

第三届中国R语言会议

R与Python在昼夜节律分析中的应用

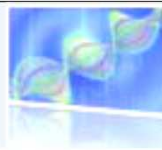
杨仁东

cauyrd@gmail.com

中国农业大学生物学院06级博士

2010.6.15





报告提纲

1

昼夜节律的定义与分析

2

昼夜节律的数学模型

3

R分析生物时间序列的周期性

4

Python调用R进行节律的建模

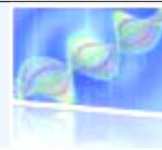
5

模拟数据与实验数据的分析结果

6

BioClock计算平台





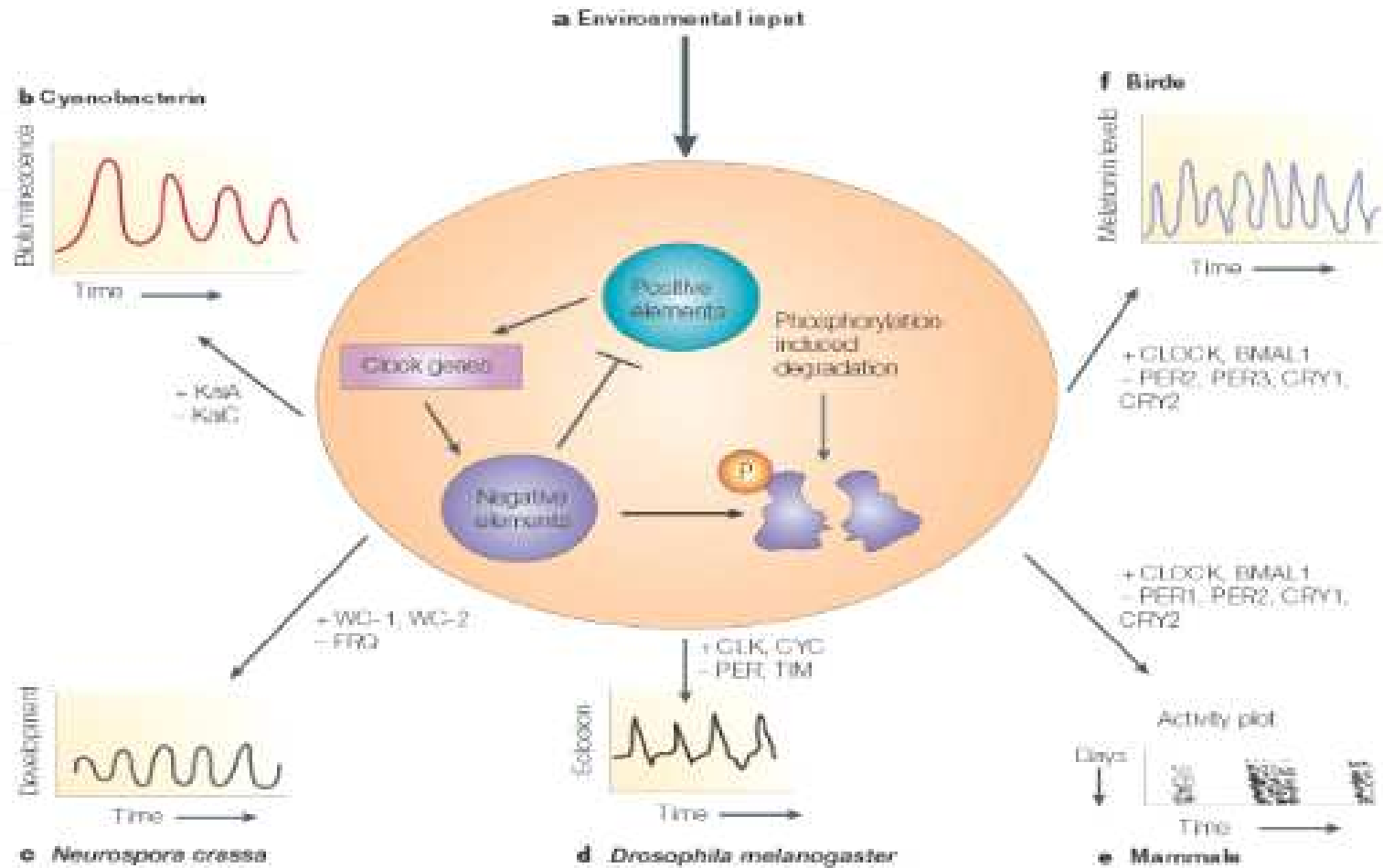
1.1 昼夜节律定义

- 昼夜节律，英文称为**Circadian Rhythm**，是一种以**24**小时左右为周期变化的生理学过程
- 昼夜节律就是我们通常意义上讲的生物钟



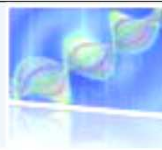


1.2 不同物种中的昼夜节律



Copyright © 2006 Nature Publishing Group
Nature Reviews | Genetics





报告提纲

1

昼夜节律的定义与分析

2

昼夜节律的数学模型

3

R分析生物时间序列的周期性

4

Python调用R进行节律的建模

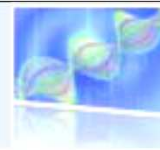
5

模拟数据与实验数据的分析结果

6

BioClock计算平台

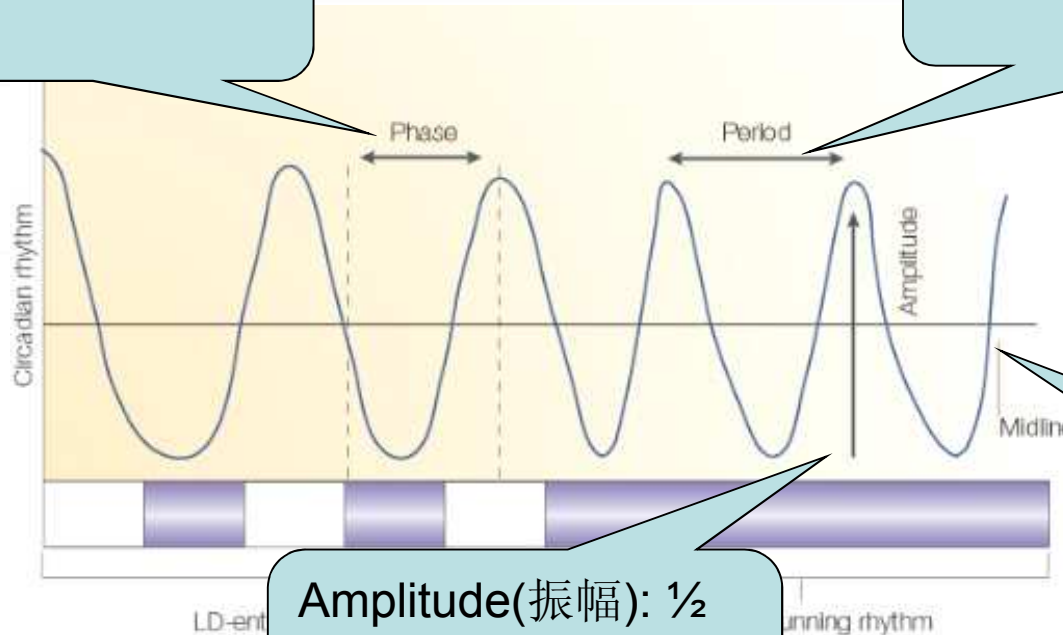




3.3 昼夜节律的数学模型

Phase(相位): peak time所在的时间点

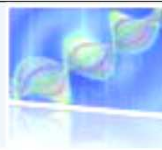
Period(周期): 两个连续峰值之间的距离



Amplitude(振幅): $\frac{1}{2}$ 波峰-波谷振荡距离

