OMEGA: viral particle tracking work- and dataflows for images stored in OMERO

Caterina Strambio De Castillia, Ph.D., et al.
Talk outline

• **Who are we:**

• **Vision (long term!):**
  – Developing bio-image informatics as a tool to enable a systems biology approach to the understanding the retroviral life cycle

• **Short term goal:**
  – Develop integrated tools to analyze viral movement in life cells

• **Approach:**
  – Develop a prototype OMEGA software tool to deliver usable viral particle tracking, movement analysis and data-exploration capabilities to end-users at the bench side

• **Progress report:** what we have done so far

• **Next steps:** where do we want to go from here
The OMEGA team

Project co-founders

software engineering

image acquisition

algorithms optimization and integration
syBIT: seed funding for the OMEGA project and integration support

OMEGA syBIT collaboration

OMEGA is a software development project lead by Caterina Strambis De Castiglia at the Laboratory of Viral Replication, Pathogenesis and Immunity of the University of Geneva and involving a multidisciplinary team of scientists from the University of Geneva, SLPSI, Emory University and ETH-ZSCE. In collaboration with syBIT, the OMEGA team will produce an OMEGA framework plugin. This OMEGA plugin version will read images and image metadata stored within the OMEGA image data repository, automatically perform the main steps of real time particle tracking utilizing a workflow defined within ANIMA and will write the results within a customized version of the openBIS metadata repository for data exploration and visualization. In addition OMEGA will produce publication quality graphs for data dissemination.
The vision: imaging as an integral part of systems virology

Systems virology: data could be represented using an hypercube

The “short” term goal: capturing the motion of viral particles within the infected cell
Emory Vaccine Center: tracking retroviral assembly

Jasmine Clark, Lara Pereira and Eric Hunter
Current situation: software-assisted manual tracking and analysis of trajectories

Manual analysis
No workflow management
No data relation
No data standards
Emory Vaccine Center: retroviral movement from the towards the cell periphery

Gag-GFP
mCherry-tubulin

Pereira et al. submitted
Emory Vaccine Center: experimental setup

- GagGFP
- EnvCherry
- CMMT

18-20h PT

10μM Nocodazole

0h, 2h, 4h

Image live cells on Deltavision System
Emory Vaccine Center: live cell imaging

Untreated

2h Nocodazole

Pereira et al. submitted
Emory Vaccine Center: trajectory analysis

A) Log-Log of Traj 23 Phase I, II, III: Transformed data

B) Slope:
- \(= 1\): Diffusion (Brownian)
- \(> 1\): Active Transport
- \(< 1\): Hindered Movement

Pereira et al. submitted

OME User’s Meeting 2011, Pasteur Institute, Paris, France, 2011_06_15-16
Emory Vaccine Center: trajectory analysis

Types of Diffusion (Particle Tracker Data)

Untreated:
- # of cells = 4
- # of Trajectories = 20
- # of Phases = 41
  *line represents median

4h Treated:
- # of cells = 2
- # of Trajectories = 7
- # of Phases = 18
  *line represents median

Pereira et al. submitted
Problem definition: the workflow from infected cells to a mathematical model of movement
What is single particle tracking?

Single Particle Tracking
Connecting the dots

Jacqman and Danuser, 2009
Trajectory segmentation:
Identifying trajectory segments using machine learning


Mathematical analysis of trajectories or trajectory segments:
Computing numerical parameters capable of discriminating between types of movement

Ewers H et al. PNAS 2005;102:15110-15115
Statistical analysis of results:
Comparing results across datasets

Types of Diffusion (Particle Tracker Data)

<table>
<thead>
<tr>
<th>Slope of Log-Log Transformed MSD plot</th>
<th>Untreated</th>
<th>4h Treated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untreated</td>
<td></td>
<td></td>
</tr>
<tr>
<td># of cells</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td># of Trajectories</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td># of Phases</td>
<td>41</td>
<td>18</td>
</tr>
<tr>
<td>*line represents median</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pereira et al. submitted
Current situation: individual steps are not integrated
One solution: integrated computational image analysis
Open Microscopy Environment integrated Analysis
OME: is the software “lego base plate” enabling integrated image analysis.
omega
Integration of the analysis process
workflow
**omega**

