Ensemble Methods for Identifying RNA Operons and Regulons in the Clock Network of Neurospora crassa

AHMAD M. AL-OMARI 1, JAMES GRIFFITH 2, 3, ASHLEY SCRUSE 4, ROBERT W. ROBINSON 5, HEINZ-BERND SCHÜTTLER 6 AND JONATHAN ARNOLD 3

1 Biomedical Systems and Informatics Engineering, Yarmouk University, Irbid, Jordan 21163
2 College of Agricultural and Environmental Sciences, University of Georgia, Athens, GA USA 30602
3 Genetics Department, University of Georgia, Athens, GA USA 30602
4 Institute of Bioinformatics, University of Georgia, Athens, GA USA 30602
5 Computer Science Department, University of Georgia, Athens GA USA 30602
6 Department of Physics and Astronomy, University of Georgia, Athens, GA 30602, USA

Corresponding author: Ahmad M. Al-Omari (aomari@yu.edu.jo) and Jonathan Arnold (arnold@uga.edu)

Funding: This research was funded by the National Science Foundation (MCB-2041546). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ABSTRACT With both light entrainment data (L/D) and free running data in the dark (D/D), Variable Topology Ensemble Methods (VTENS) were used to reconstruct the entire genome-scale clock network involving 3380 genes and their products under the hypothesis that clock-controlled genes are both transcriptionally regulated and post-transcriptionally regulated through RNA operons. The clock network has 12 hubs with target genes surrounding twelve regulators, and 694 putative clock-controlled genes provide linkage between hubs in their joint regulation. There were 71 significant feedforward loops within this clock network, the majority of which involved genes lhp-1 and NCU06108 targeting genes connected with the ribosome and ribosome biogenesis. All 12 regulators target amino acid biosynthesis. The largest hub (768 genes) surrounded the post-transcriptional regulator lhp-1 defining an RNA operon. There appears to be no partitioning of light-regulated genes and circadian genes between the 12 hubs. The regulators lhp-1, has-1, and rrp-3 appear to target putative clock-controlled genes in ribosome biogenesis. The combined hypothesis of 6 RNA operons and 5 transcriptional regulons was successful in describing the dynamics of the clock network with 3380 genes and their products with a $\chi^2 = 2.133$ per data point for 60,759 mRNA time points.

INDEX TERMS variable topology ensemble methods, supernet, RNA operon, has-1, lhp-1, ribosome biogenesis, tRNA processing, light entrainment, circadian rhythms, feedforward loop, bioinformatics, parallel processing

I. INTRODUCTION One of the major challenges of systems biology is identifying genome-scale models of regulatory networks[1, 2]. For some complex traits, such as the biological clock, thousands of genes have been implicated[3]. The challenges of identifying such networks are several fold. While omic-scale methods exist for capturing RNA or protein profiles, the experimental methods are limited by a few noisy samples in time[4, 5]. Only recently have omic measurements been possible in vivo in real time[6]. The result is that there is limited data, but the networks and their parameter space of reaction rates and initial conditions are quite large[3]. Ensemble methods were developed to address this problem[7-9]. Unfortunately these ensemble
methods are quite computationally expensive as the network grows to a genome-scale, and only recently have successful parallelization strategies emerged for these ensemble methods[10, 11]. A third challenge is that the topology of the network of a complex trait is usually unknown. To overcome this challenge a new class of variable topology ensemble methods (VTENS) was developed using the concept of the supernet to identify simultaneously the network topology as well as the rate coefficients and initial conditions of a genome-scale genetic network[3]. The VTENS method has been shown to scale to thousands of genes and their products[1].

One of the central problems in systems biology is to model and understand the network controlling the biological clock, a complex phenotype shared by most organisms[12]. This network involving thousands of genes affects metabolism[13], development[14], and aging[15]. Much of what we know at the molecular level about the clock comes from studies in the model fungal system, Neurospora crassa, because the major clock genes shared by several phyla occur in one copy in this system[16-18]. As a consequence it has been possible to elucidate the three basic properties of a biological clock at the molecular genetic level: (1) its circadian rhythm[17]; (2) its light entrainment to the light dark cycle imposed[16], and (3) its temperature compensation, the ability to maintain a constant ticking of the clock under varied temperatures[18]. In this paper for the first time we examine on a global scale a model for the first two properties of the biological clock.

One of the major contributions of N. crassa to the understanding of all living systems is the linkage between genes and metabolism[19] through the One gene-One Enzyme Hypothesis. Metabolism is the means by which an organism can rapidly tune its response in minutes or seconds to its environment[6]. A substantial portion of metabolism appears to be under clock control[1, 13, 14]. It is still an open question of how this linkage is regulated on a global scale. An examination of the clock network under different light conditions may help to elucidate further this linkage to metabolism. Another related question is whether or not the regulation by the clock of metabolism is largely transcriptional or post-transcriptional[1, 13]. Some have hypothesized that post-transcriptional mechanisms may exist through RNA operons[20] as flexible means of gene regulation in the clock[21]. Here we begin to test this hypothesis as an alternative to transcriptional regulation when there is both data from experiments on a free running clock and one subject to light entrainment. When light is included as an environmental factor, a total of 3380 genes[14] among the 11,000 genes[22] or 31% of the genes in N. crassa appear to be involved in the clock network[14].

The questions to be addressed here are: (1) what does the clock network of 3380 genes that fits available circadian and light entrainment data at the molecular level look like; (2) what is the relative importance of RNA operons in the clock network; (3) are there particular regulatory modules associated with light entrainment; (4) what are the predicted targets of the regulators in identified RNA operons involved in light entrainment vs. circadian rhythms?

II. BACKGROUND

Since the Nobel Prize winning work of Beadle and Tatum in the model filamentous fungus, N. crassa[19], genes and enzymes have been linked together with the One gene-One Enzyme Hypothesis. By culturing metabolic mutants on minimal media versus rich media, genes were assigned quickly to various enzymatic steps, and in this way Beadle and Tatum were able to map the genes involved in aromatic amino synthesis and many other pathways[23]. By 1955 Nicholson had compiled an entire map of the metabolome, and there was some concern that the demise of biochemistry[24] was imminent. Koshland[24] argued this impressive map only marked the beginning of the subject. While an enormous amount of information had been assembled in a metabolic map, later efforts found that this detailed map may only cover a small percentage of metabolism[25-27]; moreover, what was missing from this map was a detailed understanding of metabolism’s regulation by genes[24, 25]. The maps of metabolism continue to grow even today to address problems in development, health, aging, and the engineering of human tissues[28].

Out of mapping the metabolome came the reconstruction and analysis of large networks[29, 30] and how to describe flux through them through steady state assumptions[31] as well as database systems, such as BIOCYC[32] for maintaining metabolic maps. It also led to the initial development of systems biology with its focus on the theoretical analysis of pathways[29, 33] and their flux balance analysis under steady state assumptions when considering large metabolic networks. One of the challenges to understanding these large metabolic networks and their regulation is the limited amount of data available on them – many proteins but minimal information about each component over time.

Ensemble Methods. The use of ensemble methods was developed in statistical physics by Boltzmann[34] and has recently appeared in applications outside of physics[9]. Ensemble methods in systems biology were foreshadowed by work using Monte Carlo Methods to understand the dynamics of metabolic models[35]. The first application in systems biology was to understanding quinic acid metabolism as a sole carbon source and its regulation in N. crassa[7] in 2002. A little later Sethna and colleagues suggested the use of ensemble methods as a way to handle “sloppy models”[36]. As in physics, ensemble methods addressed the problem of limited data (at the molecular
level) on genes and their function with the many parameters describing gene regulation. The solution, as first suggested by Boltzmann, was to give up on finding one best model but instead to identify an ensemble of models for prediction. The ensemble was selected to be consistent with the data available on the system. In contrast to flux balance analysis, the focus was on the dynamics of the network. As an example ensemble methods were used for the first time to provide a detailed mechanism for the biological clock and its circadian oscillations in N. crassa[37]. In these examples there were a limited number of samples over time on molecular species constituting the clock mechanism, such as clock messenger RNAs and proteins, and a Monte Carlo Experiment was performed to update model parameters randomly to improve the fit to the limited available data. Typically, 40,000 sweeps (a visit on average to each model parameter once) was used to move into a region of the parameter space supported by the data and another 40,000 sweeps, to construct the model ensemble with which to make predictions about clock dynamics[37]. With a model having 42 parameters, there were then \(3.4\) million proposed solutions to construct the ensemble for the clock network. These ensemble methods have come to play a major role in signaling systems beyond the clock [38], but their major limitation is that they are computationally expensive. They could be applied to smaller dynamical systems, such as the \(q\alpha\) gene cluster[7], the biological clock mechanism[39], other signaling systems[40], early development[41], population ecology[42], and global ecosystem modeling of vegetation[43], but what about larger networks, such as the metabolic map?

