SSBD: an open public integrated database of microscopy images and quantitative data of biological dynamics

Ho, K.H.L.¹, Tohsato, Y.¹,², Kyoda, K.¹, Itoga, H.¹, Onami, S.¹
¹ Laboratory for Developmental Dynamics, RIKEN BDR
² Osaka Electro-Communication University

ホー・ケネス¹, 遠里 由佳子¹,², 京田 耕司¹, 糸賀 裕弥¹, 大浪 修一¹
¹理化学研究所生命機能科学研究センター発生動態研究チーム
²大阪電気通信大学 情報通信工学部 情報工学科
Need for public repository

- There is a need for public repositories for retaining important knowledge and reusing experimental data

Public data archives are the backbone of modern biological research. Biomolecular archives are well established, but bioimaging resources lag behind them. The technology required for imaging archives is now available, thus enabling the creation of the first public bioimage datasets. We present the rationale for the construction of bioimage archives and their associated databases to underpin the next revolution in bioinformatics discovery.
Problem: difficult to reuse

- Images and Datasets are scattered over the Internet
- Difficult to locate and search
- Most quantitative data use different formats
- Time consuming/difficult to reuse these datasets

(Bao et al., 2006) (Keller et al., 2010) (Kurotaki et al., 2007) (Kondo & Hayashi 2013) (Kunida et al. 2012)
SSBD Database

(image of SSBD Database with links and images)

(Tohsato et al, Bioinformatics, 2016)

**Image data**

- Original format
- BDML/BD5

**Quantitative data**

- BDML/BD5形式に対応

**Software tools**

- http://ssbd.qbic.riken.jp

- Image data
- Quantitative data
- Software tools
BDML/BD5 – an open unified file format
Biological Dynamics Markup Language

(Kyoda et al., Bioinformatics, 2015)

- Meta information in XML

Quantitative data in HDF

Modified from (Goldberg et al, Genome Biology 2005)
SSBD Big Data resource

586 BDML-files, 64.7M entities
1086 image sets, 2.6M slices

<table>
<thead>
<tr>
<th>Organism</th>
<th>Type</th>
<th>Basedon</th>
<th>Paper</th>
<th>Number of BDML-files</th>
<th>Number of images in SSDB</th>
<th>Released date</th>
<th>Update date</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. elegans</td>
<td>Nucleus</td>
<td>Measurement</td>
<td>Bao et al. 2006</td>
<td>1</td>
<td>2013/10/03</td>
<td>2018/11/15</td>
<td></td>
</tr>
<tr>
<td>C. elegans</td>
<td>Nucleus</td>
<td>Measurement</td>
<td>Kyoda et al. 2013</td>
<td>186</td>
<td>2013/10/02</td>
<td>2018/11/15</td>
<td></td>
</tr>
<tr>
<td>C. elegans</td>
<td>Nucleus</td>
<td>Simulation</td>
<td>Kimura &amp; Onami 2005</td>
<td>100</td>
<td>2013/10/03</td>
<td>2018/11/15</td>
<td></td>
</tr>
<tr>
<td>C. elegans</td>
<td>Nucleus</td>
<td>Measurement</td>
<td>Tsuchiya et al. 2016</td>
<td>14</td>
<td>2016/05/20</td>
<td>2018/11/15</td>
<td></td>
</tr>
<tr>
<td>C. elegans</td>
<td>Cell</td>
<td>Measurement</td>
<td>Azuma &amp; Onami 2016</td>
<td>3</td>
<td>2017/03/01</td>
<td>2018/11/19</td>
<td></td>
</tr>
<tr>
<td>D. cf. damesi</td>
<td>Individual</td>
<td>Measurement</td>
<td>Inoue &amp; Kondo 2016</td>
<td>0</td>
<td>6,096</td>
<td>2018/11/15</td>
<td></td>
</tr>
<tr>
<td>D. discoideum</td>
<td>Molecule</td>
<td>Measurement</td>
<td>Komatsu et al. 2015</td>
<td>1</td>
<td>2015/10/03</td>
<td>2018/11/15</td>
<td></td>
</tr>
<tr>
<td>D. discoideum</td>
<td>Cell</td>
<td>Measurement</td>
<td>Watabe et al. 2015</td>
<td>4</td>
<td>61</td>
<td>2018/10/15</td>
<td></td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>Nucleus</td>
<td>Measurement</td>
<td>Keller et al. 2010</td>
<td>2</td>
<td>2013/10/03</td>
<td>2018/11/15</td>
<td></td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>Cell</td>
<td>Measurement</td>
<td>Kondo &amp; Hayashi 2013</td>
<td>0</td>
<td>85,928</td>
<td>2017/10/03</td>
<td></td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>Nucleus</td>
<td>Measurement</td>
<td>Yasugi et al. 2017</td>
<td>0</td>
<td>12</td>
<td>2018/11/14</td>
<td></td>
</tr>
<tr>
<td>D. rerio</td>
<td>Nucleus</td>
<td>Measurement</td>
<td>Keller et al. 2008</td>
<td>7</td>
<td>2013/10/03</td>
<td>2018/11/15</td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>Molecule</td>
<td>Simulation</td>
<td>Arjun &amp; Tomita 2010</td>
<td>1</td>
<td>2013/03/02</td>
<td>2018/11/14</td>
<td></td>
</tr>
<tr>
<td>H. sapiens</td>
<td>Tissue</td>
<td>Measurement</td>
<td>Shiraishi et al. 2016</td>
<td>0</td>
<td>2</td>
<td>2017/10/03</td>
<td></td>
</tr>
<tr>
<td>M. musculus</td>
<td>Nucleus</td>
<td>Measurement</td>
<td>Bashardoust 2012</td>
<td>1</td>
<td>2,800</td>
<td>2017/10/03</td>
<td></td>
</tr>
<tr>
<td>M. musculus</td>
<td>Gene expression</td>
<td>Measurement</td>
<td>Masumoto et al. 2010</td>
<td>0</td>
<td>2014/10/03</td>
<td>2017/10/03</td>
<td></td>
</tr>
</tbody>
</table>
SSBD 2018 release datasets