Mean level: 时间序列样本点的平均

Copyright © 2005 Nature Publishing Group
Nature Reviews | Genetics



报告提纲

1

昼夜节律的定义与分析

2

昼夜节律的数学模型

3

R分析生物时间序列的周期性

4

Python调用R进行节律的建模

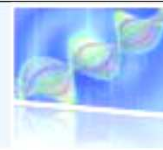
5

模拟数据与实验数据的分析结果

6

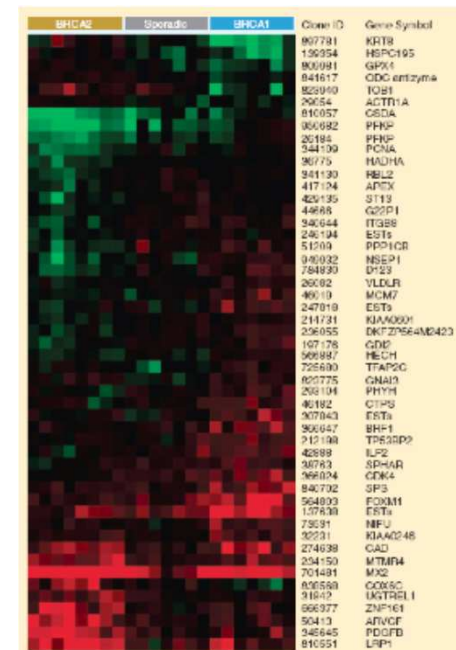
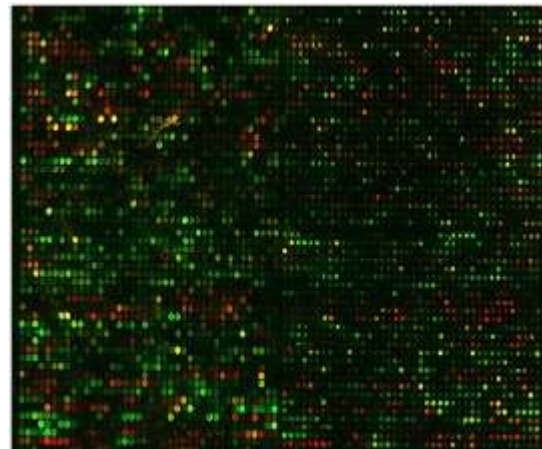
BioClock计算平台

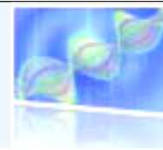




3.1 生物时间序列来源

- 数据来源
 - DNA Microarray 生物芯片数据
- 实验条件
 - 通常每隔4小时测量一次，跨越两天的时间，每条时间序列有12、13个时间点
 - 24小时弱光照
 - 昼夜恒温

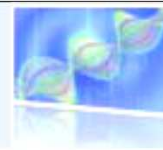




3.2 生物时间序列周期性分析

- 生物时间序列周期性的识别
 - 问题：昼夜节律时间序列时间点少，经典的周期性识别方法（如**Fourier**变换）无法适用
 - 解决：采用高分辨率的周期性频谱分析方法
 - 自回归谱分析（**Autoregressive spectral analysis**）
 - 对于**circadian expression data**的处理非常有效
 - **ARSER**算法



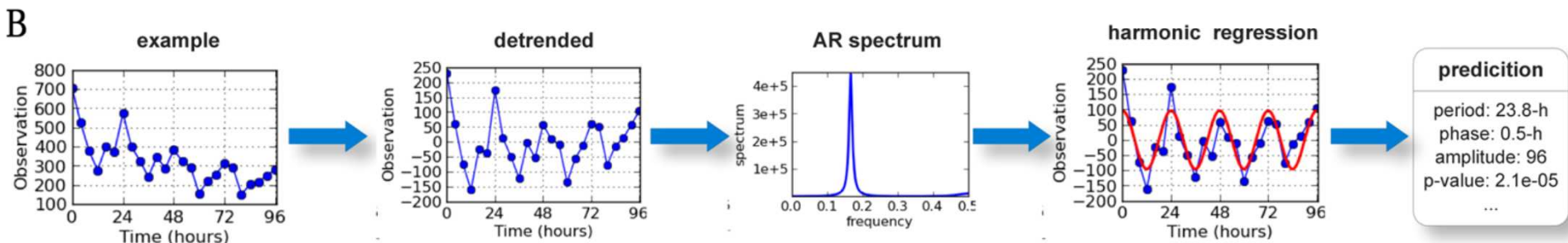


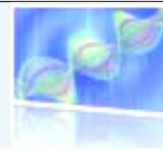
3.3 ARSER算法

- 数学原理
 - **AR(p)**模型拟合
 - 计算**AR**功率谱
- 分析流程

$$x_t = \sum_{i=1}^P \alpha_i x_{t-i} + \varepsilon_t \quad (1)$$

$$p_x(\omega) = \frac{\sigma_\varepsilon^2}{|1 + \sum_{k=1}^P \alpha_k e^{-i\omega k}|^2} \quad 0 \leq \omega < \pi \quad (2)$$





3.4 自回归谱分析的R代码

- **spec.ar(stats) :**
 - Estimate Spectral Density of a Time Series from AR Fit
- **Description:**
 - Fits an AR model to x (or uses the existing fit) and computes (and by default plots) the spectral density of the fitted model.
- **Usage**
 - `spec.ar(x, n.freq, order = NULL, plot = TRUE, na.action = na.fail, method = "yule-walker", ...)`
- **Arguments**
 - **x:** A univariate (not yet:or multivariate) time series or the result of a fit by ar.
 - **n.freq:** The number of points at which to plot. **order** The order of the AR model to be fitted. If omitted, the order is chosen by AIC.
 - **method:** method for ar fit. (Burg/Yule-Walker/mle...)