These ensemble methods have been taken one step further and fully integrated into the iterative design of experiments[14, 44-47]. In order to do model-guided discovery these ensemble methods were used to select the Maximally Informative Next Experiment (MINE). MINE as an ensemble method integrated the ensemble predictions with experimental data to design the next experiment in a series of experiments for the purposes of discovering the structure of both linear[48] and nonlinear models[14]. These new model-guided discovery tools are broadly applicable in science and engineering.

**Ensemble Methods for Large networks.** One way to overcome this limitation of computational cost is by parallelization. The most expensive part of the ensemble method is the construction of the \(3.4\) million solutions of, for example, a system of nonlinear ordinary differential equations (ODEs). Parallelization may be accomplished in several ways either by: (i) changing the method for solving ODEs[49-51]; (ii) using GPUs to carry the burden in solving the ODEs[10, 52, 53] in the Monte Carlo Experiment; (iii) using multiple agents in carrying out a search for good solutions[54]. All three approaches potentially allow ensemble methods to scale to networks involving thousands of genes and their products[55] (referred to as genetic networks hereafter) to understand the regulation of metabolism. In principle, it is then possible to use ensemble methods to understand and identify metabolism on a global scale and to identify genome scale models of its regulation.

**Variable topology ensemble methods.** While the topology of metabolic networks is generally known, the topology of a genetic network and its regulation is only known in a few special cases[7, 39]. The reason is that there are many kinds of mechanisms for the regulation of metabolic networks. Networks can be regulated transcriptionally and their topology, identified by chromatin immunoprecipitation experiments (ChIP) with chips(-chip) or sequencing (-SEQ) and related methods[56, 57], but they can also be regulated post-transcriptionally through RNA binding proteins[20] as well as by modifications of proteins[58]. The clock network through its pervasive effect on metabolism[3, 13] provides a unique doorway into understanding how genes are linked by regulation to metabolism. Variable topology ensemble methods were used previously to identify genetic networks of thousands of genes under clock control using variable topology ensemble methods using increasingly complex hypotheses about gene regulation[1, 3]. These methods can be applied to identify genome scale networks, such as the genetic network for the development of the echinoderm’s mesoderm[59] or whole genome networks in bacterial systems[2, 60].

**Analysis of network structure.** With the advent of identifying large network architectures both from network reconstructions and from experimental approaches, the question naturally arises as to their building blocks. Just as an integrated circuit is made up of components, such as capacitors, resistors, amplifiers, and inverters, large biochemical and genetic networks may be composed of building blocks or network motifs[61, 62]. These network motifs are patterns of interconnected nodes, such as gene A regulating gene B, gene B regulating gene C, and gene A regulating gene C, that occur repeatedly in a large network. A formal definition is given[63]. As in integrated circuits, they may offer clues as to how large networks function[64]. These network motifs have been found in a variety of regulatory networks, such as yeast[65] as well as many other large network structures[63].

**Filling in the gaps left by Beadle and Tatum, between metabolism and gene regulation.** Genes can be linked to metabolism[19]. Little is known about allosteric regulation of metabolism on a global scale. The biological clock by virtue of its pervasive effects on metabolism[1] could help to fill the gap left by Beadle and Tatum. To fill these gaps requires the ability to identify a very large clock genetic network of unknown topology under varied hypotheses about gene regulation (described in Section III).
Variable Topology Ensemble Methods and their parallelization provide the tools to chip away at the allosteric regulation of metabolism (described in Section IV).

III. MODEL

The model consists of two components, the topology of the genetic network with 12 regulatory genes (Figure 1A) and the genetic subnetwork of each putative clock-controlled gene (ccg), the targets of the regulators (Figure 1B). The core clock mechanism involving the positive regulatory elements (wc-1 and wc-2) and the negative regulatory element (frq) is taken as given and has been specified in previous work by ensemble methods applied to four microarray experiments[14]. The regulatory topology is structured into three layers, a master regulator (WCC=WC-1/WC-2), regulatory targets, which include both transcription factors and RNA binding proteins, and finally their nonregulatory targets, the nonregulatory putative ccgs. Many of these nonregulatory putative ccgs are involved in metabolism[3].

The regulators fall into two classes, transcription factors and RNA binding proteins. Each regulator (Regk) has a binding strength μk to a target ccg. Under the supernet strategy for specifying the topology, each regulator, Regks, is in principle allowed to bind to all of the available nonregulatory putative ccgs. As the ensemble method unfolds, these binding strengths are determined by Monte Carlo[1]. At the end of the fitting procedure only some of the connections between regulator and target are maintained. The binding strengths to the putative ccg determine the strength of the activation signal affecting the ccg. For the transcriptional regulators the activation signal S(t) is given by:

\[ S(t) = \mu_0 [\text{Reg}_0(t)]^m + r_0^m \sum_{k=1}^{5} \left( \frac{1}{n_k} \right)^m \mu_k [\text{Reg}_k(t)]^m \]  (1)

The concentrations [Reg\textsubscript{s}(t)] (k = 1, ..., 5) are for positive transcription factors (NCU07392, NCU01640, NCU06108, and NCU07155) and one repressor (NCU00045). The protein profiles of these transcription factors were captured experimentally by Westerns[1] with one exception. The WCC profile was determined by a series of 4 Maximally Informative Next Experiment (MINE) experiments including the protein profile of WC-1[14]. The binding strength \( \mu_0 \) is that of WCC, and \( \mu_k, k = 1, ..., 5 \), the binding strengths of the remaining transcriptional regulators, and all of the binding strengths to a putative ccg are constrained to be non-negative and sum to: \( \sum_{k=1}^{5} \mu_k = 1 \), with the exception of the negative regulator (NCU00045). This repressor is unconstrained in its binding strength \( \mu_4 \). The quantity \( r_0 \) is the time average concentration of the regulator Reg\textsubscript{0}(t).

In addition, there are 6 post-transcriptional regulators of 6 hypothesized RNA operons (Figure 1A). The protein profiles of two of these regulators (HAS-1 and LHP-1) were determined here by Westerns (See Materials and Methods). The remaining protein profiles were determined from their RNA profiles as described in a previous work[3]. The post-transcriptional signal of regulators \( S_p \) is given by:

\[ S_p(t) = \sum_{k=6}^{11} \left( \sum_{l=1}^{5} \mu_k [\text{Reg}_k(t)]^m \right) \]  (2)

The protein concentrations [Reg\textsubscript{k}(t)] of these 6 post-transcriptional regulators Reg\textsubscript{k}, k=6,...,11, additively determine the strength of the signal \( S_r(t) \).

All of the regulators, Reg\textsubscript{k}, k=1,...,11, are regulated by the master regulator WCC (Reg\textsubscript{0}) with a fixed binding constant \( \mu_0 = 1 \), and in turn there are putative regulatory ccgs targeted only by WCC[14].
With these signals (S(t) and S_p(t)) the subnetwork of each putative ccg is then regulated (Figure 1B). Boxes represent genes and their products. Circles represent the reactions. Each gene can be transcriptionally off (g_0 genes and their products. Circles represent the reactions. Each gene can be transcriptionally off (g_0) or on (g_1). In turn the gene can be transcribed at a rate S(t) into its mRNAs (g_2 and g_3), which in turn can be translated at a rate L(t) into a polypeptide (g_4). The transcriptional signal S(t) enters to activate a gene; the post-transcriptional signal(S_p(t)) acts to affect the stability of the cognate mRNA. All gene products have decay reactions, such as D_C for its mRNA. The diagram (Figure 1B)[66] specifies a system of ordinary differential equations (ODEs) presented previously in equation (1)[66].