Integration of the data pipeline

*Dataflow*
Goals of the OMEGA prototype
What is openBIS? a generic repository and management system for biological data

http://www.cisd.ethz.ch/software/openBIS
openBIS: Support for distributed storage of analysis results

http://www.cisd.ethz.ch/software/openBIS
The concrete plan
Integrated work- and data-flows for viral particle tracking

- Image pixel data/metadata management
- Image processing
- Work- and data-flow management
- Single Particle Tracking
- Trajectory segmentation
- Trajectory segments classification
- Analysis results visualization and exploration
- Statistical analysis of analysis results
OMERO installations: the starting point

✓ Microbiology Institute of Canton Ticino: image data management

✓ University of Applied Sciences of Southern Switzerland (SUPSI): benchmarking installation for software development

✗ Emory Vaccine Center: image data management
OMERO (1): environmental microbiology

Istituto Cantonale di Microbiologia del Canton Ticino
OMERO (2 and 3): viral particle tracking
Emory University and SUPSI

Once they are in OMERO, what do we do with the images?
Goals of the OMEGA prototype
OMEGA prototype: phase I of implementation

OMEGA Java client

Single Particle Tracking

Trajectory Segmentation

Results Visualization and Exploration

File Type 1: Trajectories geometric description

File Type 2: Trajectory segments definition

JNI

metadata

image data

Image data management and rendering

OpenBIS

to be integrated into
OpenBIS

to be integrated into
OpenBIS

DLL - API

 DLL - API

analysis definition

mathematical analysis of trajectories and charts production

OME User's Meeting 2011, Pasteur Institute, Paris, France, 2011_06_15-16
OMEGA prototype: phase II of implementation
OMEGA: progress report
Single Particle Tracking
Sbalzarini’s algorithm

Particle Tracker 2D / 3D

- Global-nearest-neighbor approach
- Less accurate in sub-optimal conditions
- Does not allow for trajectory gaps
- Already available in C
- Easier to integrate

- Re-engineering of sequential application in C → DONE
- Multi-core optimization → DONE
- GPU optimization → NEXT RELEASE
Sbalzarini’s SPT: significant speed-ups will facilitate work on lap-top computers commonly available to individual scientists

<table>
<thead>
<tr>
<th></th>
<th>Particle Radius 6 pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Run time (s)</td>
</tr>
<tr>
<td>Original version (C)</td>
<td>197.65</td>
</tr>
<tr>
<td>Optimized sequential version (C)</td>
<td>133.01</td>
</tr>
<tr>
<td>Multi-cores parallel version (8 cores)</td>
<td>18.18</td>
</tr>
<tr>
<td>Multi-cores parallel version (12 cores)</td>
<td>12.21</td>
</tr>
<tr>
<td>GPGPU parallel version (8 cores)</td>
<td>12.10</td>
</tr>
</tbody>
</table>

- Software re-engeneering of sequential version afforded significant performance improvements
- Multi-core parallelization on 12 cores achieves 16x speed-up
- The GPU preliminary version achieves 16x improvements already on 8 cores
SPT output: File Type 1

particle $p_k$

$\begin{array}{c|cc}
  t_0 & x_0 & y_0 \\
  t_1 & x_1 & y_1 \\
      & \vdots & \vdots \\
  t_n & x_n & y_n \\
\end{array}$

$(x_0, y_0)$

$(x_1, y_1)$

$(x_n, y_n)$
OMEGA: progress report

[Diagram showing the integration of Single Particle Tracking with openBIS, JRI, and Omeiro for trajectory analysis and visualization.]
Trajectory segmentation:

Identifying trajectory segments using machine learning


The trajectory segmentation algorithm consists of four steps:

(1) **Trajectory preprocessing**: remove position noise

(2) **Feature extraction**: A set of 7 features is used to identify the motion type of each segment. Transform trajectory segments to low-dimensional feature vectors.