报告提纲

1

昼夜节律的定义与分析

2

昼夜节律的数学模型

3

R分析生物时间序列的周期性

4

Python调用R进行节律的建模

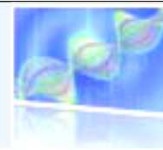
5

模拟数据与实验数据的分析结果

6

BioClock计算平台

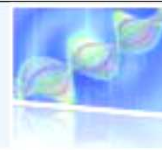




4.1 Python调用R进行节律建模

- **R**的**spec.ar**函数用来识别时间序列的周期性，完成算法的统计计算部分；数据的处理、与回归部分用**Python**程序完成
- **Python**语言有丰富的扩展包，**Python**可以通过**Rpy**包调用**R**函数
 - **Rpy**包逐渐被其升级版本**Rpy2**所替代





4.2 代码细节

Import numpy as np

from rpy import r # 导入Rpy包

x = r.seq(0,48,4) # x <- seq(0,48,4)

x = np.array(x)

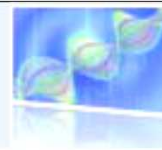
y = r.cos(2*r.pi/24*x) + r.rnorm(r.length(x))

arse = r.sepc_ar(y) # 自回归波谱分析的计算 arse<- spec.ar(y)

r.print(arse['freq']) # print(arse\$freq)

r.print(arse['spec'])





4.3 Rpy2

- Python通过rpy2.robjects中r实例访问R的对象
- 调用**R**函数
 - 以属性的形式
 - 以关键字的形式
 - 直接作为R的解释器
- 返回多为**RVector**对象
- **Rpy2**比**Rpy**在逻辑上有更好的组织
- 更多内容参见**Rpy2**官方文档

```
In [7]: from rpy2.robjects import r
In [8]: print r.pi
Out[8]: 3.141593
In [9]: print r['pi']
Out[9]: 3.141593
```

```
In [8]: rscript = '''
...: x <- c(5,1,3,9,0)
...: print(sort(x))
...: '''
In [9]: r(rscript)
Out[9]: <RVector - Python:0x039C0620 / R:0x06E0F448>
```

```
In [6]: r.pi
Out[6]: <RVector - Python:0x03C33378 / R:0x06CC2F68>
In [7]: len(r.pi)
Out[7]: 1
In [8]: r.pi[0]
Out[8]: 3.1415926535897931
```



Introduction
Pre-installation
Usage
Input/Output Files
Download
Additional Tools
Contact

Rendong Yang and Zhen Su, **Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation** *Bioinformatics*. 2010 Jun 15;26(12):i168-74. [link]

Introduction

ARSER is a Python package for identifying periodic expression profiles in analyzing circadian microarray data and has been released under the GPL.

Pre-installation

ARSER package is implemented by Python calling R program.

Before using the package, please install the following software and packages first:

- Programing environments:
 1. Python version 2.5 or later
 2. R version 2.7 or later
 - Packages:
 3. scipy version 0.7 or later
 4. numpy version 1.1 or later
 5. matplotlib version 0.99 or later
 6. Rserve
- Tips.

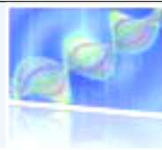
To avoid wading through all the details (and potential complications) on Installation, the easiest thing for you to do is use one of the pre-packaged python distributions that already provide scipy/numpy/matplotlib built in. The Enthought Python Distribution (EPD) for Windows, OS X or Redhat is an excellent choice that "just works" out of the box. Another excellent alternative for Windows users is Python (x, y).