The putative clock-controlled genes fall into three classes as experimentally determined from RNA profiling experiments[14]: (1667) genes that display a circadian rhythm; (974) genes that display a light entrainment response; (739) genes that display both a circadian rhythm and light response (Figure 1A). It is an open question how this phenotypic response, a circadian rhythm or light response, partitions across the total genetic network involving all 3380 genes in the clock network[1].

The only other component of the clock genetic network is the clock mechanism itself. The clock mechanism with its three genes (wc-1, wc-2, and frq) has been successfully identified using ensemble methods[14, 39] based on microarray experiments and other data in the literature from a total of 3007 measurements. The ensemble of models for the clock mechanism yielded a mean trajectory of WCC, which was used here to describe how WCC regulated each of the 11 regulators in the clock network (Figure 1A). Using the notation previously[39], the relevant concentrations of the transcriptionally active oscillator gene, mRNA, and protein are \( f_1, f_r, \) and \( f_p \), respectively; the relevant concentrations of the activator gene, unstabilized mRNA, stabilized mRNA, and protein are \( [WC - 1^{0r}] \), \( [WC - 1^{rp}] \), and \([WC - 1] \), respectively. An excellent approximation to the clock mechanism is given by the following system of 7 ODEs[39, 49] in which the protein concentration \([WC - 1] \) is taken to be large and constant:

\[
\begin{align*}
   f_1' &\equiv d[f r q^1]/dt = A(f_C - f_1) w^n - A' f_1 \\
   f_r' &\equiv d[f r q^r]/dt = S_3(f_2 - f_1) + S_4 f_1 - D_3 f_r \\
   f_p' &\equiv d[f r q^p]/dt = L_3 f_r - D_6 f_p \\
   u_{r0}' &\equiv d[w c - 1^{0r}]/dt = V_1 - D_1 u_{r0} - C_1 u_{r0} f_p \\
   u_{r1}' &\equiv d[w c - 1^{1r}]/dt = C_1 u_{r0} f_p - D_7 u_{r1} \\
\end{align*}
\]

\(\begin{align*}
   u_p' &\equiv d[W C - 1]/dt = L_1 u_{r1} - D_4 u_p - E_2 u_p - V p h o t(t) \\
   w_r' &\equiv d[W C C]/dt = E_2 u_p - D_6 w - nA(f_C - f_1) w^n + nA' f_1 - P w f_p^m + V p h o t(t)
\end{align*}\)

The last two equations in red introduce the light dependence. The photon concentration over time for light intensity \( I(t) \) is given by:

\[
V p h o t(t) \equiv C_3 I(t) * [W C - 1(t)], \quad \text{with} \quad C_3 = R_{23} * C_2.
\]

The protein [WC-1] provides the light response[16]. Application of the ensemble method to this system of ODEs was used to generate the mean (across the ensemble) of trajectories of [WCC] under varied light conditions for use in the genome-scale network (Figure 1). The values of the Hill Cooperativities \( n \) and \( m \) are set at \( n = m = 4 \) here.

IV. METHODS

**RNA profiling Data:** The RNA profiling data were described[14]. In the cycle 1 experiment characterizing the free running circadian rhythms of 2406 mRNAs, 13 RNA samples were collected every 4 hours over 48 h in the dark (D/D). In the cycle 2 light entrainment experiment the light response of 1713 mRNAs were found to vary in expression in both the light and dark (L/D)[14]. There were 17 time points over 48 hours with the first 7 time points taken in the dark, and the last 10 time points were taken in the light[14]. In addition to the time points every 4 h as in the D/D experiment, additional time points at 24 h + 20 m, 24 h + 40 m, 25 h and 26 h were taken near the switch from dark to light. RNA profiling data was obtained on all 11,000 genes using microarrays[14].

**Protein Extraction:** Protein extraction was carried out as previously described using a kit YT-105(Invent Biotechnologies, Inc. Eden Prairie, MN)[1].

**Antibody Production:** Polyclonal antibodies were produced to HAS-1 (encoded by NCU09349) and LHP-1 (NCU08295) with the PolyExpress Good Package 1649 by GenScript USA, Inc. (860 Centennial Ave, Pascataway, NJ 08854).

**Protein Profiling of RNA-binding proteins:** Western blots were used to produce protein profiles on two RNA binding proteins, HAS-1 and LHP-1, as well as a standard, HIS-3.

**Protein Gels:** Pre-cast “ExpressPlus PAGE Gels” from “GenScript” were run at 140 volts for approximately 70
minutes in Tris-MOPS (SDS) from “Genscript” as previously described[1].

Gel Transfer: Gels were electroblotted using the “GenScript” “eBlot” electroblotter as described previously[1]. Following this transfer, the Western membrane was treated with the “Genscript ONE-HOUR western Advanced Kit.”[1]

Western Blot: a Western blot was carried out in 6 steps described previously[1]. Bands were quantified using IMAGEJ[67], and each band was normalized with respect to HIS-3 band intensity.

Determining protein concentration of remaining post-transcriptional regulators: For the remaining 4 post-transcriptional regulators, protein concentrations were inferred from their RNA profiling data with a previously developed method[3]. The method is briefly recapitulated. Using the ODEs specified by Figure 2B, the protein trajectory of regulators Reg, were solved with initial condition \( r_k \). This initial condition was matched to the time average of \( \left[ Reg_k \right] \).

VTENS ensemble method: The VTENS has been described in detail[3] and is briefly recapitulated here. The ensemble of model parameters is calculated by Markov Chain Monte Carlo (MCMC)[7]. Model parameters fall into two classes: (1) \( \Theta \)-variables consisting of rate constants and initial conditions (Figure 1B); (2) binding strengths (\( \mu_k \)) specifying the topology of the network (Figure 1A). Each class of parameters has a distinct updating rule during MCMC. The \( \Theta \)-variables were updated with a traditional Metropolis-Hastings Algorithm with variable step-width[3, 68]. The binding strengths were updated with a measure preserving transformation so that the number of parameters did not vary in the MCMC run. During MCMC each update was sent to a GPU, where the system of 4,143 subnetworks each with 22 variables was solved. In the MCMC a total of 10,000 equilibration sweeps followed by 40,000 accumulation sweeps were performed, a sweep being a visit once on average to all parameters in the model. Wall time for the MCMC run was 61 days on 2 NVIDIA K40 processors (NVIDIA, Inc., Santa Clara, Ca).

An ensemble method with heterogeneous solver (ARK and GQ) on the GPU: The systems of ODEs specified by Figure 1B were integrated to give an exact integral solution formula[3]. This formula has two properties:

1) It was necessary to solve each of these systems of ODEs nearly 8 billion times during MCMC; therefore, it was necessary to develop a fast ODE implementing Gauss Quadrature integrator (GQ) with 100 integration points to solve the 4,143 ODE subnetworks (Figure 1B).
2) The ODE solutions are stiff (i.e., sudden change in the solution) near time point 0. To sidestep this problem Adaptive Runge-Kutta (ARK) was used near 0, and the GQ solver was used away from zero. The result was a heterogeneous solver[3].

The heterogeneous solver was implemented on a single K40 GPU (NVIDIA, Inc., Santa Clara, CA) in the equilibration phase and on two K40 GPUs in the accumulation phase, with one of the GPUs with 2418 threads and the other GPU with 1725 threads. The threads were divided into blocks of size 32 threads to implement the heterogeneous solver. To increase the speed of the heterogeneous solver on the GPUs, the solver used registers on the GPU to define variables. To cut down on excessive use of registers, all constant data, such as the interpolation files of \([\text{Reg}(t)]\) and ARK and GQ constants, were stored in constant memory on the GPU. The software was implemented with considerations in mind to overlap the GPU and CPU work to maximize the speed up. The software and data in this paper are stored in https://sourceforge.net/projects/identifying-rna-operons. Visualization of the network was accomplished in Cytoscape[69].

