seen from the results of the analysis provided in the next section, for a given data set, it is possible that our algorithm may declare a subset of these genes to have same order relative to other genes (see Eq (5) in the next section).

If one is interested in the testing for the conservation order of groups of genes (or orthologs) rather than singletons as above, then our methodology can be easily extended to test orders among groups of genes. More precisely, our methodology can be extended to test the order

$$\{g_{11}, g_{12}, \dots, g_{1r_1}\} \preceq \{g_{21}, g_{22}, \dots, g_{2r_2}\} \preceq \dots \preceq \{g_{n1}, g_{n2}, \dots, g_{nr_n}\},$$

where the order among the genes (or orthologs) within {} is irrelevant but as a group they are ordered with the previous and the next group. Thus our method can handle situations where a biologist may be interested in studying the relative order of groups of cell-cycle genes. For



**Table 1. Evolutionarily conserved human cell-cycle genes along with corresponding *S. pombe* and *S. cerevisiae* orthologs.**

Genes	<i>S. pombe</i>	<i>S. cerevisiae</i>	Humans
1	<i>ace2</i>	<i>ACE2</i>	<i>ZNF367</i>
2	<i>cdc18</i>	<i>cdc18</i>	<i>CDC6</i>
3	<i>mik1</i>	<i>SWE1</i>	<i>PKMYT1</i>
4	<i>hhf1</i>	<i>HHF1</i>	<i>HIST2H4B</i>
5	<i>hta2</i>	<i>HTA2</i>	<i>H2AFX</i>
6	<i>fkh2</i>	<i>FKH1</i>	<i>FOXM1</i>
7	<i>kfp5</i>	<i>KIP3</i>	<i>KIF10</i>
8	<i>cig2</i>	<i>CLB1</i>	<i>CCNB1</i>
9	<i>plp1</i>	<i>CDC5</i>	<i>PLK1</i>
10	<i>slp1</i>	<i>CDC20</i>	<i>CDC20</i>
11	<i>rad21</i>	<i>MCD1</i>	<i>RAD21</i>

doi:10.1371/journal.pone.0124842.t001

example, several cell-cycle genes encode proteins that make up large protein assemblies and since all of the subunits within each assembly would be needed for the function of that assembly to be carried out, one may be interested in testing for the order among such large assemblies and not interested in the order among the elements within each assembly.

### Determination of the common temporal order across species

We used the publicly available time course gene expression microarray data on humans (Hela cell), the budding yeast and fission yeast. Specifically, we used the four human data obtained from [40]; six budding yeast data (one from [44], another from [45], two from [46] and two from [20]) and ten fission yeast data (five by [9], three by [6] and two by [7]). Thus we had access to data from 20 experiments conducted on 3 different species. We focused on the expression of 11 cell-cycle genes that are periodic in all 3 species (see Table 1). We estimated the phase angle of each gene within each experiment by fitting the RPM [26]. These estimates, known as the unconstrained estimates because they are obtained with no constraints of the phase angles, are reported in Table A in S2 File. The  $\kappa_j$  values used to determine the  $\omega_j$  weights have been obtained using the procedure developed in [4] and appear in Table B in S2 File.

To determine whether the temporal order is conserved across the 3 species, we first tested the following hypotheses using all 11 genes:

$$\begin{aligned}
 H_0 &: O_{\text{fission yeast}} = O_{\text{budding yeast}} = O_{\text{humans}} \\
 H_1 &: H_0 \text{ not true.}
 \end{aligned}
 \tag{4}$$

Our resampling procedure rejected the null hypothesis with a p-value of 0.0045. This suggests that at least one of the 11 genes was out of order in at least one pair of species. In order to determine a maximum size subset of genes for which the three species share a common order we applied the forward procedure described in the S4 File.