Usage

- Command-line running:
usage: python arser.py input_file_name output_file_name
example: \$ python arser.py data.txt output.txt >& log.txt

<http://bioinformatics.cau.edu.cn/ARSER>





报告提纲

1

昼夜节律的定义与分析

2

昼夜节律的数学模型

3

R分析生物时间序列的周期性

4

Python调用R进行节律的建模

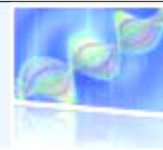
5

模拟数据与实验数据的分析结果

6

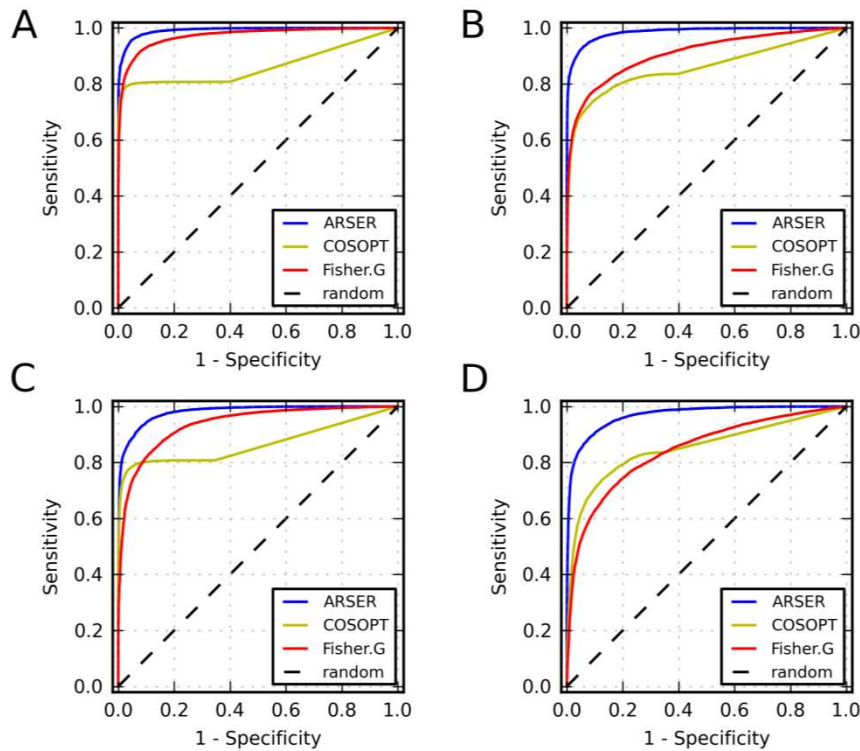
BioClock计算平台





5.1 模拟数据实验1

- 对周期性与非周期性数据的预测
 - 周期性信号分为平稳的与非平稳的
 - 非周期性信号分为白噪声与**AR(1)**随机过程



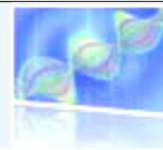
ROC曲线评估周期性预测

A 图 平稳周期信号与白噪声信号数据集

B 图 非平稳周期信号与白噪声信号数据集

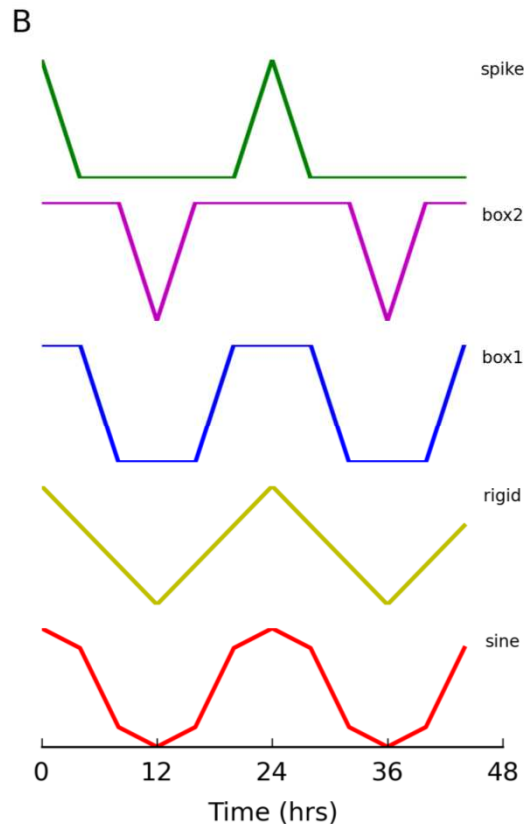
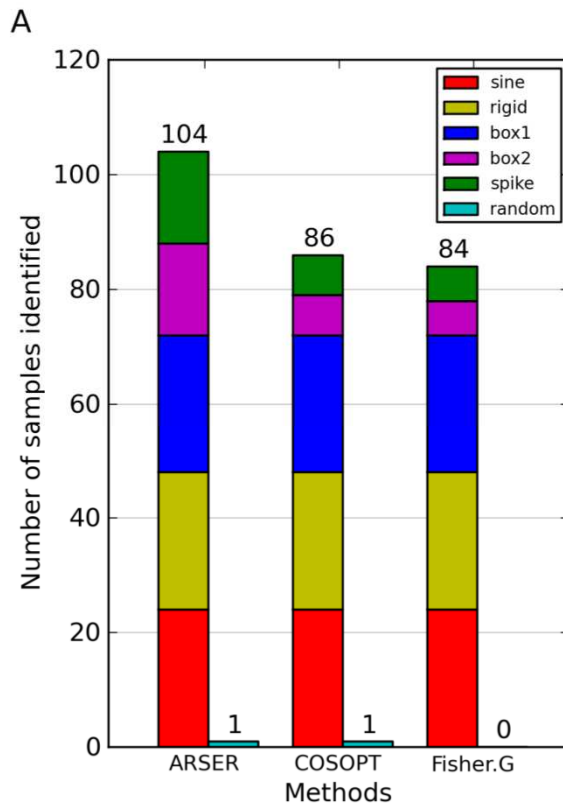
C 图 平稳周期信号与AR(1)随机信号数据集

