

The Kullback-Leibler divergence between attractor reconstructions: a new means of distinguishing noisy biological time series

Emma Granqvist^{1,2*}, Saul Hazledine¹, Derin Wysham¹,
Giles E.D. Oldroyd² and Richard J. Morris¹

¹ Department of Computational and Systems Biology, John Innes Centre,
Norwich Research Park

² Department of Disease and Stress Biology, John Innes Centre,
Norwich Research Park

Abstract

We present a novel method for comparing biological time series which has been successfully used to distinguish between differentially induced calcium oscillations. Within many branches of the natural sciences and engineering, the extraction of useful information from noisy data presents a huge challenge. Many biological processes give rise to data which appear random but may be chaotic, such as the calcium oscillations that we investigate here. Distinguishing between different processes in a chaotic system can be a highly non-trivial exercise in the presence of even moderate noise levels. In such cases, classical approaches might often not be able to pinpoint key differences between data sets. The trajectories of any dynamical system have a unique structure, and to our knowledge a quantitative exploitation of differences in attractor shapes has not been developed. Here we introduce an approach to separate different processes by comparing their attractor reconstructions. The method was tested in the symbiosis signaling pathways of the legume plant *Medicago truncatula*.

1 Introduction

In any biological system, the flow of information is a key process. Signaling pathways transfer information about both the surrounding environment and the internal state of an organism. In symbioses these communication pathways play an essential part of the interaction between the species.

Legume plants that have developed several symbiotic; they form associations with both fungi and bacteria. The fungi, arbuscular mycorrhiza, aid the plant mainly in phosphate uptake. The symbiotic bacteria on the other hand form nitrogen-fixing nodules in the plant roots. Both symbionts receive carbon in return (Oldroyd and Downie, 2004). First, the microbes recognize plant compounds in the soil, e.g. flavanoids (Harrison, 2005), and the bacteria then release nodulation factors (Nod from now on), which diffuse to the plant root and activate the signaling pathway required for nodulation (Oldroyd and Downie, 2004). In the fungi symbiosis a different diffusible signal molecule is hypothesised to have the same function as Nod, activating the signaling pathway leading to mycorrhizal infection (Harrison, 2005).

Even though these two pathways produce fundamentally different end results, it is known that they have several key signaling components in common (Kosuta et al., 2008). In *Medicago truncatula* the two genes DMI1 and DMI2 are required in both pathways to activate calcium spiking. Downstream of the spiking, DMI3 is necessary in both symbioses, but further down the pathways diverge and results in either nodules or mycorrhizal infection (Fig. 1).

Calcium, known to be a secondary messenger in many biological processes, plays a central role in this symbiotic signaling pathway (Sanders et al., 2002, Oldroyd and Downie, 2006). In *M. truncatula*, the calcium oscillations from both symbioses have been analyzed mathematically, and the results support the hypothesis that the Nod spiking might be chaotic (Kosuta et al., 2008). A key feature of chaos is a great sensitivity to initial conditions, so that a small difference in input gives a large difference in output, a cause of the exponential divergence of trajectories (Strogatz, 1994). However, chaos can also produce almost periodic and very robust patterns, which can be stabilized through the process of chaos control (Ott et al., 1990). Biological systems may take advantage of these features and use them efficiently to generate multiple responses.

Chaotic systems are deterministic, yet their behavior often appears random and can thus easily be mistaken for being stochastic. Analyzing such dynamics can become a time-intensive and challenging undertaking, especially when the noise level is high and the availability of data is limited. In this symbiotic signaling pathway, a better method was needed to distinguish between different calcium oscillations.

Time delay embedding of a time series enables phase space dynamics to be characterized. For chaotic dissipative systems this process results in an attractor. Chaotic attractors often have a complex geometry with fractal dimensions, are often called *strange*, and possess a unique shape. A classical example is the Lorenz attractor, with a shape similar to twisted butterfly wings (Kantz and Schreiber, 1997). However, to our knowledge a quantitative shape comparison of these attractors has not been exploited to aid the analysis.

We have analyzed time series of calcium oscillations induced by two different symbioses, nitrogen-fixing bacteria and mycorrhizal fungi. The two symbionts seem to elicit differing calcium oscillations, which have been shown previously to be chaotic (Kosuta et al., 2008). We then attempted to distinguish between them by comparing the unique shapes of the attractor reconstructions using an entropy measure, the Kullback-Leibler divergence. This could be a useful method in other noisy, dynamical systems.

