

bioWeb3D: an online webGL 3D data visualisation tool

Jean-Baptiste Pettit*¹ and John C. Marioni*¹

¹EMBL-EBI, European Molecular Biology Laboratory - European Bioinformatics Institute, Cambridge, CB10 1SD, UK

Email: Jean-Baptiste Pettit*- jbp Pettit@ebi.ac.uk; John C. Marioni*- marioni@ebi.ac.uk;

*Corresponding author

Abstract

Background: Data visualization is an important part of biology that sometimes proves to be a bottleneck for non trained researchers. This is especially true when it comes to three dimensional (3D) data representation. Plenty of existing software provide all necessary functionalities to represent and manipulate biological 3D datasets but very few are easily accessible (browser based), cross platform and created for non expert users to quickly and efficiently view 3D data before an eventual finer analysis. **Results:** An online HTML5/webGL based 3D visualisation tool has been developed to allow biologists to quickly and easily view interactive and customizable three dimensional representations of their data along with multiple layers of information. Using the WebGL library Three.js written in Javascript, bioWeb3D allows the simultaneous visualisation of multiple large datasets inputted via a simple JSON file, which can be read and analysed locally thanks to HTML5 capabilities. **Conclusions:** Using basic 3D representation technique in a technologically innovative context we are able to provide a program that is not intended to compete with professional 3D representation software but can be very useful for a quick and intuitive representation of reasonably large 3D datasets.

1 Background

Visualisation is a key feature in the analysis of large biological datasets, especially when analysing organized structures with distinct sub-clusters [1]. This is particularly important when analysing 3-Dimensional (3D) datasets. In a situation where a biological process or feature has been described spatially by a set of 3D referenced points, either via laboratory work (confocal microscopy for example) or generated within

a simulation, with some data attached to each point in space, the first approach in order to analyse the results is to visualise it. Often a second step towards the full analysis will be to cluster the observations in different classes based upon the information attached, those results will need visualisation as well.

While various 3D visualisation tools have been developed, they have typically been available via a locally installed piece of software such as BioLayout Express^{3D} [2], Arena3D [3], 3D Genome Tuner [4], the Allen Brain Atlas [5] or Cytoscape [6]. Other 3D visualisation tools have been built online and are accessible through the browser directly, such as AstexViewer [7], which is utilised by the Protein Databank Europe via a Java Applet. More recently, visualisation tools developed using HTML5/WebGL capabilities have been described, although they have focused on very specific applications, such as analysing radiology data [8].

In all those cases no tool has yet allowed biologists to view their own 3D data directly online in an easy, fast and interactive and secure way. Using WebGL and the JavaScript 3D library Three.js, bioWeb3D aims to be a simple, generic, tool for visualising such data.

2 Implementation

bioWeb3D allows the representation of any 3D dataset by defining only two file formats, which enables the user to quickly view their own datasets to their browser. The format of these files is either based on JSON, a widely used structured format on the web [9] or directly through Comma Separated Values format (CSV). The first file used by the application, referred to as “dataset file”, contains the coordinates of every point in the dataset. The second type of file used, the “information layer” file, describes one or several information layers that are associated with the points defined in the first file. For example, if each point defines the location of a cell within a tissue, the second file could describe whether a particular gene is expressed in each cell. That way the tissue expression profile can be represented in the spatial context of the tissue.

Datasets can be viewed and compared in up to four “worlds” (each world refers to a separate visualisation sub-window) at the same time. Although browser based, the application, fully written in Javascript, does not need to send any data to the host server. Instead the modern internet browser’s local file system reading capabilities are used through the HTML 5 FileReader functionality. This allows the application to handle, in a very short period of time, large datasets while ensuring that the privacy of the data is maintained.

Although the focus is on making bioWeb3D simple and easy to use, some options are available to customise how datasets are represented. The application can be used to visualise sequential information, such as 3D protein structures, in which case links can be drawn between the points. In other cases, such as when a

population of cells is considered, the points can be left unlinked as individual particles. The information layers are visualised by colouring the 3D points according to the class that each point belongs to.

2.1 technological overview

bioWeb3D is fully written in HTML/Javascript. It relies heavily upon a relatively recent 3D javascript library called Three.js [10]. This library is used as the main interface between WebGL (cross-platform, royalty-free web standard for a low-level 3D graphics API) [11]. More specifically bioWeb3D creates a simple way to generate and manipulate some simple Three.js objects and to interact with them. In fact most of the development work associated with the creation of bioWeb3D has been to design interactions between the 3D visualisation and the user interface in the most efficient way.

2.2 Defining the input files formats

Using JSON format to input files into bioWeb3D is recommended because of its rigorous structure, which allows fast Javascript object generation within the browser interpreter. Compared to other data-interchange languages, such as XML, JSON is also easily human readable thanks to a light-weight syntax. It is also supported by all of the primary internet browsers.

However, much data generated in the biological sciences is stored within CSV files. Converting CSV documents to the JSON format used in this application is not always trivial. In order to facilitate this process, the application is also able to interpret simple CSV files following a certain format as an input. Finally, CSV to JSON converters are made available as a separate tool.