D 图 非平稳周期信号与AR(1)随机信号数据集

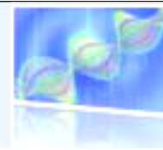


5.2 模拟数据实验2

- 对于特殊波形的识别

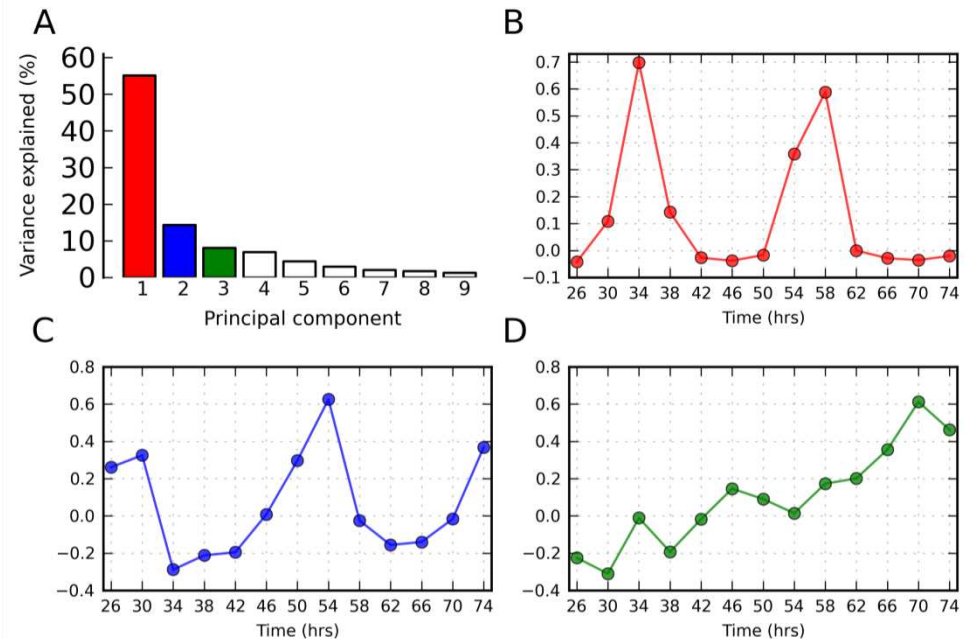
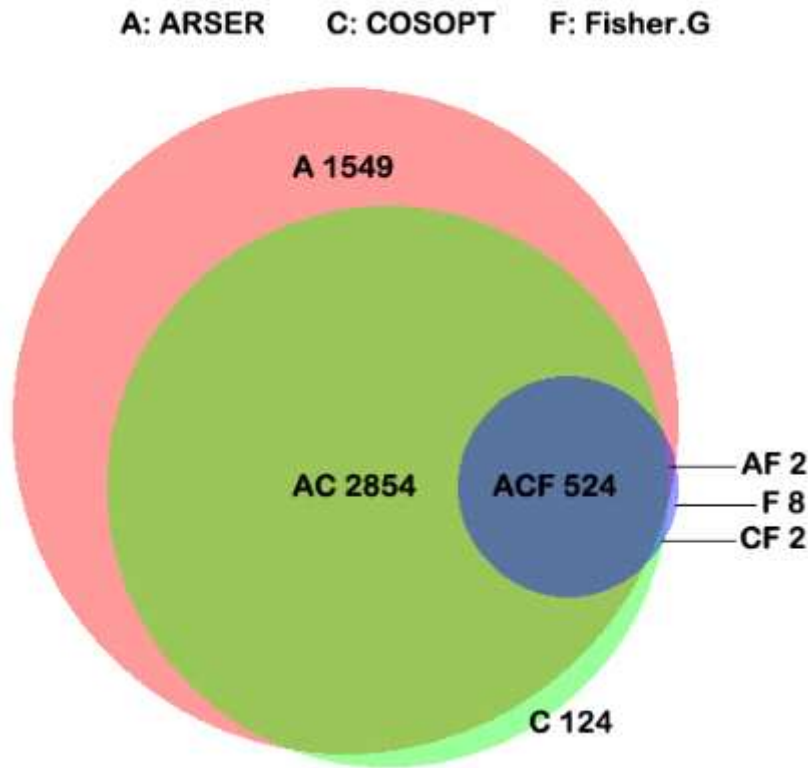