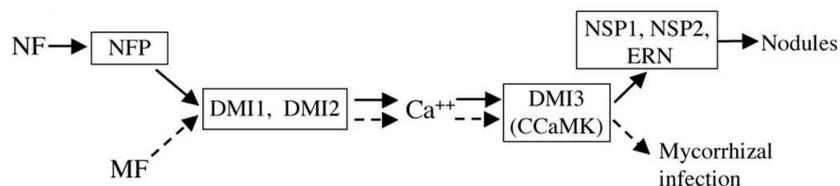


Figure 1: The symbiosis signaling pathway of *M. truncatula*. DMI1, DMI2 and DMI3 are necessary genes for both Nod factor and mycorrhizal factor signalling. NF, Nod factor; MF, mycorrhizal factor (adopted from Kosuta *et al.*, 2008).

2 Method

The data came from the model legume *M. truncatula* containing the calcium reporter gene cameleon (Miwa *et al.*, 2006). Root hair cells were exposed to either Nod or mycorrhizal fungi (Myc from now on), and the resulting calcium oscillations were measured as described in Sun *et al.* (2008). Six Nod traces and six Myc traces were used in the analysis. These time series were detrended with a moving average (Brockwell and Davis, 2002), tested for stationarity (Kennel, 1997), and noise reduced (Schreiber, 1993). The optimal embedding dimension was found to be 6, which is appropriate for an attractor with up to three dimensions (Kantz and Schreiber, 1997). In this analysis we used 2D projections of the attractors, and the time delay was chosen to be fifteen seconds by a drop in autocorrelation to $(1-(1/e))$ (Rosenstein *et al.*, 1993).

A script was written in the statistical package R (<http://www.r-project.org/>) to compare the attractor reconstructions. It places a grid over the phase space plot, and assigns a probability to each grid box of it containing a data point. The result is a 2D matrix of probabilities. We used a 100 x 100 grid and employed the Parzen Window technique using 2D Gaussians of variance equal to the grid size to smooth the distributions. To compare two such shapes, the Kullback-Leibler divergence was used (Fig. 2), which measures the difference between two probability distributions P and Q (Burnham and Anderson, 1998). The result is given in the units bits, where identical distributions have a difference of zero bits. As a reference point, a matrix of random probabilities drawn from a uniform distribution was created and compared to the data.

$$D_{\text{KL}}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

Figure 2: The Kullback-Leibler divergence, comparing two probability distributions (P and Q) and calculating the difference between them, D_{KL} .

3 Results

The attractor reconstructions in 2D of Nod show a clear deterministic shape, while the Myc shape look more irregular and were much more varied between the individual time series (Fig. 3).

The average Kullback-Leibler divergence is summarized in Table 1. There is a three-fold difference when comparing Nod to Nod over Nod to the random probability matrix, which shows the high degree of similarity in the Nod attractor reconstructions. The Myc to Myc comparison is consistent with the higher variability shown in the Myc oscillations. The divergence between Myc and the random matrix resulted in the relatively high value of 2.91 bits. This is due to the fact that the Myc attractor reconstructions often have a small area with very high probabilities, contrasting to the smooth distributions of a random matrix. The Nod attractor reconstructions on the other hand are more regular. Comparing Nod and Myc further shows that there is a higher similarity among the Nod traces compared to the Myc.

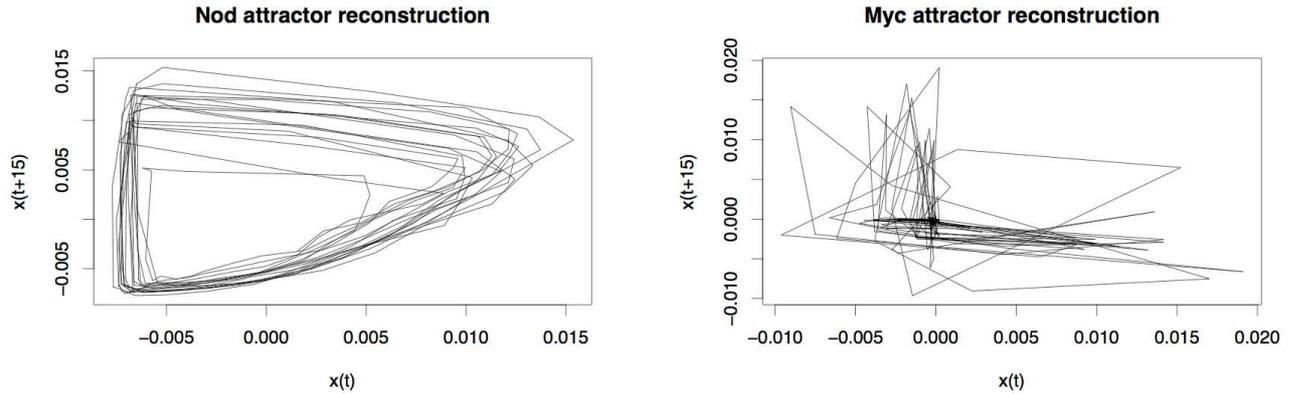


Figure 3: Examples of attractor reconstructions of calcium oscillations in root hair cells of *M. truncatula*. Nod: Oscillations induced in bacterial symbiosis. Myc: Oscillations induced in fungal symbiosis.

