

# Utilizing Social Navigation and Visualization Tools for the Creation of a Cooperative Genome Platform

Lindsey Bieda and Patrick Dudas

Consults: Dr. Petr Pancoska/Dr. Brusilovsky

Keyword: Social Navigation, Visualizations, and Genomics

**ABSTRACT** Navigating in a platform can take many forms and can be a hindrance more than it can be beneficial in getting the meaning you are trying to portray. Finding areas of interests and the novelty of a project can be left to the user and if you do not provide a good enough map, they can become lost and frustrated with your design to the point of nullification of its novelty. This can be exacerbated when the interface in your user will be utilizing your application is a visualization. Like trying to interpret the meaning of a painting, this is left to the interpretation of the beholder. This may be fine for art, but for scientific research, there is a necessity to get people to exactly where you want them to be. We provide a usability study and developed timeline in this pursuit of such a goal in the field genomic research and we examine the viewpoint of social navigation [1].

## Section 1 - Introduction

For the past two and half years, Patrick has been working with the Department of Medicine, more specifically the Clinical Pharmacology Department under the advisement of Dr. Petr Pancoska. The project, called Entromics [2], is a new approach to the way genome sequences are being examined. Applications were developed for genome analysis to show changes in the genome sequence over the evolution of the organism. The intuitive notion is that evolution has left its footprint in the optimization of the genome, and throughout time the structure of the sequence has been tailored to produce the exact organism and organism phenotype simply by coding a long, linear string of [A,T,C,G]. This is one of the main thermodynamic properties of genomic analysis - that over time the sequence started

as a non-evolved sequence. The evolutionary energy or the energy to incorporate all necessary optimization changes from the initial state of the DNA sequence to form a biologically fully active state in which the sequence, the one it is today, is still encoded in the sequence. This property is defined by the entropy and enthalpy (energy of incorporation) values for the genome structures, showing where high levels of energy incorporation have occurred and how this relates to the evolutionary change of the organisms. What we have developed is a method of calculating both entropy (eS) and enthalpy (eH) values for a given sequence by taking advantage of this evolutionary fact.

Even though the project is heavily influenced by biology. The main subject matter of the visualization falls into a category of social navigation, where we define social navigation in terms of collaboration, an ideology of "learning-in-doing [3]," and application of affordance. Simplistically, getting the meaning of the data you are trying to represent without the necessity of explanation. The social aspect takes form in that the user has the ability to annotate information to the visualizations and the visualization will alter every time there is an annotation made to direct the user to most important points of interests. Also, because of the novelty of Entromics there is an added component, best described as networks, that directs the user to other areas of interests that they may not be aware of.

The paper is situated as followed: section 2 covers the visualizations themselves, section 3 looks at the usability study both pre and post

testing, and section 4 ties this together with its significant and future works.

## Section 2 – Constructions of visualizations

There were two platforms created one with a 2D plot representation and one with a 3D plot representation. See appendix (*Preliminary Survey*) for complete survey. These were both tested against one another to find the most optimal solution.

### 3D Visualization

The 3D visualization was developed in C++ using the libraries: OpenGL, OpenInventor, and QT and produces a 3D plot of the contour (so no data is lost). Figure 1 shows the genome representation, which was idea presented in MizBee [4], and a movable slider that alters the genome visualization. Figure 2 shows the network of correlations between various positions around the genome sequence (Entromics novel contribution).

### 2D Visualization

The 2D visualization was created utilizing a spatial database (OpenGIS [5]) and processed using AJAX and PHP to create SVG visualizations in very short amount of time (roughly seconds to 1 minute). Figure 3 (in Appendix) shows an example output of the visualization. The 2D approach takes the original three dimensional matrix output of the Entromics application and cuts it into layers of even size. Areas above a certain threshold are represented as “islands” and will change as more annotations are added. The green notches represent the positions with information annotated to the sequence. For both the “islands” and notches, the color variation showcases the amount of information available with the prognosticated that research will explorer areas of higher intensity. Figure 4 shows the network created and the resulting information when queried.