利用自回归谱分析的方法针对除了除了正弦曲线的其他特殊周期波形的数据的预测



5.3 生物芯片数据计算

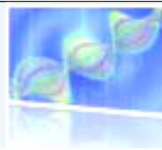
- 研究模式植物拟南芥的生物钟基因



主成分分析(PCA) 鉴定新识别出的生物钟基因

比较三种算法识别的生物钟基因





报告提纲

1

昼夜节律的定义与分析

2

昼夜节律的数学模型

3

R分析生物时间序列的周期性

4

Python调用R进行节律的建模

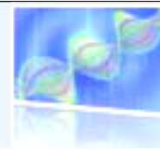
5

模拟数据与实验数据的分析结果

6

BioClock计算平台





BioClock 计算平台

<http://bioinfo.cau.edu.cn/BioClock/index.php>

The screenshot shows the BioClock System website. At the top, there's a header with the title "BioClock System" and the tagline "interpret circadian rhythm". Below this is a navigation bar with links: Home, Analysis, Browse, Jobs, Search, Download, and FAQ. The main content area is divided into two columns. The left column contains sections for "News" (with a date 2010/4/10 and text "Analysis module available"), "Statistics" (showing "Jobs runned: 47"), and "Related Links" (listing Cycbase, CIRCA, Diurnal, and CCDB). The right column contains sections for "Introduction and Motivation" (describing the platform's purpose and citing Wikipedia), "Abilities" (listing functions of the pages), and a list of 7 numbered items detailing the site's features. At the bottom, there's a footer with contact information: "ZhenSu.Lan | TaoWang.Lab | Bioinformatics Center | China Agricultural University" and a copyright notice for 2010.

BioClock System
interpret circadian rhythm

Home Analysis Browse Jobs Search Download FAQ

News
2010/4/10: Analysis module available.

Statistics
Jobs runned: 47

Related Links
Cycbase
CIRCA
Diurnal
CCDB

Introduction and Motivation
We would like BioClock to represent a platform for analysis of biological clock behaviour in living organisms. BioClock could analysis time-course microarray data to find probe sets (genes) behaviour in circadian mode. *A circadian rhythm is a roughly 24-hour cycle in the biochemical, physiological, or behavioural processes of living entities (from Wikipedia).* In addition, genes involved in circadian rhythm process from published microarray experiments were collected and stored in BioClock for users to browse.

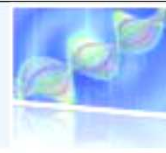
Abilities
Functions of pages listed below:
1. **Home**: this page. A brief introduction of this site. :)
2. **Analysis**: analysis of time-course microarray data; support file uploading.
3. **Browse**: circadian data collection from published works.
4. **Jobs**: check status of submitted analysis job.
5. **Search**: search job quickly by job code generated when submitted analysis. (other functions are being developed.)
6. **Download**: download our collection; support search by keyword.
7. **FAQ**: frequently asked question — a help file! :)

ZhenSu.Lan | TaoWang.Lab | Bioinformatics Center | China Agricultural University
Copyright © 2010. All Rights Reserved.

后台算法由
Pyhton与R实现

前台界面由PHP
实现





Q&A

谢谢