The process ended with the 6 genes, *kfp5*, *fkh2*, *cdc18*, *mik1*, *hhf1* and *hta2*, that failed to reject the null hypothesis with a p-value of 0.488 (see Table 2). Thus we conclude that the temporal order among these genes is evolutionary conserved from yeast to humans with the

**Table 2. Summary of the 4 comparisons.**

	#genes	P-value	Confidence Coefficient
All 3 species together	6	0.488	100%
<i>S. pombe</i> — <i>S. cerevisiae</i>	10	0.336	72.31%
<i>S. pombe</i> —Humans	8	0.436	92.6%
<i>S. cerevisiae</i> —Humans	6	0.119	99.15%

doi:10.1371/journal.pone.0124842.t002

following partial order,

$$cdc18 \preceq mik1 \preceq hhf1 \preceq hta2 \preceq \{fkh2, klp5\} \preceq cdc18 \quad (5)$$

Using the estimation and the resampling methodology described in this article, we estimated that the confidence of this partial order Eq (5) is 100%. The most frequent simple circular order  $cdc18 \preceq mik1 \preceq hhf1 \preceq hta2 \preceq klp5 \preceq fkh2 \preceq cdc18$  had an estimated confidence coefficient of 76.06%.

The two yeasts shared a common ancestor nearly a billion years ago and neither is closer to human beings more than the other [47]. However, according to [48] and [49], while *S. pombe* and metazoan cell-cycle genes retained some of the functions from their common ancestor, the budding yeast cell-cycle genes may have lost them. In fact, relative to *S. cerevisiae* there are proportionally more *S. pombe* genes conserved in metazoans [48, 50]. There are other similarities between *S. pombe* and higher order animals including stress response pathways. For a review one may refer to [47–50]. In view of the above discussion, we performed pairwise comparisons between the 3 species starting with the 6 genes discovered above.

The pairwise forward selection analysis between the two yeasts (*S. pombe* and *S. cerevisiae*) revealed that the relative order of peak expression among 10 out of the 11 genes was conserved with an associated p-value of 0.336. The relative was determined to be  $cdc18 \preceq rad21 \preceq mik1 \preceq \{ace2, hhf1, hta2, cig2\} \preceq \{fkh2, klp5\} \preceq slp1 \preceq cdc18$  with a confidence coefficient of 72.31%. In the case of *S. pombe* and humans the relative order was conserved among 8 of the 11 genes with an associated p-value of 0.436, with relative order  $\{ace2, cdc18\} \preceq mik1 \preceq hhf1 \preceq hta2 \preceq plo1 \preceq \{fkh2, klp5\} \preceq \{ace2, cdc18\}$ . The confidence coefficient associated with this order was estimated to be 92.6%. However, in the case of *S. cerevisiae* and humans we discovered that the order conserved only among the original 6 genes whose order was conserved among the 3 species, namely, *cdc18*, *mik1*, *hhf1*, *hta2*, *klp5* and *fkh2*. Thus, we did not find any additional genes unlike the other 2 pairwise analyses. The p-value associated with these 6 genes in the *S. cerevisiae* and humans pair was 0.119 and the relative order was essentially same as when all three species were considered together but slightly perturbed. The estimated relative order among these 6 genes in the pair *S. cerevisiae* and humans was estimated to be  $cdc18 \preceq mik1 \preceq hhf1 \preceq hta2 \preceq \{fkh2, klp5\} \preceq cdc18$  with a confidence coefficient of 99.15%. These results are summarized in Table 2. Full details of each of the steps in the procedure can be found in the Supporting Information.

Using published phases of these 6 genes in the literature, we summarize the phases of these 6 genes in the 3 species in Table 3. Note that while the phase order of the 6 genes is same across the 3 species their phases are not same across species.