V. RESULTS

Fitting the genetic network ensemble fitting to D/D and L/D profiling data. A VTENS ensemble method was used to fit both the Dark/Dark(D/D) data on a free running clock as in previous work[3] as well as new Light/Dark (L/D) data in light entrainment experiments[14] (see Materials and Methods). A total of 4,143 gene profiles were successfully fitted on GPUs (see Materials and Methods) by a VTENS ensemble method both separately (D/D or L/D) and together with the respective chi-squared statistic per data point being 1.914, 2.368, and 2.133. The ensemble equilibrated quickly in less than 3000 sweeps, although equilibration was carried out to 10,000 sweeps[7] to insure that the ensemble was accurately identified (Figure 2). The wall time on the calculation was 61 days.
FIGURE 2. The $\chi^2$ goodness of fit statistics decrease and improve with sweep in the Markov Chain Monte Carlo run to achieve equilibration (see Materials and Methods).

FIGURE 3. The predictions of the genetic network with transcriptional control and 6 RNA operons (in purple, red, orange, and yellow) (averaged over the model ensemble) fitted microarray data over a 48 hour window with timepoints every at 0 h, 4 h, 8 h, 12 h, 16 h, 20 h, 24 h, 24 h + 20 m, 24 h + 40 m, 25 h, 26 h, 28 h, 32 h, 36 h, 40 h, 44 h, and 48 h on a L/D cycle of 24 h of dark followed by 24 h of light. Two views of the model ensemble fitted to the L/D data are presented: (A) side view; (B) overview. The black dots are the data from the L/D experiment. The surface is the prediction by the model ensemble.

To create a smooth surface for the fitted model ensemble the 1725 light responsive genes were first ordered by an average linkage clustering algorithm using Euclidean distance between the profiles of each gene centered by its mean of log-expression[70, 71]. In the overview (Panel A) the black stripes represent the time points sampled in the RNA profiling. The black cloud (Panel B) in the neighborhood of 24 h encompasses time points 24 h, 24 h + 20 m, 24 h + 40 m, 25 h, and 26 h. The measured log-expression levels blanketed the predicted surface with some scatter about the surface predicted by the ensemble.
If an average is taken over the predicted surface (from the ensemble) with respect to the genes, then the fit can be summarized in a 2-dimensional projection onto the time axis. For the D/D free running experiment the two circadian oscillations can be seen averaged over all genes in a 48 h window (Figure 4A). An examination of the L/D experiment revealed the effect of turning the light back on after 24 hours (Figure 4B). There was a rise in the average concentration of genes as they entrained to the light signal on a genome scale. The model made one excursion to a high concentration when not near a data point.

FIGURE 4. An average over the ensemble and genes (in blue) predicts the average trajectory of a gene (dots in red) under clock control in the genome in: (A) D/D experiment; (B) L/D experiment. Each graph is a projection onto the time axis of one of the panels in Figure 3.

In these experiments the Western profiles of HAS-1 and LHP-1 were measured to constrain further the role of the post-transcriptional regulators. Both of these regulators appeared to have a circadian rhythm both in their protein levels and in their RNA levels[14] (Figure 5). These observed protein levels were then used to regulate all of the genes assigned to the HAS-1 and LHP-1 RNA operons.

FIGURE 5. Western profiles of HAS-1 and LHP-1 used in VTENS model ensemble fitting.

All of the regulators were tested previously to be both circadian, light-responsive, and WCC-responsive, i.e., to be under WCC control[14]. A more detailed examination of the fit was obtained by examining the trajectories of the regulators (Figure 7). All of the regulators showed a circadian response both through the lense of the model ensemble and the data directly. The light entrainment response was more varied (Figure 8). The transcription factor rpn-4 (NCU01640) had a strong positive response to light, while the regulator NCU06108 had a strong negative response to light.
FIGURE 7. The model ensemble well predicts the RNA profile trajectories of the regulators for the D/D data. The blue curve is the ensemble average; the yellow and grey curves are the 95% confidence bands about the model average; and the red dots are the RNA profiles on the regulators. *lhp-1* (NCU08295) and *has-1* (NCU09349) are regulators Reg6 and Reg11, respectively.

FIGURE 8. The model ensemble well predicts the RNA trajectories of the regulators for the L/D data. The blue curve is the ensemble average; the yellow and grey curves are the 95% confidence bands about the model average; and the red dots are the RNA profiles on the regulators. The regulator NCU01640 is the *rpn-4* transcription factor.
FIGURE 9. Each family of targets of transcriptional and post-transcriptional regulators varies in size. (A) The number of targets for each regulator is reported for both the D/D and L/D experiments. (B) The number of targets of each regulator is further broken down in the D/D experiment fitted with the previous model ensemble[72], in the D/D experiment fitted with the current ensemble, in the L/D experiment fitted with the current ensemble for the L/D experiment, and the combined results of the D/D and L/D experiments fitted with the current ensemble. The regulators (12) are included in these counts in panel B.

There are a comparable number of targets in RNA operons and transcriptional regulons.

If one simply counts the number of target genes assigned to each regulator, the sizes of the transcriptional regulons and RNA operons also varied substantially based on both the D/D and L/D data (Figure 9-panel A). The smallest regulatory cluster was the rok-1 RNA operon with 54 predicted targets, while the largest RNA operon of lhp-1 was predicted to have 768 target genes. The inclusion of the D/D and L/D data together increased the identification of targets in each RNA operon from 850 genes to 1,419 genes[1]. This total of 1,419 targets was comparable to the 1973 targets of transcriptional regulators (Figure 9-Panel A). In fact the RNA operon of lhp-1 had more targets than any other regulator.

No regulator was exclusively controlling targets that are principally light-regulated (L/D) or circadian (D/D), although WCC as a known light responsive regulator has many light-regulated targets (Panel B). What is surprising is that LHP-1 protein had more targets (768) in the L/D entrainment experiment than any other regulator including WCC.

Transcriptional regulons and RNA operons are enriched for certain functional classes of genes.

The transcriptional regulons and RNA operons had distinct functional profiles. For example, the rpn-4 (NCU01640) regulon was predicted to be enriched for carbon metabolism, amino acid metabolism, glycerophospholipid metabolism, and oxidative phosphorylation genes (Table 1). Another transcriptional regulator pro-1 (NCU07392) had a very similar set of targets to rpn-4, but also was predicted to pick up target genes connected with the ribosome, ribosome biogenesis, meiosis, and recombination.

The RNA operons also had distinct functional profiles as well. For example, the lhp-1 (NCU08295) operon was predicted to consist of target genes with products involved in the ribosome, ribosome biogenesis, proteasome, protein export, and amino acid biosynthesis. The has-1 (NCU09349) RNA operon in contrast was predicted to involve target genes whose products are involved in not only ribosome biogenesis, but also meiosis, RNA transport, and the spliceosome [73].

| Function Regulators | NCU | NCU | NCU | NCU | NCU | NCU | NCU | NCU | NCU | NCU |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| RiboCenone         | 0.90 | 0.89 | 0.88 | 0.87 | 0.86 | 0.85 | 0.84 | 0.83 | 0.82 | 0.81 |
| RiboCenone in subclusters | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 |
| LHC                  | 0.43 | 0.42 | 0.41 | 0.40 | 0.39 | 0.38 | 0.37 | 0.36 | 0.35 | 0.34 |
| Proteasome          | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Carbon metabolism   | 0.91 | 0.90 | 0.89 | 0.88 | 0.87 | 0.86 | 0.85 | 0.84 | 0.83 | 0.82 |
| Glycolysis           | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Synthesis and degradation of amino bodies | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Biosynthesis of amino acids | 0.53 | 0.52 | 0.51 | 0.50 | 0.49 | 0.48 | 0.47 | 0.46 | 0.45 | 0.44 |
| Peptide metabolism  | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Oxidative phosphorylation | 0.45 | 0.44 | 0.43 | 0.42 | 0.41 | 0.40 | 0.39 | 0.38 | 0.37 | 0.36 |
| Peptic export        | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Homoglycan recombination | 0.40 | 0.39 | 0.38 | 0.37 | 0.36 | 0.35 | 0.34 | 0.33 | 0.32 | 0.31 |
| RNA transport        | 0.71 | 0.70 | 0.69 | 0.68 | 0.67 | 0.66 | 0.65 | 0.64 | 0.63 | 0.62 |
| O-Glycosylated amino acids | 0.43 | 0.42 | 0.41 | 0.40 | 0.39 | 0.38 | 0.37 | 0.36 | 0.35 | 0.34 |
| Endothelial cell      | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Splicing             | 0.47 | 0.46 | 0.45 | 0.44 | 0.43 | 0.42 | 0.41 | 0.40 | 0.39 | 0.38 |
| RNA polymerase       | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ribosomal repair     | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

The RNA helicase encoded by gene rrp-3 (NCU04504) had similar target genes to HAS-1 including genes involved with the ribosome, ribosome biogenesis, RNA transport as well as carbon metabolism and amino acid biosynthesis. All of the RNA operons were predicted to have some overlapping targets with one exception – gene gab-1 (NCU04799) encoding a poly-A binding protein was uniquely involved in targeting the genes homologous to those involved in the synthesis and degradation of ketone bodies.