Increases in 2018

- BDML files: 103
  - 85,540 entities
- Image sets: 453
  - 58,919 slices
SSBD - examples of quantitative data

- \textit{C. elegans} embryogenesis
  - (Bao et al., 2006)

- \textit{D. melanogaster} embryogenesis
  - (Keller et al., 2010)

- \textit{E. coli} single-molecules
  - (Arjunan & Tomita, 2010)

- \textit{C. elegans} embryogenesis
  - (Kyoda et al., 2013)

- Zebrafish embryogenesis
  - (Keller et al., 2008)

- \textit{C. elegans} pronuclear migration
  - (Kimura & Onami, 2005)
SSBD – examples of image data

- Dictyostelium cell
- C. elegans neuronal cell
- C. elegans oocyte
- Mouse ES cell
- Mouse iPS cell
- Mouse cultured cell
- Hamster cultured cell

- Mouse
- D. cf. damesi
- X. laevis
- D. melanogaster
- Rat
- Dog
- Human

Analysis — REST API

Data can be accessed directly from SSBD REST API

7. Apply active contour model (snake)

```python
# Data can be accessed directly from SSBD REST API
```

Get the title, organism and contact name of the dataset

```python
Get the title, organism and contact name of the dataset
```

Plotting the curve

```python
Plotting the curve
```
SSBD - Visualization

BDML対応ツール

SSBD DBツール

ImageJプラグイン

JavaScriptアプリケーション

http://ssbd.qbic.riken.jp/image/publicKey/
SSBD repository service

- Submitting publications to journals with image and numerical data
- Reviewer has access to datasets before publications
- After publications, SSBD will include them in our release.

(Toyoshima et al. 2016)

(Iwata et al. 2017)

(Azuma & Onami 2017)
Collaborations in Japan & Global partners

Working with various Japanese societies, institutes and research centers.
SSBD

Benefits of Open Science

• Public repository for sharing and reuse of data by biologists
• Open collaboration globally in developing computational tools/mathematical models
Acknowledgements

- Project group
  Yuichi Iino (Univ. Tokyo)
  Sigeo Ihara (Univ. Tokyo)
  Hiroki Ueda (RIKEN)
  Masahiro Ueda (RIKEN)
  Ryoichiro Kageyama (Kyoto Univ.)
  Kunihiko Kaneko (Univ. Tokyo)
  Sinya Kuroda (Univ. Tokyo)
  Minoru Ko (Keio Univ.)
  Shigeru Kondo (Osaka Univ.)
  Makoto Taiji (RIKEN)
  Shinichi Tate (Hiroshima Univ.)
  Michiyuki Matsuda (Kyoto Univ.)

- Database development
  OMERO team
  DBCLS
  Hiroshi Masuya (RIKEN)
  Norio Kobayashi (RIKEN)
  Kai Lenz (RIKEN)

- Onami Lab members (RIKEN)