MATH:7450 (22M:305) Topics in Topology: Scientific and Engineering Applications of Algebraic Topology

Nov 6, 2013: Stable Persistence and time series.

Fall 2013 course offered through the University of Iowa Division of Continuing Education

Isabel K. Darcy, Department of Mathematics
Applied Mathematical and Computational Sciences,
University of Iowa

http://www.math.uiowa.edu/~idarcy/AppliedTopology.html
Stability of Persistence Diagrams
David Cohen-Steiner, Herbert Edelsbrunner, John Harer

Given sets $X$, $Y$ and bijection $\beta: X \rightarrow Y$,

Bottleneck Distance: $d_B(X, Y) = \inf \sup_{\beta} \max \{|x_1 - y_1|, \ldots, |x_n - y_n|\}$
Stability theorem:

Let $X$ be a triangulable space with continuous tame functions $f, g : X \to \mathbb{R}$. Then the persistence diagrams, $D(f)$ and $D(g)$, satisfy

$$d_B(D(f), D(g)) \leq \| f - g \|_\infty = \sup_{x} \{ |f(x) - g(x)| \}$$
Given sets $X$, $Y$ and bijection $\beta : X \rightarrow Y$,

Wasserstein distance:

$$W_q(X, Y) = \left[ \inf_{\beta} \sum_{x \in X} || x - \beta(x) ||_{\infty}^q \right]^{1/q}$$
Stability theorem:

Let $X$ be a triangulable space whose triangulations grow polynomially with constant exponent $j$. Let $f, g : X \rightarrow \mathbb{R}$ be tame Lipschitz functions.

Then there are constants $C$ and $k > j$ no smaller than 1 such that persistence diagrams, $D(f)$ and $D(g)$, satisfy

$$W_q(D(f), D(g)) \leq C \| f - g \|_{\infty}^{1 - k/q}$$

for every $q \geq k$. 
And section 9.1 in
Computational Topology: An
Introduction By Herbert
Edelsbrunner, John Harer
Goal: To determine what genes are involved in a particular periodic pathway.

Application: segmentation clock of mouse embryo.

- 1 somite develops about every 2 hours

What genes are involved in somite development?
Persistence:

For each of 7549 genes, create

\[ f_k : S^1 \rightarrow R, \ k = 1, \ldots, 7549 \]

\[ f_k \ (\text{time point } i) = \text{amount of RNA at time point } i \text{ for gene } k \]
Figure 8. Function $g(x)$ for the expression pattern of Axin2.

http://www.plosone.org/article/info:doi/10.1371/journal.pone.0002856
Data from:

Published Online November 9 2006
Science 8 December 2006:
Vol. 314 no. 5805 pp. 1595–1598
DOI: 10.1126/science.1133141

REPORT

A Complex Oscillating Network of Signaling Genes Underlies the Mouse Segmentation Clock

Mary-Lee Dequéant\textsuperscript{1,2,3}, Earl Glynn\textsuperscript{3}, Karin Gaudenz\textsuperscript{3}, Matthias Wahl\textsuperscript{1,3}, Jie Chen\textsuperscript{4}, Arcady Mushegian\textsuperscript{3}, Olivier Pourquié\textsuperscript{1,2,3,*}
During the formation of each somite, Lfng is expressed in the PSM as a wave that sweeps across the tissue in a posterior-to-anterior direction (1). Therefore, by visually comparing the anteroposterior position of the Lfng expression stripes in the PSM in stained embryos, it is possible to define an approximate chronological order of the embryos along the segmentation clock oscillation cycle (3, 4). We collected PSM samples from 40 mouse embryos ranging from 19 to 23 somites and used their Lfng expression patterns as a proxy to select 17 samples covering an entire oscillation cycle.

Indeed, due to technical issues, the right PSM samples of the time series were dissected from mouse embryos belonging to five consecutive somite cycles, and they were ordered based on their phase of Lfng expression pattern (revealed by in situ hybridization on the left PSM of each dissected mouse embryo) to reconstitute a unique oscillation cycle [5].
Fig. 2. Identification of cyclic genes based on the PSM microarray time series.

M Dequéant et al. Science 2006;314:1595-1598
ArrayExpress - functional genomics data

ArrayExpress is a database of functional genomics experiments that can be queried and the data downloaded. It includes gene expression data from microarray and high throughput sequencing studies. Data is collected to MIAME and MINSEQE standards. Experiments are submitted directly to ArrayExpress or are imported from the NCBI GEO database.