The clock regulates diverse metabolic functions including ribosome biogenesis, carbon metabolism, RNA transport, and biosynthesis of amino acids.
The clock regulated diverse metabolic functions (Figure 10). No regulator had exclusive control of particular modules of metabolism. The number of clock-controlled genes as regulators ($Reg(t), k=0,...,11; k≠4$) involved in a particular metabolic function varied from a low of 3 regulators of genes homologous to those involved in synthesis and degradation of ketone bodies to a high of 11 regulators in amino acid biosynthesis. The effects of the clock network pervaded metabolism and reached into the control of the ribosome (6 regulators), ribosome biogenesis (9), meiosis (7), homologous recombination (6), cell cycle (5), and RNA transport (9), as examples. The close ties of the clock network to ribosome biogenesis have been reported previously[14].

The view of the clock network in toto with its 12 hubs

The entire clock network was visualized in Cytoscape[69] (Figure 11), capturing genes that were both circadian and light-regulated. The estimated binding strengths (1) and (2) were used to assign each regulator to a target gene to visualize the network edges (Figure 11). The 12 hubs were defined by the regulators. The lhp-1 RNA operon (in yellow) was the largest hub with shared connections to other hubs, such as NCU07155, NCU04504, NCU01640 ($rpm-4$), and NCU07392 ($pro-1$). Each of the regulators shared a total of 3380 targets visualized in this network. The degree distribution of the network was typical in having a long right tail with power law exponent (-1.4)[74]. The most connected node in the network was lhp-1 (NCU8295) with 768 connections. The least connected regulatory node was rok-1 (NCU00919) with 54 connections. The average number of connections for a node in this network was 2.23, but there were 2686 target genes assigned to one and only one regulator. This implies there are $694 = 3380 – 2686$ genes with at least 2 regulators assigned.

One of the characteristic motifs of this clock network are feedforward loops

Genetic networks like the one constructed here in Figure 11 may have network motifs as building blocks for...
function\[61\]. One of the simplest motifs found in fungal transcriptional networks is the feedforward loop (FFL)\[75\]. An FFL has one regulator (e.g., WCC) regulating a second regulator (e.g., Regs), and both of these regulators regulate a target gene (e.g., ccg). An FFL introduces the possibility of a delayed response in expression of a ccg by virtue of the target gene receiving dual inputs from two regulators that may need to reinforce each other in their activation of the ccg. The fitted network structure in Figure 1A was relatively simple – a FFL to a target clock-controlled gene must involve WCC and another activator Regs. There were only 649 target genes with at least 2 regulators.

To calculate FFLs an 11 x 11 matrix R with 0 diagonal is introduced that describes how the master regulator controls each regulatory ccg (Figure 1A) with the exception of the repressor (NCU00045). Leaving out self-regulation and the repressor, the matrix R has 1’s indicating activation by WCC in the first row and 0’s indicating no activation elsewhere. In the network the regulatory CCG proteins do not regulate each other. This R matrix summarizes in part the topology of the regulatory network. A second 11 x 3402 matrix T is also introduced that describes how the master regulator and each regulatory ccg targets each nonregulatory gene for activation (Figure 1A). This matrix is referred to as the Target matrix T. The row labels correspond to the regulators, and the columns, to the targets. A 1 in the T-matrix indicates regulation, and a 0 indicates no regulatory effect. The row labels of the Target Matrix T are in the same order as the R matrix. All of the feedforward loops (FFLs) can be identified by taking the Hadamard product of RT with T, i.e., \(RT \odot T\). The sum of the entries of \(RT \odot T\) is a count of the FFLs. If A regulates B and C, and B regulates C as an FFL, gene A must be the master regulator WCC, B, a ccg regulator, and C, a target nonregulatory gene. Entries of \(RT \odot T\) greater than 1 count alternative regulatory ccgs (Bs) of the same target gene C. An inner product of the first row of the T matrix with any of the successive 10 rows identifies the number of regulators that appear as a B gene in a feedforward loop. As a note, this method of counting FFLs applies to any R matrix with zero diagonal, not just this special case.

There were 71 feedforward loops in the clock network (Figure 11). As a basis of comparison the clock network is about the same size (3380 genes) as the E. coli transcriptional network (4000 genes), but the clock network had more feedforward loops (71 > 40)\[61\]. In that the probability of a FFL in a randomized network does not depend strongly on network size and since the E. coli and Sacharomyces cerevisiae transcriptional networks have fewer significant FFLs\[61\] than the clock network, the 71 FFLs in the clock network were significant and not likely to have arisen by chance. Of these 71 FFLs in the clock network of N. crassa, 5 of them involved shared network motifs through a different “B gene”. In other words there were 2 regulatory motifs assigned to each of the 5 target genes that only differ in the B gene but overlap in the master regulator A and target gene C.

The 5 target genes sharing 2 FFLs were mrp-180 (NCU08120), trax (NCU06059), emp-8 (NCU04728), NCU2213, and NCU08166. Three of the gene products were annotated as belonging to the mitochondrion (mrp-180, emp-8, and NCU08166). The protein MRP-180 is associated with the ribosome, and TRAX, with tRNA binding. Both EMP-8 and the product of NCU08166 are associated with glucose and calcium ion homeostasis, respectively. The products of NCU02213 and NCU08166 are annotated as membrane bound.

Both LHP-1 and the product of NCU6108 shared targets associated with the ribosome and ribosome biogenesis (Figure 10). Not surprisingly, they both regulated mrp-180 in overlapping feedforward loops. LHP-1 is thought to be involved in tRNA processing, and this is consistent with it targeting TRAX as well. These two regulators were involved in all 5 pairs of overlapping motifs. What is striking is that these two regulators that target the ribosome and ribosome biogenesis were found in 39 out of the 71 feedforward loops in the clock network.

VI. CONCLUSION

The ability to identify genetic networks involving post-transcriptional regulators is a challenging problem, where a variety of ensemble methods are needed\[76, 77\]. Using VTENS methods we have successfully identified the dynamics of 3380 genes involved in the clock network out of 11,000 genes in the genome of N. crassa as well as the unknown transcriptional and post-transcriptional regulatory structure of these 3380 genes (Figure 3) from both RNA and protein profiling data. In the process of using the VTENS method we were able to identify both the dynamics and topology of the network as well as how each gene in the genetic network was regulated, i.e., the network topology\[1, 3\] (Table 1). To carry out this task it was necessary to use GPUs to parallelize the Monte Carlo method for ensemble model identification\[1\]. Other dynamic genetic networks of comparable size identified to varying extents are the genetic network of the pathogenic bacterium\[2\], Mycoplasma genitalium, and the sea urchins’ developmental network of the mesoderm reconstructed by the Davidson laboratory\[59\].

Post-transcriptional mechanisms can operate by a variety of mechanisms\[78\], including affecting: i) mRNA stability; ii) translation initiation; iii) transcriptional elongation; and iv) phosphorylation and dephosphorylation of network components\[58\]. The regulatory mechanism emphasized here is mRNA stability (Figure 1). The importance of these mechanisms is rapidly adjusting gene expression in a
changing environment, such as the daily light/dark cycle of the planet. RNA operons have been hypothesized to play a key role in clock regulation[21]. One of these is the large RNA operon for *lhp-1* (Figure 9) first documented in *S. cerevisiae*[79]. This particular RNA operon appeared to be the largest regulatory module in the clock network (Figure 9).