In the case of the two yeasts it is well known that the yeast orthologs of *fkh2* and *ace2* participate in a regulatory network loop where *fkh2* regulates the expression of *ace2* which in turn regulates *fkh2* [51]. Furthermore *fkh2*, the *S. pombe* ortholog of *fkh2*, is one of the regulators of

**Table 3. Phases of the 6 cell-cycle genes whose circular ordered is conserved in the 3 species according to [www.cyclebase.org](http://www.cyclebase.org).**

<i>S. pombe</i>		<i>S. cerevisiae</i>		Humans	
Gene	Phase	Gene	Phase	Gene	Phase
<i>klp5</i>	G2	<i>KIP3</i>	S/G2	<i>KIF10</i>	S
<i>fkh2</i>	G2	<i>FKH1</i>	G2/M	<i>FOXM1</i>	S/G2
<i>cdc18</i>	G1/S	<i>cdc18</i>	M	<i>CDC6</i>	M
<i>mik1</i>	S	<i>SWE1</i>	M	<i>PKMYT1</i>	G1/S
<i>hhf1</i>	S	<i>HHF1</i>	G1/S	<i>HIST2H4B</i>	S
<i>hta2</i>	S/G2	<i>HTA2</i>	G1/S	<i>H2AFX</i>	S

doi:10.1371/journal.pone.0124842.t003

the Cdc15 clusters which peak in late G2 or M phase. In fact, according to [6] its expression peaks prior to 94% of the genes in the Cdc15 cluster, implying that it potentially regulates most of the genes in the cluster. Gene *ace2*, belongs to the Eng1 cluster which contains genes that regulate cell separation. These genes peak after the Cdc15 cluster of genes.

Interaction between the proteins of *cdc18* and *mik1* are well-known [52]. Furthermore, according to the Human Protein-Protein Interaction Prediction software [53, 54], the proteins *cdc18* and *mik1* are highly interactive. The probability that they interact with each other is 17.80 times the probability that they do not. Thus our method not only validates some of the well-known relationships and interactions but also provides the direction of the interaction, suggesting that possibly one gene regulates the other which may lead to new hypotheses for biologists to investigate.

## Discussion

Often biological processes involve complex network of inter-relationships among the components of the process (e.g. genes). Biologists have been interested in deriving such networks and using them for drawing inferences regarding the underlying biological process. In the case of an oscillatory system, such as the cell-cycle or circadian clock, these networks are intrinsically dynamic in nature with the system going through different states or phases (e.g. phases in cell-cycle) over time before returning to the original state. At each state, due to the underlying biology, a subset of the components plays a prominent role. For example, only those genes that are involved in DNA synthesis are likely to express during the S-phase of the cell-cycle and the others may not. However, once S-phase is completed, the next wave of genes that are involved in the G2 phase express, and so on. It is of interest for biologist to understand the temporal order of how genes regulate each other as the cell goes through various phases. Thus, in an oscillatory system it is of interest to determine the temporal order among the components. Because of the structure of oscillatory system, underlying statistical parameters of interest (e.g. phase angles of cell-cycle genes) are points on a unit circle and not the entire Euclidean space. Focus of this research is to determine the temporal order with confidence and to compare the temporal orders among various study groups. Because of the underlying geometry of the circle, standard Euclidean space based methods are not suitable and until [4] there did not exist any rigorous statistical framework to analyze such data. Although [4] take important first step towards this problem, their methodology cannot be used to estimate the underlying order among the components. Secondly, their methodology does not allow a researcher to simultaneously test for the equality of the order among 3 or more populations. Lastly, when comparing two

populations, their methodology assumes that the order of expression among the components of one of the populations is known with certainty, an unreasonable assumption in practice. We not only overcome the above deficiencies of [4] but we also provide a novel method to estimate the common temporal order among a set of oscillatory genes across multiple populations, along with the associated confidence coefficient. Using the proposed methodology we successfully demonstrated that the temporal order of 6 cell-cycle genes is conserved in the two species of yeast and the humans. The proposed methodology can potentially be extended to develop dynamic networks for oscillatory systems where a biologist may be interested in not only inferring gene networks at a given time point but draw inferences across time points.

## Supporting Information

### S1 File. Angular mean versus Arithmetic mean for circular data.

(PDF)

### S2 File. Unconstrained estimates of phase angles and the concentration parameters.

(PDF)

### S3 File. Operating characteristics of the test statistic.

(PDF)

### S4 File. Gene forward selection procedure and results.

(PDF)

## Acknowledgments

The authors thank Drs. Xinping Cui and Delong Liu for careful reading of the manuscript and their helpful comments.