E-TABM-163 - Transcription profiling of murine presomitic mesoderms of 17 samples at various time points to identify cyclic genes of the mouse segmentation clock

Organism: Mus musculus

Status: Released on 6 November 2006, last updated on 27 March 2012

Samples (23) - Click for detailed sample information and links to data

Arrays (2)
- A-AFFY-6 - Affymetrix GeneChip Murine Genome U74Av2 [MG_U74Av2]

Description: A microarray time series was generated to identify cyclic genes of the segmentation clock in the mouse. The right posterior half of the embryo was excised (E10.5) from 17 embryos at 4 time points which were sampled at the right posterior half of the embryo at equivalent intervals. Seven controls from the wild type were used.
E-TABM-163 - Transcription profiling of murine presomatic mesoderms of 17 samples at various time points to identify cyclic genes of the mouse segmentation clock

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Age</th>
<th>Developmental Stage</th>
<th>Organism</th>
<th>Strain/Line</th>
<th>Individual</th>
<th>Organism Part</th>
<th>Links to Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charles River # 022</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 1</td>
<td>right or left presomatic mesoderm (posterior half)(a)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 023</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 1</td>
<td>right or left presomatic mesoderm (posterior half)(b)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 024</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 2</td>
<td>right or left presomatic mesoderm (posterior half)(a)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 026</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 2</td>
<td>right or left presomatic mesoderm (posterior half)(b)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 027</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 3</td>
<td>right or left presomatic mesoderm (posterior half)(a)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 028</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 3</td>
<td>right or left presomatic mesoderm (posterior half)(b)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 029</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 01</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 030</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 02</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 031</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 03</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 032</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 04</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 033</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 05</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 034</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 06</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 035</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 07</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 036</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 08</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
<tr>
<td>Charles River # 037</td>
<td>9.5 days</td>
<td>embryo</td>
<td>Mus musculus</td>
<td>CD-1</td>
<td>Embryo 09</td>
<td>right presomatic mesoderm (posterior half)</td>
<td>...</td>
</tr>
</tbody>
</table>
Persistence:

For each of 7549 genes, create

\[ f_k: S^1 \to R, \quad k = 1, \ldots, 7549 \]

\[ f_k \text{ (time point } i) = \text{ amount of RNA at time point } i \text{ for gene } k \]
17 time points $\Rightarrow$ 17 equally space time points

microarray expression of gene $k$ at time $i$ $\Rightarrow$ ranked order of microarray expression of gene $k$ at time $i$

Ex: $(0.41, 0.63, 0.11, 0.23, 0.59) \Rightarrow (3, 5, 1, 2, 4)$. 
\( f_k \) (time point \( i \)) = RNA intensity at time point \( i \) for gene \( k \).

\( \pi(f_k) \) = replace RNA intensity with rank order.

\[
g(x_i) = \left[ \pi(f_k)(x_i) - 1 \right] / (17 - 1), \quad \text{for} \quad i = 1, \ldots, 17.
\]

\( g(x) \) obtained by linear interpolation for \( x \neq x_i \) for some \( i \).

Note: \( 0 \leq g(x) \leq 1 \).
Figure 8. Function $g(x)$ for the expression pattern of Axin2.

http://www.plosone.org/article/info:doi/10.1371/journal.pone.0002856
Not period 2:
Not period 2:
Not period 2:
Not period 2:
Period 2
\[ \Phi^q(f) = \sum_{i=1}^{n} (d_i - b_i)^q \]

**Stability Theorem for Total Persistence.** Let \( f, g : S^1 \to \mathbb{R} \) be Lipschitz functions with Lipschitz constant one and let \( q \geq 2 \). Then

\[ |\Phi^q(f) - \Phi^q(g)| \leq 4q\pi^{q-1} \cdot \|f - g\|_{\infty}. \]
Figure IX.3: The step functions characterizing the distribution of the verified genes in the ordered lists generated with the periodicity measures $\mu_q$, for $q = 0, 1, 2, 3, 4$. For better visibility, we truncate the lists after the first three and a half thousand genes. Moving northwest, toward the upper left corner of the rectangle, we first cross the graph of $f_0$, then that of $f_1$, and finally the graphs of $f_2$, $f_3$, $f_4$ in an order that depends on the exact route we choose.
Figure 1. Identification of benchmark cyclic genes in the top 300 probe set lists of the five methods.


http://www.plosone.org/article/info:doi/10.1371/journal.pone.0002856
Figure 2. Comparison of the intersection of the top 300 ranked probe sets from the five methods.

http://www.plosone.org/article/info:doi/10.1371/journal.pone.0002856