The widespread importance of post-transcriptional mechanisms in yeast was first demonstrated in *S. cerevisiae* by comparing protein profiles of gene products to the corresponding mRNA profiles[5]. A more recent analysis of mRNA and protein levels lend support to the hypothesis of post-transcriptional response by “translating proteins on demand” in different cellular compartments[80]. Direct evidence for RNA operons included work on *lhp-1*[79] and the Puf family of RNA-binding proteins[81] in *S. cerevisiae*. In addition to transcriptional control here we used successfully the post-transcriptional mechanism of an RNA operon to explain the dynamic behavior of the clock network in *N. crassa* (Figure 3). Goodness of fit (Figure 4) was 1.91 (D/D experiment) and 2.37 (L/D experiment) per data point and outperforms previously published work[14, 39].

Consideration of both the free running circadian rhythm data and the light entrainment data shifts the view on the importance of RNA operons in the regulation of the clock network[21]. Putative regulators, such as *lhp-1*, are as important as any of the previously identified transcriptional regulators including the master regulator WCC (Figure 9). Some of these post-transcriptional regulators appeared to control more target genes than other transcriptional regulators.

The circadian rhythm and light entrainment properties of the biological clock were not decomposable across regulatory hubs (Figure 9B). The light entrainment response, for example, involved genes in each transcriptional regulon and each RNA operon (Figure 9B). The two responses of a clock and its entrainment were linked together by involving genes in all hubs (Figure 12). Also what is striking is that the regulatory modules were also all linked together. While there appeared to be 2686 genes in the clock network with only one regulator, all of the hubs shared regulation between some of the 694 genes. The clock network does not constitute independent regulator modules, but is tied together in all likelihood to coordinate and propagate its effects[5]. There were 71 feed forward loops within the interconnected clock network[58, 82] that may propagate the clock’s effect into a number of processes within a cell including the ribosome and tRNA binding (Figure 10). The majority of these feedforward loops (39 out of 71) involved *lhp-1* and NCU06108 connected with the ribosome and ribosome biogenesis.

In that there is no demarcation of circadian and light-regulated genes in Figure 11, this raises the question of what constitutes a clock-controlled gene. The strict definition used previously is that the gene is: (1) circadian; (2) light responsive; and (3) WCC regulated[14]. A more inclusive definition might be any gene that is in the clock network (Figure 11). Such genes have a binding strength assigned to quantify how strongly they are regulated in the network. They are associated with other genes that have varying degrees of circadian or light-responsive expression depending on their physical properties, such as their mRNA or protein stability. This inclusive definition of a clock-controlled gene has been used here throughout.

There is yet to be explored one further post-transcriptional regulatory mechanism, clock control of ribosome biogenesis through, for example, HAS-1. The HAS-1 protein in *S. cerevisiae* is the only DEAD box helicase required for both 40S and 60S ribosome subunit biogenesis[83]. In that ribosome biogenesis appears to be under clock control in *N. crassa*[14] and in mammals[84], this raises the possibility that the clock can modulate the cell’s ribosome concentration and thereby indirectly modulate the translation rate of any other protein, including other regulators. The hypothesis of the RNA operon and the clock control of ribosome biogenesis (Figure 10) together may serve to explain the widespread effects of the clock network on 3380 genes and associated products.

The clock network exerts its role over many aspects of metabolism[1, 13]. In addition to its role in ribosome biogenesis, *lhp-1* appeared to play a role in carbon metabolism, glycerophospholipid metabolism, and purine metabolism (Table 1). Another example was *rpm-4* regulating carbon metabolism, valine, leucine, and isoleucine biosynthesis, glycerophospholipid metabolism, amino acid biosynthesis, and oxidative phosphorylation (Table 1). It was typical that *lhp-1* and *rpm-4* have overlapping regulatory profiles (Figure 12). All twelve of the regulators shared in their control of metabolism and in particular the biosynthesis of amino acids (Figure 10). The first linkage of genes and metabolism occurred in *N. crassa*[19], but the missing component even now is how genes regulate metabolism. The new VTENS approach here combining transcriptomics with the new approach of real time *in vivo* metabolomics[6] will allow fresh insights into the regulation of metabolism on a genomic scale.

**ACKNOWLEDGMENT**

This research was funded by the National Science Foundation (MCB-2041546). We thank the Georgia Advanced Computing Resource Center for computing.

**REFERENCES**

[1] A. Al-Omari, J. Griffith, C. Caranica, T. Taha, H. Schütter, and J. Arnold, "Discovering Regulators in Post-Transcriptional Control of the Biological
Clock of Neurospora crassa Using Variable Topology Ensemble Methods on GPUs," IEEE Access, vol. 6, pp. 54582-54594, 2018, doi: 10.1109/ACCESS.2018.2871876.

Jonathan R. Karr et al., "A Whole-Cell Computational Model Predicts Phenotype from Genotype," Cell, vol. 150, no. 2, pp. 389-401, 2012/07/20/ 2012, doi: https://doi.org/10.1016/j.cell.2012.05.044.

A. Al-Omari, J. Griffith, M. Judge, T. Taha, J. Arnold, and H. Schuttler, "Discovering Regulatory Network Topologies Using Ensemble Methods on GPGPUs With Special Reference to the Biological Clock of Neurospora crassa," IEEE Access, vol. 3, pp. 27-42, 2015.

J. L. DeRisi, V. R. Iyer, and P. O. Brown, "Exploring the metabolic and genetic control of gene expression on a genomic scale," Science, vol. 278, no. 5338, pp. 680-686, Oct 24 1997. [Online]. Available: <Go to ISI>://A1997YCY32300053.

T. Ideker et al., "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network," Science, vol. 292, no. 5518, pp. 929-934, 2001.

M. T. Judge et al., "Continuous in vivo Metabolism by NMR," Frontiers in Molecular Biosciences, 10.3389/fmolv.2019.00026 vol. 6, p. 26, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmolv.2019.00026.

D. Battogtokh, D. K. Asch, M. E. Case, J. Arnold, and H. B. Schuttler, "An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of Neurospora crassa," Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. 26, pp. 16904-9, Dec 24 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12477937

E. O. Voit, "The best models of metabolism," WIREs Systems Biology and Medicine, https://doi.org/10.1002/wsbm.1391 vol. 9, no. 6, p. e1391, 2017/11/01 2017, doi: https://doi.org/10.1002/wsbm.1391.

D. P. Landau and K. Binder, "Monte Carlo simulations at the periphery of physics and beyond," A Guide to Monte Carlo Simulations in Statistical Physics; Cambridge University Press: New York, NY, USA, pp. 13-22, 2014.

A. Al-Omari, J. Arnold, T. Taha, and H. Schuttler, "Solving Large Nonlinear Systems of First-Order Ordinary Differential Equations with Hierarchical Structure Using Multi-GPGPUs and an Adaptive Runge Kutta ODE Solver," IEEE Access, vol. 1, pp. 770-777, 2013.

E. Zamora-Sillero, M. Hafner, A. Ibig, J. Stelling, and A. Wagner, "Efficient characterization of high-dimensional parameter spaces for systems biology," BMC Systems Biology; vol. 5, no. 1, p. 142, 2011/09/15 2011, doi: 10.1186/1752-0509-5-142.

J. C. Dunlap, "Molecular bases for circadian clocks," Cell, vol. 96, no. 2, pp. 271-90, Jan 22 1999. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9988221

J. M. Hurley et al., "Analysis of clock-regulated genes in Neurospora reveals widespread posttranscriptional control of metabolic potential," Proceedings of the National Academy of Sciences, vol. 111, no. 48, pp. 16995-17002, 2014.

W. Dong et al., "Systems biology of the clock in Neurospora crassa," PLoS One, vol. 3, no. 8, p. e3105, 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18769678

M. Judge, J. Griffith, and J. Arnold, "Aging and the biological clock," Healthy Aging and Longevity, vol. Circadian Rhythms and Their Impact on Aging, pp. 211-234, 2017.