2.2.1 Dataset files specification

When the user adds a new *Dataset* file, a new Dataset section is created in the “Data” panel of the application. One raw data file contains one dataset.

JSON format The *dataset* file should have a root object called “dataset” which contains:

- The “name” property of the dataset (*e.g.*, “my dataset”);
- The “chain” parameter which has to be set to *true* if the points should be linked (the default value is *false*). The data will be considered sequentially, each point will be linked according to its order in the dataset file;

- The “points” property, which is a two dimensional array representing a list of (x,y,z) vectors that define the co-ordinates of the points.

Listing 1 is an example of a minimal 3 points dataset file:

CSV format To read CSV dataset files, bioWeb3D needs them to follow a particular organisation. Each line represents a point and the three coordinates on each line must be separated by “tabulation” characters.

Listing 4 carries the same information as Listing 1. Please note that although the spatial information remains the same it is not possible to set a name or to link the points within a CSV file input.

2.2.2 *Information layer file specification*

The *Information layer* file contains information about the points described in the Dataset file. The information in this file has to be present in the same order as the points defined in the Dataset file.

JSON format The *information layer* files must have a root element named “information”. Since one information file can define multiple information sets, the structure below “information” is a list. Each element of the list is structured as follows:

- The “name” property (optional);
- The “numClass” property, which indicates the number of different classes the data will be assigned to;
- The “labels” property, which defines a list of names for the “numClass” classes previously defined (optional);
- The “values” property, which defines the class of each point in the dataset. As points do not have single IDs, this property must be in the same order and have the same length as the points defined in the *dataset* file.

For example coming back to the 3 points defined previously, two information layers could correspond to:

- one clustering algorithm that puts the first two points together in class one and the third point alone in a second class
- a second clustering algorithm that puts each point in a separate class

In this case the Information layer file would look like Listing 3.

CSV format The first element of each column will be the name of the information layer, and the rest of the column represents in which class each point belongs. The separation character between columns must be a “tabulation”. Listing 2 carries the same information as Listing 3. Note that it is not possible to use the “labels” property available in Listing 3 within a CSV information layer file.

2.3 User interface

The user can interact with the visualisation via an interface on the right of the screen (Figure 1), which contains three panels. In the “dataset” panel, the user can choose the *datasets* and *information layer* files that should be represented in each world. This panel also allows the user to show/hide specific classes of the selected information layers. Each dataset file entered will create a new sub-panel where the user can input *information layer* files for each world. Selecting an *information layer* in the drop-down list will display the data in the current world and generate a list of classes that the user can modify regarding their visibility and colour. The “View” panel enables the user to choose which of the worlds are shown on the screen, ranging from 1 to 4 simultaneous worlds (Supplementary Fig 1). Finally, the “Settings” panel provides the user with a number of options that affect all the worlds and all datasets, such as allowing the axes scales to be modified.

3 Results and Discussion

3.1 bioWeb3D and local software

Many 3D visualisation software tools, most of which require local installation, exist and provide similar functionalities with standard 3D format input such as Wavefront .OBJ. Some are extremely generic and powerful like Blender. However, these tools are not typically oriented towards a scientific audience. Moreover, those that are more focused on science are often targeted towards a very specific application, especially in medical sciences [4]. In this context we believe that bioWeb3D can be useful as it is completely generic and browser based. It should also be noted that recent browser improvements regarding GPU acceleration through the WebGL paradigm allow bioWeb3D to visualise several hundred thousand points. Additionally, local software is usually platform specific, which is not the case for browser based applications.

3.2 bioWeb3D and Java Applets

As mentioned previously, browser based 3D visualisation tools currently exist mainly in the form of Java Applets. This technology has attracted much criticism in 2012 regarding security flaws, leading the “United

States Computer Emergency Readiness Team” to advise that all Java Applets should be disabled due to current and future Java vulnerabilities [12]. The development of WebGL technology is viewed by many as a candidate for replacing Applets.

3.3 Current limitations

The main current limitation of a WebGL based application is the machine and browser compatibility. Only computers with fairly recent graphic cards will be able to run a 3D environment. It should also be noted that Microsoft has notified the developer community that Internet Explorer is not scheduled to support WebGL in the near future. However, importantly, Chrome, Firefox, Safari and Opera all now support WebGL applications (<http://caniuse.com/webgl>). It could also be important to mention for eventual future developments that WebGL is supported on mobile platforms such as iOS or Android.

4 Conclusions

bioWe3D is designed to be a simple and quick way to view simple 3D data typically in a biological context. Being browser-based, the software can be easily used from one computer to the other without the need for installation. Everything has been designed in order to offer the most gentle learning curve possible. Despite the current limitations in terms of compatibility or rendering performances for large numbers of points, it is the authors’ belief that such tool can be used by non-experts in 3D data representation to have a quick access to visualising their data and the information attached to it in many biological context.

5 Availability and requirements

The full source is available on the github page of the project [13]. A live version of the software is online [14]. You will require a graphical card and a browser with WebGL capabilities to run bioWeb3D.