(3) **Feature classification**: evaluate whether a feature vector corresponds to a certain class of motion or not in a binary classification problem. Uses user-trained Support Vector Machines (SVM) as classifiers.

(4) **Label assignment**: classification ambiguities are resolved and motion types labels are assigned to each individual trajectory segment.
Trajectory Segmentation
MATLAB TOOLBOX
Jo Helmuth and Ivo Sbalzarini

TRAJECTORY_SEGMENTATION
http://www.mosaic.ethz.ch/

SPIDER
http://people.kyb.tuebingen.mpg.de/spider/main.html

SVMlight
http://svmlight.joachims.org/

• The engine of the application was ported to C
• The GUI necessary for supervised machine learning has been integrated into the main OMEGA Java client
TS output: File Type 2

<table>
<thead>
<tr>
<th>coord</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_0$</td>
<td>$y_0$</td>
</tr>
<tr>
<td>$x_1$</td>
<td>$y_1$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$y_2$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>$y_3$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>$y_4$</td>
</tr>
<tr>
<td>$x_5$</td>
<td>$y_5$</td>
</tr>
<tr>
<td>$x_6$</td>
<td>$y_6$</td>
</tr>
<tr>
<td>$x_7$</td>
<td>$y_7$</td>
</tr>
<tr>
<td>$x_8$</td>
<td>$y_8$</td>
</tr>
<tr>
<td>$x_9$</td>
<td>$y_9$</td>
</tr>
<tr>
<td>$x_{10}$</td>
<td>$y_{10}$</td>
</tr>
<tr>
<td>$x_{11}$</td>
<td>$y_{11}$</td>
</tr>
</tbody>
</table>

$$(x_0, y_0)$$

$$(x_1, y_1)$$

$$(x_2, y_2)$$

$$(x_n, y_n)$$
OMEGA: progress report
OMEGA: connecting with OMERO
OMEGA: images review
OMEGA: running image analysis
OME OmegA: progress report

Diagram showing the integration of different components into an OmegA framework, indicating progress on file types and integration with openBIS.
OMEGA: phase II will be completed by August 2011...
OMEGA – openBIS integration: data tree for particle tracking
OME data model

http://www.openmicroscopy.org.uk/site/support/legacy/ome-server/conceptual-framework/object-hierarchy
openBIS data model

http://www.cisd.ethz.ch/software/openBIS
OMEGA – openBIS integration: mapping between the two models
OMEGA – openBIS integration: particle tracking data model
Where do we go from here?
Complexity hinders data integration

• The fundamental complexity of biological systems limits scale and productivity
• Results are difficult to compare across experimental models, set-ups, data types, time and space
• Each question requires new output data structures
• Data structures are difficult to standardize

The current “solution” is twofold:

1. data bases are used as “static” data repositories, which allow open access to current scientific knowledge and long term archiving of such knowledge
2. individual data acquisition centers (i.e. large labs, bioimaging centers, high-through put screening facilities) develop ad hoc solutions for analysis with little cross-integration of results
The size of the problem is staggering...

1. Usable systems capable of scaling for increasing amount of data
2. Enterprise level systems available in a distributed manner at individual laboratories and facilities
3. Systems capable of summarizing results while at the same time revealing the full detail of individual samples
4. Compare results between samples, within groups and across multiple levels of complexity
5. Integrate several multi-dimensional data types
6. *Data must be structured, indexed, and annotated*
We need to start working on a solution…

Can DBs be used as “dynamic” data exploration engines capable of integrating results from different sources, different experimental set ups, different experimental models?

Can the data models be abstracted enough without losing meaning?

Are OLAP cubes a solution?

Would SDCubes that combine HDF + Xml files allow the level of cross-correlation required for capturing biological complexity?

We open to give a meaningful contribution to some of these questions…
GRAZIE!
Caterina Strambio De Castillia

Lello Giulietti  Vanni Galli  Nicholas Vecchietti
Mario Valle  Jeremy Luban  Tiziano Leidi
Andrea Danani  Federico Santoni  Max Maiolo
Eric Hunter  Jasmine Clark  Lara Pereira