This work was supported by Spanish Ministerio de Ciencia e Innovación grant (MTM2012-37129 to S.B, M.A.F and C.R) and Junta de Castilla y León, Consejería de Educación and the European Social Fund within the Programa Operativo Castilla y León 2007–2013 (to S.B.) and the Intramural Research Program of the National Institute of Environmental Health Sciences (Z01 ES101744-04 to S.D.P.).

## Author Contributions

Conceived and designed the experiments: CR MAF SDP. Performed the experiments: SB. Analyzed the data: SB. Contributed reagents/materials/analysis tools: SB. Wrote the paper: SB CR MAF SDP.

## References

1. Cermakian N, Lamont EW, Bourdeau P, Boivin DB. 2011. Circadian clock gene expression in brain regions of alzheimer's disease patients and control subjects. *J. Biol. Rhythms* 26:160–170. PMID: [21454296](#)
2. Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, et al. 2009. Harmonics of circadian gene transcription in mammals. *PLoS Genetics* 5(4): e1000442. doi: [10.1371/journal.pgen.1000442](#) PMID: [19343201](#)
3. Kondratova AA, Kondratov R. 2012. The circadian clock and pathology of the ageing brain. *Nature Reviews Neuroscience* 13(5):325–335. doi: [10.1038/nrn3208](#) PMID: [22395806](#)
4. Fernández MA, Rueda C, Peddada SD. 2012. Identification of a core set of signature cell-cycle genes whose relative order of time to peak expression is conserved across species. *Nucl. Acids Res.* 40(7):2823–2832. doi: [10.1093/nar/gkr1077](#) PMID: [22135306](#)
5. Jensen JL, Jensen TS, Lichtenberg U, Brunak S, Bork P. 2006. Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature* 443:594–597. PMID: [17006448](#)

6. Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, et al. 2005. The cell-cycle-regulated genes of *Schizosaccharomyces pombe*. *PLoS Biology* 3:1239–1260. doi: [10.1371/journal.pbio.0030225](https://doi.org/10.1371/journal.pbio.0030225)
7. Peng X, Karutury RKM, Miller LD, Kui L, Yonghui J, et al. 2005. Identification of cell-cycle-regulated genes in fission yeast. *Mol. Biol. Cell* 16:1026–1042. doi: [10.1091/mbc.E04-04-0299](https://doi.org/10.1091/mbc.E04-04-0299) PMID: [15616197](https://pubmed.ncbi.nlm.nih.gov/15616197/)
8. Rueda C, Fernández MA, Peddada SD. 2009. Estimation of parameters subject to order restrictions on a circle with application to estimation of phase angles of cell-cycle genes. *J. Am. Stat. Assoc.* 104(485):338–347. doi: [10.1198/jasa.2009.0120](https://doi.org/10.1198/jasa.2009.0120) PMID: [19750145](https://pubmed.ncbi.nlm.nih.gov/19750145/)
9. Rustici G, Mata J, Kivinen K, Lió P, Penkett CJ, et al. 2004. Periodic gene expression order of the fission yeast cell-cycle. *Nature Genetics* 36:809–817. doi: [10.1038/ng1377](https://doi.org/10.1038/ng1377) PMID: [15195092](https://pubmed.ncbi.nlm.nih.gov/15195092/)
10. Xiao E, Xia-Zhang L, Barth A, Zhu J, Ferin M. 1998. Stress and menstrual cycle: Relevance of cycle Quality in the short- and long-term response to a 5-day endotoxin challenge during the follicular phase in the rhesus monkey. *J. Clin. Endocrinol. Metab.* 88:2454–2460. doi: [10.1210/jcem.83.7.4926](https://doi.org/10.1210/jcem.83.7.4926)
11. Slavov N, Airoidi EM, van Oudenaarden A, Botstein D. 2012. A conserved cell growth cycle can account for the environmental stress responses of divergent eukaryotes. *Mol. Biol. Cell* 23:1986–1997. doi: [10.1091/mbc.E11-11-0961](https://doi.org/10.1091/mbc.E11-11-0961) PMID: [22456505](https://pubmed.ncbi.nlm.nih.gov/22456505/)
12. De Quadros-Wander S, Stokes M. 2007. The effect of mood on opposite-sex judgments of males commitment and females sexual content. *Evol. Psychol.* 4:453–475.
13. Russell JA. 1980. A circumplex model of affect. *J. Pers. Soc. Psychol.* 39(6):1161–1178. doi: [10.1037/h0077714](https://doi.org/10.1037/h0077714)
14. Baayen C, Klugkist IG, Mechsner F. 2012. A test for the analysis of order constrained hypotheses for circular data. *J. Mot. Behav.* 44(5):351–363. doi: [10.1080/00222895.2012.709549](https://doi.org/10.1080/00222895.2012.709549) PMID: [22974062](https://pubmed.ncbi.nlm.nih.gov/22974062/)
15. Hastings MH, Reddy AB, Maywood ES. 2003. A clockwork web: circadian timing in brain and periphery, in health and disease. *Nat. Rev. Neurosci.* 4:649–661. doi: [10.1038/nrn1177](https://doi.org/10.1038/nrn1177) PMID: [12894240](https://pubmed.ncbi.nlm.nih.gov/12894240/)
16. Moller-Levet CS, Archer SN, Bucca G, Laing EE, et al. 2013. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc. Natl. Acad. Sci. USA* 110(12): 1132–1141. doi: [10.1073/pnas.1217154110](https://doi.org/10.1073/pnas.1217154110)