M. Judge, J. Griffith, and J. Arnold, "Aging and the biological clock," Healthy Aging and Longevity, vol. Circadian Rhythms and Their Impact on Aging, pp. 211-234, 2017.

S. K. Crosthwaite, J. C. Dunlap, and J. J. Loros, "Neurospora wc-1 and wc-2: transcription, photoresponses, and the origins of circadian rhythmicity," Science, vol. 276, no. 5313, pp. 763-9, May 2 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9115195

B. D. Aronson, K. A. Johnson, J. J. Loros, and J. C. Dunlap, "Negative feedback defining a circadian clock: autoregulation of the clock gene frequency," Science, vol. 263, no. 5153, pp. 1578-84, Mar 18 1994. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8128244

Y. Liu, N. Y. Garceau, J. J. Loros, and J. C. Dunlap, "Thermally regulated translational control of FRQ mediates aspects of temperature responses in the neurospora circadian clock," Cell, vol. 89, no. 3, pp. 477-86, May 2 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9150147

G. W. Beadle and E. L. Tatum, "Genetic Control of Biochemical Reactions in Neurospora," Proceedings of the National Academy of Sciences of the United States of America, vol. 27, no. 11, pp. 499-506, Nov 15 1941. [Online]. Available:
[43] L. Tang et al., "Optimization of Multi-Ecosystem Model Ensembles to Simulate Vegetation Growth at the Global Scale," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 962-978, 2021, doi: 10.1109/TGRS.2020.2993641.

[44] J. C. W. Locke et al., "Extension of a genetic network model by iterative experimentation and mathematical analysis," *Molecular Systems Biology*. https://doi.org/10.1038/msb4100018 vol. 1, no. 1, p. 2005.0013, 2005/01/01 2005, doi: https://doi.org/10.1038/msb4100018.

[45] R. L. McGee and G. T. Buzzard, "Maximally informative next experiments for nonlinear models," *Mathematical Biosciences*, vol. 302, pp. 1-8, 2018/08/01/ 2018, doi: https://doi.org/10.1016/j.mbs.2018.04.007.

[46] J. P. Townsend and F. Lopez-Giraldez, "Optimal Selection of Gene and Ingroup Taxon Sampling for Resolving Phylogenetic Relationships," *Systematic Biology*, vol. 59, no. 4, pp. 446-457, 2010, doi: 10.1093/sysbio/syq025.

[47] J. P. Townsend and C. Leuenberger, "Taxon Sampling and the Optimal Rates of Evolution for Phylogenetic Inference," *Systematic Biology*, vol. 60, no. 3, pp. 358-365, 2011, doi: 10.1093/sysbio/sys097.

[48] A. M. Bouffier, J. Arnold, and H. B. Schüttler, "A MINE Alternative to D-Optimal Designs for the Linear Model," *PloS one*, vol. 9, no. 10, p. e110234, 2014, doi: 10.1371/journal.pone.0110234.

[49] A. Al-Omari, H. B. Schüttler, J. Arnold, and T. Taha, "Solving Nonlinear Systems of First Order Ordinary Differential Equations Using a Galerkin Finite Element Method," *IEEE Access*, vol. 1, pp. 408-417, 2013, doi: 10.1109/ACCESS.2013.2269192.

[50] J. M. M. Ruiz, J. O. Lopera, and J. A. Carrillo, "Exploiting the multilevel parallelism and the problem structure in the numerical solution of stiff ODEs," in *Proceedings 10th Euromicro Workshop on Parallel, Distributed and Network-based Processing*, 9-11 Jan. 2002 2002, pp. 173-180, doi: 10.1109/EMPDP.2002.994262.

[51] B. R. C. Magalhães, M. Hines, T. Sterling, and F. Schüermann, "Exploiting Flow Graph of System of ODEs to Accelerate the Simulation of Biologically-Detailed Neural Networks," in 2019 *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 20-24 May 2019, pp. 176-187, doi: 10.1109/IPDPS.2019.00028.

[52] J. Alikhani, A. Massoudieh, and U. K. Bhowmik, "GPU-Accelerated Solution of Activated Sludge Model's System of ODEs with a High Degree of Stiffness," in 2016 *International Conference on Computational Science and Computational Intelligence (CSCI)*, 15-17 Dec. 2016 2016, pp. 555-560, doi: 10.1109/CSCI.2016.0111.

[53] F. Hegediůs, "Program package MPGOS: Challenges and solutions during the integration of a large number of independent ODE systems using GPUs," *Communications in Nonlinear Science and Numerical Simulation*, vol. 97, p. 105732, 2021/06/01 2021, doi: https://doi.org/10.1016/j.cnsns.2021.105732.

[54] C. Caranica, A. Al-Omari, H. B. Schüttler, and J. Arnold, "Identifying a stochastic clock network with light entrainment for single cells of Neurospora crassa," *Scientific Reports*, vol. 10, no. 1, p. 15168, 2020/09/16 2020, doi: 10.1038/s41598-020-72213-1.

[55] C. Caranica et al., "Ensemble methods for stochastic networks with special reference to the biological clock of Neurospora crassa," *PloS one*, vol. 13, no. 5, p. e0196435, 2018, doi: 10.1371/journal.pone.0196435.

[56] C. T. Harbison et al., "Transcriptional regulatory code of a eukaryotic genome," *Nature*, vol. 431, no. 7004, pp. 99-104, Sep 2 2004. [Online]. Available: <Go to ISI>:://000223641500050.

[57] B. Ren et al., "Genome-Wide Location and Function of DNA Binding Proteins," *Science*, vol. 290, no. 5500, pp. 2306-2309, 2000/12/22 2000, doi: 10.1126/science.290.5500.2306.

[58] M. Brunner and T. Schafmeier, "Transcriptional and post-transcriptional regulation of the circadian clock of cyanobacteria and Neurospora," *Genes & Development*, 10.1101/gad.1410406 vol. 20, no. 9, pp. 1061-1074, 2006, doi: 10.1101/gad.1410406.

[59] E. H. Davidson et al., "A genomic regulatory network for development," *Science*, vol. 295, no. 5560, pp. 1669-78, Mar 1 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11872831.

[60] N. Macklin Derek et al., "Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation."

[61] N. Macklin Derek et al., "Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation."

[62] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network Motifs: Simple Building Blocks of Complex Networks."

[63] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network Motifs: Simple Building Blocks of Complex Networks."

[64] R. Milo et al., "Superfamilies of Evolved and Designed Networks," *Science*, vol. 303, no. 5663, pp. 1538-1542, 2004/03/05 2004, doi: 10.1126/science.1089167.
[63] S. Yu, J. Xu, C. Zhang, F. Xia, Z. Almakhadmeh, and A. Tolba, "Motifs in Big Networks: Methods and Applications," IEEE Access, vol. 7, pp. 183322-183338, 2019, doi: 10.1109/ACCESS.2019.2960044.

[64] U. Alon, "Network motifs: theory and experimental approaches," Nature Reviews Genetics, vol. 8, no. 6, pp. 450-461, 2007/06/01 2007, doi: 10.1038/nrg2102.

[65] T. I. Lee et al., "Transcriptional regulatory networks in Saccharomyces cerevisiae," (in eng), Science, vol. 298, no. 5594, pp. 799-804, Oct 25 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12399584

[66] J. Arnold, T. R. Taha, and L. Deligiannidis, "GKIN: a tool for drawing genetic networks," Network Biology, vol. 2, no. 1, p. 26, 2012.

[67] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," Nature Methods, vol. 9, no. 7, pp. 671-675, 2012.

[68] J. Dai et al., "Exact Sample Size Needed to Detect Dependence in 2 × 2 × 2 Tables," Biometrics, vol. 63, no. 4, pp. 1245-1252, 2007, doi: 10.1111/j.1541-0420.2007.00801.x.