Author’s contributions

JBP developed the presented software and drafted the manuscript. JCM has been involved in writing and editing the manuscript. He has also given the final approval of this version to be submitted.

Acknowledgements

The authors would like to acknowledge Samuel Croset, Stijn van Dongen, Konrad Rudolph and Sergio Martinez Cuesta for helpful discussion and criticisms.

References

1. Rubel O, Weber GH, Huang MY, Bethel EW, Biggin MD, Fowlkes CC, Luengo Hendriks CL, Keranen SV, Eisen MB, Knowles DW, Malik J, Hagen H, Hamann B: **Integrating data clustering and visualization for the analysis of 3D gene expression data.** *IEEE/ACM Trans Comput Biol Bioinform* 2010, **7**:64–79.
2. Freeman TC, Goldovsky L, Brosch M, van Dongen S, Mazière P, Grocock RJ, Freilich S, Thornton J, Enright AJ: **Construction, Visualisation, and Clustering of Transcription Networks from Microarray Expression Data.** *PLoS Computational Biology* 2007, **3**(10), [http://dblp.uni-trier.de/db/journals/ploscb/ploscb3.html#FreemanGBDMGFTE07].
3. Pavlopoulos GA, O'Donoghue SI, Satagopam VP, Soldatos TG, Pafilis E, Schneider R: **Arena3D: visualization of biological networks in 3D.** *BMC Syst Biol* 2008, **2**:104.
4. Wang Q, Liang Q, Zhang X: **3D Genome Tuner: Compare Multiple Circular Genomes in a 3D Context.** *Genomics Proteomics Bioinformatics* 2009, **7**(3):143–146, [http://linkinghub.elsevier.com/retrieve/pii/S1672022908600431].
5. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR, Fliss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR, Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawal R, Kidney JM, Knapik RH, Kuan CL, Lake JH, Laramée AR, Larsen KD, Lau C, Lemon TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE, Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA, Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova NV, Sivisay T, Slaughterbeck CR, Smith SC, Smith KA, Smith BL, Sodt AJ, Stewart NN, Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson CL, Varnam LR, Visel A, Whitlock RM, Wornoutka PE, Wolkey CK, Wong VY, Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B, Zwingman TA, Jones AR: **Genome-wide atlas of gene expression in the adult mouse brain.** *Nature* 2007, **445**(7124):168–176.
6. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Research* 2003, **13**(11):2498–504, [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=403769&tool=pmcentrez&rendertype=abstract].
7. Hartshorn MJ: **AstexViewer: a visualisation aid for structure-based drug design.** *J. Comput. Aided Mol. Des.* 2002, **16**(12):871–881.
8. Kulkarni DB, Doijade MM, Devrukhkar CS, Zilpe GR, Surana RR: **Article: NetraRIS - a Web based DI-COM Viewer.** *International Journal of Computer Applications* 2012, **48**(24):40–44. [Published by Foundation of Computer Science, New York, USA].
9. Wilde E: **Putting Things to REST.** *Transport* 2007, **15**(November):567–583, [http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.96.7936].
10. **Three.js - JavaScript 3D library** [http://mrdoob.github.com/three.js/].
11. **WebGL 1.0 Specification** [https://www.khronos.org/registry/webgl/specs/1.0/].
12. **United States Computer Emergency Readiness Team** [http://www.kb.cert.org/vuls/id/636312].
13. **bioWeb3D on github** [http://github.com/jbogg/bioWeb3D].
14. **bioWeb3D online** [http://www.ebi.ac.uk/~jbpettit/bioWeb3D].
15. Tomer R, Denes AS, Tessmar-Raible K, Arendt D: **Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium.** *Cell* 2010, **142**(5):800–809, [http://dx.doi.org/10.1016/j.cell.2010.07.043].

```

{ "dataset" : {
  "name" : "my superb dataset",
  "chain" : true,
  "points" :
  [
    [
      0.5,
      100,
      -50.5
    ],
    [
      200,
      10,
      0.0
    ],
    [
      3,
      250.15,
      15
    ]
  ]
}

```

Listing 1: Json dataset file

```

clustering_algo_1 clustering_algo_2
1 1
1 2
2 3

```

Listing 2: CSV information layer file

Figures

Figure 1 - An example of the application of bioWeb3D.

The 3D location of cells within the brain of the marine annelid *Platynereis dumerilii* is shown. Two classes are displayed (in green and blue) along with the shadow of the remaining cells. The User interface is visible on the right of the screen and can be hidden. Data for this figure was taken from [15]


```

{ "information" :
  [
    {
      "name": "clustering algo 1",
      "numClass": "2",
      "labels" : [
        "Category 1",
        "Category 2"
      ],
      "values": [
        1,
        1,
        2
      ]
    },
    {
      "name": "clustering algo 2",
      "numClass": "3",
      "values": [
        1,
        2,
        3
      ]
    }
  ]
}

```

Listing 3: JSON information layer file

```

0.5 100 -50.5
200 10 0.0
3 250.15 15

```

Listing 4: CSV dataset file