17. Liu D, Peddada SD, Li L, Weinberg CR. 2006. Phase analysis of circadian-related genes in two tissues. *BMC Bioinformatics*, 7:87. doi: [10.1186/1471-2105-7-87](https://doi.org/10.1186/1471-2105-7-87) PMID: [16504088](https://pubmed.ncbi.nlm.nih.gov/16504088/)
18. Storch KF, Lipan O, Leykin I, Viswanathan N, Davis FC, et al. 2002. Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417: 78–83. doi: [10.1038/nature744](https://doi.org/10.1038/nature744) PMID: [11967526](https://pubmed.ncbi.nlm.nih.gov/11967526/)
19. Caretta-Cartozo C, de los Rios P, Piazza F, Lio P. 2007. Bottleneck genes and community structure in the cell-cycle network of *S. pombe*. *PLoS Comput. Biol.* 3:968–976. doi: [10.1371/journal.pcbi.0030103](https://doi.org/10.1371/journal.pcbi.0030103)
20. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, et al. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*. 9(12):3273–3297. doi: [10.1091/mbc.9.12.3273](https://doi.org/10.1091/mbc.9.12.3273) PMID: [9843569](https://pubmed.ncbi.nlm.nih.gov/9843569/)
21. De Lichtenberg U, Wernersson R, Jensen TS, Nielsen HB, Fausbøll A, et al. 2005. Comparison of computational methods for the identification of cell cycle-regulated genes. *Bioinformatics* 21(7):1164–1171. doi: [10.1093/bioinformatics/bti093](https://doi.org/10.1093/bioinformatics/bti093) PMID: [15513999](https://pubmed.ncbi.nlm.nih.gov/15513999/)
22. Pihur V, Datta S, Datta S. 2007. Weighted rank aggregation of cluster validation measures: a monte carlo cross-entropy approach. *Bioinformatics* 23:1607–1615. doi: [10.1093/bioinformatics/btm158](https://doi.org/10.1093/bioinformatics/btm158) PMID: [17483500](https://pubmed.ncbi.nlm.nih.gov/17483500/)
23. Lehmann EL, Machné R, Georg J, Benary M, Axmann I, Steuer R. 2013. How cyanobacteria pose new problems to old methods: challenges in microarray time series analysis. *BMC Bioinformatics* 14:133. doi: [10.1186/1471-2105-14-133](https://doi.org/10.1186/1471-2105-14-133) PMID: [23601192](https://pubmed.ncbi.nlm.nih.gov/23601192/)
24. Fisher NI. 1993. *Statistical Analysis of Circular Data*. Cambridge University Press.
25. Mardia K, Jupp P. 2000. *Directional Statistics*. John Wiley & Sons, New York.
26. Liu D, Umbach DM, Peddada SD, Li L, Crockett PW, et al. 2004. A random periods model for expression of cell-cycle genes. *Proc. Natl. Acad. Sci. USA* 101(19):7240–7245. doi: [10.1073/pnas.0402285101](https://doi.org/10.1073/pnas.0402285101) PMID: [15123814](https://pubmed.ncbi.nlm.nih.gov/15123814/)
27. Bartholdi J, Tovey CA, Trick MA. 1989. Voting schemes for which it can be difficult to tell who won the election. *Soc. Choice Welf.* 6:157–165. doi: [10.1007/BF00303169](https://doi.org/10.1007/BF00303169)
28. Borda JC. 1781 *Memorie sur les elections au scrutin*. *Historie de l'Academie*.
29. Condorcet MJ. 1785. *Essai sur l'application de l'analyse a la probabilité des decisions rendues a la pluralité des voix*.
30. Diaconis P, Graham RL. 1977. Spearman's footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B* 39(2):262–268.