Comparison	Bits
Nod - Nod	0.51 ± 0.17
Random - Nod	1.51 ± 0.11
Myc - Myc	0.86 ± 0.63
Random - Myc	2.91 ± 0.19
Nod - Myc	0.81 ± 0.27

Table 1: Results from the comparison of probability distributions in phase space, averages \pm standard deviation. A trace compared to itself i.e. zero difference always returns zero bits. Nod-Myc: Nod and Myc traces compared to each other. Random: A matrix of probabilities generated with random numbers.

4 Discussion

A number of other natural systems have been shown to be chaotic, including weather, population dynamics and diffusion. The human heart beat and neuron signaling in the brain has also been suggested to behave chaotically. In all these fields of study, separating noise and stochastic effects from the deterministic dynamics of the system is a complex problem, and methods such as the one presented here could provide useful information.

The Kullback-Leibler divergence shows clear differences between the calcium oscillations induced by Nod or Myc. It also confirms that the Nod data are very similar to each other, while the Myc traces are much more variable. This gives a quantification of the visual difference seen in the shapes of the attractor reconstructions. Overall, this supports that the signaling pathway is activated differently in the two symbioses. It is possible that the oscillations induced by the fungi does not follow any attractor, since they are so varied and the trajectories does not show a clear deterministic shape in phase space. Since the correct dimension of the attractors is indicated to be 3D, our method is currently being extended to apply to 3D reconstructions. Then, the attractors could also be twisted in phase space to minimize the Kullback-Leibler divergence.

In the analysis of time series such as these it is difficult to pinpoint differences due to the high level of noise and variability in the oscillations, effects that could make the system seem stochastic. However, by attractor reconstruction and comparison, the differences of the deterministic dynamics emerge.

References

- Brockwell P.J., Davis R.A. (2002). *Introduction to Time Series and Forecasting*. Springer-Verlag.
- Burnham, K.P. and Anderson, D.R. (1998). *Model selection and multimodel inference*. New York, USA. Springer Science + Business Media, LLC. Second edition.
- Harrison, M.J. (2005). *Signaling in the Arbuscular Mycorrhizal Symbiosis*. *Annu. Rev. Microbiol.* **59**, 19–42.
- Kantz, H. and Schreiber, T. (1997) *Nonlinear time series analysis*. New York, NY, USA: Cambridge University Press.
- Kennel, M.B. (1997). *Statistical test for dynamical nonstationarity in observed time-series data*. *Phys Rev E* **56**, 316–321.
- Kosuta *et al.* (2008). *Differential and chaotic calcium signatures in the symbiosis signalling pathway of legumes*. *PNAS* **105**, 9823–9828.
- Miwa, H., Sun, J., Oldroyd, G., and Downie, A. (2006). *Analysis of calcium spiking using a cameleon calcium sensor reveals that nodulation gene expression is regulated by calcium spike number and the developmental status of the cell*. *The Plant J* **48**: 883–894.
- Oldroyd and Downie (2004). *Calcium, kinases and nodulation signalling in legumes*. *Nat Rev Mol Cell Biol.* **5**, 566–76.
- Oldroyd and Downie (2006) *Nuclear calcium changes at the core of symbiosis signalling*. *Curr. Op. Plant Bio.* **9**, 351–357.
- Ott E, Grebogi C, Yorke JA (1990). *Controlling chaos*. *Phys Rev Lett* **64**, 1196–1199.
- Rosenstein MT, Collins JJ, Luca CJD (1993). *A practical method for calculating largest Lyapunov exponents from small data sets*. *Physica D* **65**, 117–134.
- Sanders *et al.* (2002). *Calcium at the crossroads of signalling*. *The Plant Cell*. S401-S417 (Supplement).
- Sun J, Cardoza V, Mitchell DM, Bright L, Oldroyd G, *et al.* (2006). *Crosstalk between jasmonic acid, ethylene and Nod factor signaling allows integration of diverse inputs for regulation of nodulation*. *The Plant Journal* **46**, 961–970.
- Schreiber, T (1993). *Extremely simple nonlinear noise-reduction method*. *Phys Rev E* **47**, 2401–2404.
- Strogatz, S. H. (1994). *Nonlinear Dynamics and Chaos*. Perseus Book Publishing, USA.