[69] P. Shannon et al., "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," Genome Research, 10.1101/gr.1239303 vol. 13, no. 11, pp. 2498-2504, 2003. [Online]. Available: http://genome.cshlp.org/content/13/11/2498.abstract

[70] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," Proceedings of the National Academy of Sciences, vol. 95, no. 25, p. 14863, 1998. [Online]. Available: http://www.pnas.org/content/95/25/14863.abstract

[71] M. J. L. de Hoon, S. Imoto, J. Nolan, and S. Miyano, "Open source clustering software," Bioinformatics, vol. 20, no. 9, pp. 1453-1454, 2004, doi: 10.1093/bioinformatics/bth078.

[72] A. Al-Omari, J. Griffith, C. Caranica, T. Haha, H. SchüTtler, and J. Arnold, "Discovering Regulators in Post-Transcriptional Control of the Biological Clock of SNeurospora—crassa Using Variable Topology Ensemble Methods on GPUs," IEEE Access, vol. 6, pp. 54582-54594, 2018, doi: 10.1109/ACCESS.2018.2871876.

[73] X. H. Liang and M. J. Fournier, "The helicase Has1p is required for snoRNA release from pre-rRNA," Molecular and cellular biology, vol. 26, no. 20, pp. 7437-50, Oct 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16908538

[74] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, "The large-scale organization of metabolic networks," Nature, vol. 407, no. 6804, pp. 651-654, 2000/10/01 2000, doi: 10.1038/35036627.

[75] I. Lee Tong et al., "Transcriptional Regulatory Networks in Saccharomyces cerevisiae," Science, vol. 298, no. 5594, pp. 799-804, 2002/10/25 2002, doi: 10.1126/science.1075090.

[76] R. Armañanzas, "Revealing post-transcriptional microRNA–mRNA regulations in Alzheimer’s disease through ensemble graphs," BMC Genomics, vol. 19, no. 7, p. 668, 2018/09/24 2018, doi: 10.1186/s12864-018-5025-y.

[77] R. Armañanzas, "Ensemble graphs to reveal post-transcriptional regulatory networks in Alzheimer’s disease," in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 13-16 Nov 2017 vol. 298, no. 5594, pp. 297-304, doi: 10.1109/BIBM.2017.8217667.

[78] E. Van Assche, S. Van Puyvelde, J. Vanderleyden, and H. P. Steenackers, "RNA-binding proteins involved in post-transcriptional regulation in bacteria," Frontiers in Microbiology, 10.3389/fmicb.2015.00141 vol. 6, p. 141, 2015. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmicb.2015.00141

[79] M. Inada and C. Guthrie, "Identification of Lhp1p-associated RNAs by microarray analysis in Saccharomyces cerevisiae reveals association with coding and noncoding RNAs," (in eng), Proceedings of the National Academy of Sciences of the United States of America, vol. 101, no. 2, pp. 434-9, Jan 13 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14704279

[80] A. Beyer, J. Hollunder, H.-P. Nasheuer, and T. Wilhelm, "Post-transcriptional Expression Regulation in the Yeast Saccharomyces cerevisiae on a Genomic Scale," Molecular & Cellular Proteomics, vol. 3, no. 11, pp. 1083-1092, 2004/11/01 2004, doi: https://doi.org/10.1074/mcp.M400099-MCP200

[81] A. P. Gerber, D. Herschlag, and P. O. Brown, "Extensive association of functionally and cyttopically related mRNAs with Puf family RNA-binding proteins in yeast," PLoS biology, vol. 2, no. 3, p. E79, Mar 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15024427

VOLUME XX, 2017
AHMAD M. AL-OMARI
Associate Professor (with Tenure) since Jan., 2020 and Erasmus+ certified trainer. Became the first Jordanian Nvidia Deep Learning Institute (DLI) Certified Instructor and DLI University Ambassador in April, 2020. He received the B.Sc. degree of electrical and computer engineering from Yarmouk University, Irbid, Jordan, in 2004. He was a teaching assistant and a lab engineer with the department of electrical and computer engineering, Yarmouk university, from 2004 to 2010. He was involved in more than 30 different projects on computers, communication, and electronic engineering. He received a grant of over $150,000 from Yarmouk University to pursue the Ph.D. degree of bioinformatics and in the same theme he received a Teaching Assistant grant of over $133,000 from The University of Georgia/USA, which is considered the oldest public university in the US and one of the biggest and the topmost ranked university as well. Dr. Al-OMari was a member of professor Jonathan Arnold's research group with the university of Georgia, Athens, GA, USA, on systems biology and genetic networks.

James Griffith received his B.S. in Animal and Dairy Science from the University of Georgia in 1984. Following several years in areas of bovine genetics, he joined the Genetics Department at the University of Georgia in 1989. He started out as a Research Technician, and is currently the lab manager and Research Professional for Dr. Jonathan Arnold. He has been heavily involved in the physical mapping of several fungi (Aspergillus nidulans, Aspergillus flavus and Neurospora crassa), and for the last seven years with longevity and the biological clock of Neurospora crassa. He has published over 20 publications in fungal genomics and systems biology.

Robert W. Robinson received his A.B. in Mathematics and his A.M. in Physics from Dartmouth College in 1963, followed by his Ph.D in Mathematics from Cornell University in 1966. Since then he has been an NSF Postdoctoral Fellow at the Institute for Advanced study in Princeton, an Assistant Professor of Mathematics at U. C. Berkeley and at U. Michigan, and a Professor of Mathematics at Newcastle U. (Australia) and at Southern Illinois U. (Carbondale). In 1984 he became the founding Head of the Computer Science Dept. at U. Georgia, where he retired as Emeritus Professor in 2010. He has authored or co-authored some 95 research papers in mathematical logic, combinatorics, graph theory, and applications.

Heinz-Bernd Schüttler received his Diplom in Physics from the Technische Universität München, Germany, in 1981; and his Ph.D. in Physics from the University of California Los Angeles in 1984. After post-doctoral appointments at the University of California Santa Barbara and at the Argonne National Laboratory, he joined the faculty of the Department of Physics and Astronomy at the University of Georgia in 1987 where he presently serves as a Professor of Physics. His research interests include applications of computational statistical mechanics in biology and condensed matter physics. His recent notable contributions in computational biology include the development of the ensemble network simulation (ENS) method for the reconstruction of biological circuits; and ENS-based methods for maximally informative next experiment (MINE) design.

Ashley Scruse received her B.S. in mathematical sciences from Clark Atlanta University in 2017. Her interest in mathematics and how it applies to biological sciences lead her to the University of Georgia’s Institute of Bioinformatics where she is a 5th year PhD candidate. In 2018 Ashley received the NIH T32 genomics training grant to help her merge her mathematical skillset with her interest in biological network motifs. Her research interest includes genetic regulatory networks and the discovery of network motifs using mathematical analysis.

[82] M. Judge, J. Griffith, and J. Arnold, "Aging and the biological clock," vol. Circadian Rhythms and Their Impact on Aging., pp. 211-234, 2017.

[83] B. Emery, J. de la Cruz, S. Rocak, O. Deloche, and P. Linder, "Has1p, a member of the DEAD-box family, is required for 40S ribosomal subunit biogenesis in Saccharomyces cerevisiae," Molecular microbiology, vol. 52, no. 1, pp. 141-58, Apr 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15049817

[84] C. Jouffe et al., "The circadian clock coordinates ribosome biogenesis," PLoS biology, vol. 11, no. 1, p. e1001455, 2013.
Jonathan Arnold received his B.S. in Mathematics from Yale University in 1975 and subsequently his M. Phil. and Ph.D. in Statistics from Yale University in 1978 and 1982, respectively. In 1982, he joined the Departments of Statistics and Genetics in the College of Arts and Sciences at the University of Georgia, where he has risen through the ranks to professor of Genetics, Statistics, and Physics & Astronomy. He has authored and co-authored over 140 papers in journals, book chapters, and conference proceedings in the subjects of statistical genetics, population genetics, computational biology, fungal genomics, genetics of aging, and systems biology. He has served on the NSF Computational Biology Panel, DOE Genome Panel, NSF Systems and Synthetic Biology Panel, and as a charter member on the NIH Genetic Variation and Evolution Study Section. His research interests include the development and identification of genetic networks of fundamental processes in the model system *Neurospora crassa*, i.e., Computing Life. He was elected an AAAS fellow in 2011.