31. Schalekamp F, Zuyley A. 2009. Rank aggregation: Together we are strong. In Proc. of 11th ALENEX 38–51.
32. Karp RM. 1972. Complexity of Computer Computations. The IBM.
33. Papadimitriou CH, Steiglitz K. 1998. Combinatorial Optimization: Algorithms and Complexity. Dover Publications.
34. Hahsler M, Hornik K. 2011. Traveling Salesperson Problem (TSP). R package version 1.0-6. <http://CRAN.R-project.org>.
35. Lawler EL, Lenstra JK, Rinnooy Kann AHG, Shmoys DB. 1985. The Traveling Salesman Problem. John Wiley and Sons.
36. Reinelt G. 1994. The Traveling Salesman. Computational solutions for TSP applications. Springer-Verlag.
37. Chartrand G, Johns GL, Tian S, Winters SJ. 1993. Directed distance in digraphs: centers and medians. *J. Graph Theory.* 17(4):509–521. doi: [10.1002/jgt.3190170408](https://doi.org/10.1002/jgt.3190170408)
38. Dwork C, Kumar R, Naor M, Sivakumar D. 2001. Rank aggregation methods for the Web. *Proc. 10th International WWW Conf.* 613–622.
39. Bushel P, Heard NA, Gutman R, Liu L, Peddada SD, et al. 2009. Dissecting the fission yeast regulatory network reveals phase-specific control elements of its cell-cycle. *BMC Syst. Biol.* 3:93. doi: [10.1186/1752-0509-3-93](https://doi.org/10.1186/1752-0509-3-93) PMID: [19758441](https://pubmed.ncbi.nlm.nih.gov/19758441/)
40. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, et al. 2002. Identification of genes periodically expressed in the human cell-cycle and their expression in tumors. *Mol. Biol. Cell.* 13:1977–2000. doi: [10.1091/mbc.02-02-0030](https://doi.org/10.1091/mbc.02-02-0030) PMID: [12058064](https://pubmed.ncbi.nlm.nih.gov/12058064/)
41. Grosheva I, Shtutman M, Elbaum M, Bershadsky AD. 2001. p120 catenin affects cell motility via modulation of activity of Rho-family GTPases. A link between cell-cell contact formation and regulation of cell locomotion. *J. Cell Sci.* 114:695–707. PMID: [11171375](https://pubmed.ncbi.nlm.nih.gov/11171375/)
42. Sardet C, Vidal M, Cobrinik D, Geng Y, Onufryk C, et al. 1995. E2F-4 and E2F-5, two members of the E2F family, are expressed in the early phases of the cell-cycle. *Proc. Natl. Acad. Sci. USA* 92:2403–2407. doi: [10.1073/pnas.92.6.2403](https://doi.org/10.1073/pnas.92.6.2403) PMID: [7892279](https://pubmed.ncbi.nlm.nih.gov/7892279/)
43. Gauthier N, Larsen ME, Wernersson R, de Lichtenberg U, Jensen LJ, Brunak S, Jensen TS. 2008. Cyclebase.org—A comprehensive multi-organism online database of cell-cycle experiments. *Nucl. Ac. Res.* 36:854–859. doi: [10.1093/nar/gkm729](https://doi.org/10.1093/nar/gkm729)
44. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, et al. 1998. A genome-wide transcriptional analysis of the mitotic cell-cycle. *Mol. Cell.* 2(1):65–73. doi: [10.1016/S1097-2765\(00\)80114-8](https://doi.org/10.1016/S1097-2765(00)80114-8) PMID: [9702192](https://pubmed.ncbi.nlm.nih.gov/9702192/)
45. De Lichtenberg U, Wernersson R, Jensen TS, Nielsen HB, Fausbøll A, et al. 2005. New weakly expressed cell cycle-regulated genes in yeast. *Yeast* 22(5):1191–1201. doi: [10.1002/yea.1302](https://doi.org/10.1002/yea.1302) PMID: [16278933](https://pubmed.ncbi.nlm.nih.gov/16278933/)
46. Pramila T, Wu W, Miles S, Noble WS, Breeden LL. 2006. The forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle. *Genes Dev.* 22(16):2266–2278. doi: [10.1101/gad.1450606](https://doi.org/10.1101/gad.1450606)
47. Forsburg SL. 1999. The best yeast? *Trends Genet.*, 15:340–344. doi: [10.1016/S0168-9525\(99\)01798-9](https://doi.org/10.1016/S0168-9525(99)01798-9) PMID: [10461200](https://pubmed.ncbi.nlm.nih.gov/10461200/)
48. Aravind L, Watanabe H, Lipman DJ, Koonin EV. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl. Acad. Sci. USA* 97:11319–11324. doi: [10.1073/pnas.200346997](https://doi.org/10.1073/pnas.200346997) PMID: [11016957](https://pubmed.ncbi.nlm.nih.gov/11016957/)
49. Roux AE, Chartrand P, Ferbeyre G, Rokeach LA. 2010. Fission yeast and other yeasts as emergent models to unravel cellular aging in eukaryotes. *J. Gerontol. A. Biol. Sci. Med. Sci.* 65:1–8. doi: [10.1093/gerona/glp152](https://doi.org/10.1093/gerona/glp152) PMID: [19875745](https://pubmed.ncbi.nlm.nih.gov/19875745/)
50. Forsburg SL. 2005. The yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*: models for cell biology research. *Gravit. Space Biol. Bull.* 18:3–9. PMID: [16038088](https://pubmed.ncbi.nlm.nih.gov/16038088/)
51. Bähler J. 2005 Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet.* 39:69–94. doi: [10.1146/annurev.genet.39.110304.095808](https://doi.org/10.1146/annurev.genet.39.110304.095808) PMID: [16285853](https://pubmed.ncbi.nlm.nih.gov/16285853/)
52. Chu LH, Chen BS. 2008. Construction of a cancer-perturbed protein-protein interaction network for discovery of apoptosis drug targets. *BMC Syst. Biol.*, 2:56. doi: [10.1186/1752-0509-2-56](https://doi.org/10.1186/1752-0509-2-56) PMID: [18590547](https://pubmed.ncbi.nlm.nih.gov/18590547/)
53. McDowall MD, Scott and Barton, 2007 MS, Barton GJ. 2009. PIPs: Human protein-protein interactions prediction database. *Nucl. Acids Res.*, 37:D651–D656. doi: [10.1093/nar/gkn870](https://doi.org/10.1093/nar/gkn870) PMID: [18988626](https://pubmed.ncbi.nlm.nih.gov/18988626/)
54. Scott MS, Barton GJ. 2007. Probabilistic prediction and ranking of human protein-protein interactions. *BMC Bioinformatics* 2007 8:239–260. doi: [10.1186/1471-2105-8-239](https://doi.org/10.1186/1471-2105-8-239) PMID: [17615067](https://pubmed.ncbi.nlm.nih.gov/17615067/)