## Section 3 – Usability Testing

We looked at both the front-end analysis and the prototype usability studies to create the most optimal design for our applications. The front-end analysis was completed using surveys which were sent out to 10 experts/5 non-experts. We classify our experts as PhDs/MDs in Clinical Pharmacology or Computational Biology at the University of Pittsburgh and the non-experts as students/grad students in Clinical Pharmacology at the University of Pittsburgh. We received 8 experts’ surveys and 2 non-experts’ surveys at the conclusion of this paper. The surveys covered topics such as:

- How navigating in a 2D or 3D space can vary user experience when searching for information (points of interests)
- How best to represent the genome sequence
- What type of information they would like to annotate to the sequence (text, papers, ratings, etc.)
- Who would be using this application and do they have experience with other types of applications like ours

The results of these studies produced some interesting preliminary results such as:

- When navigating a 2D or 3D plot, they would prefer 2D. However, for finding POIs on a 2D or 3D plot, they would prefer 3D
- When it came to finding applications that allowed researchers to view, edit, or annotate visualizations for genome sequence. There was zero overlap between any responses (everyone uses different applications for different functions).

For the prototype testing we got our high-fidelity prototypes in front of four expert users. Each user was shown how to use each prototype and

Each user was given 5 tasks for each prototype. After each task, user was able to voice their opinion about the experience. Users were shown 2 alternative designs and asked to pick their favorite of the 3 (see Figure 5) and a survey was given after all tasks were completed (Appendix: *Post-Prototype Usability Survey*). Some of the interesting results/comments are as followed:

- “2D version would be easier to learn, but the 3D version may be more beneficial in the long run”
- All the users were able to complete all the tasks
- Even though they liked both designs, most agreed that a hybrid (2D and 3D) would be the best final design
- We need to be careful not to create something that may be more visually pleasing, but misleading in terms of the biological property

#### Section 4 – Conclusion and Future Works

Overall, we are pleased with the preliminary results of the visualizations. Users were able to complete tasks associated with the design and implementation and find important bits of

information on their own without the necessity of being told directly where to find them.

From here we are looking to incorporate the suggestions made about the design to create a more intuitive and domain specific visualization for a genomic sequence.

#### References

1. Farzan, R. and P. Brusilovsky, *AnnotatEd: A social navigation and annotation service for web-based educational resources*. New Review of Hypermedia and Multimedia, 2008. **14**(1): p. 3-32.
2. Petr Pancoska, R.A.B., Patrick M. Dudas, *Discovery of  $\tau$ -homology in a set of sequences and production of lists of  $\tau$ -homologous sequences with predefined properties.*, U.o.P.C.S.o.H. Education, Editor. 2010.
3. Pea, R., *The collaborative visualization project*. Communications of the ACM, 1993. **36**(5): p. 60-63.
4. Meyer, M., T. Munzner, and H. Pfister, *MizBee: a multiscale synteny browser*. IEEE Transactions on Visualization and Computer Graphics, 2009. **15**(6): p. 897-904.
5. (OGC), O.G.C. *OpenGIS*. 1994 [cited 2010 December 16th, 2010].

# Appendix

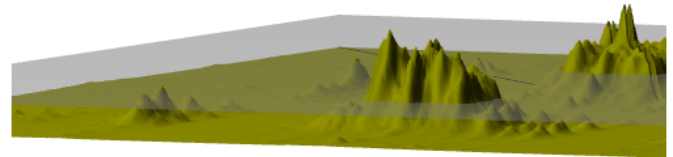
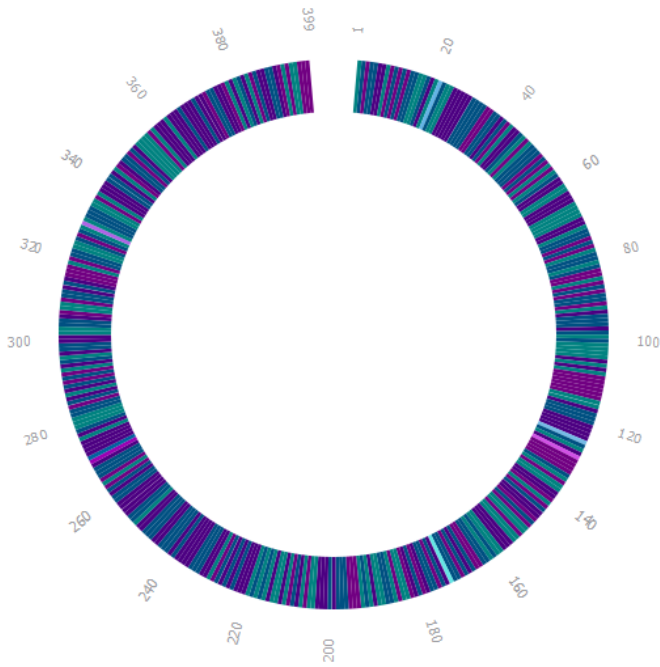


Figure 1 - 3D Contour Visualization

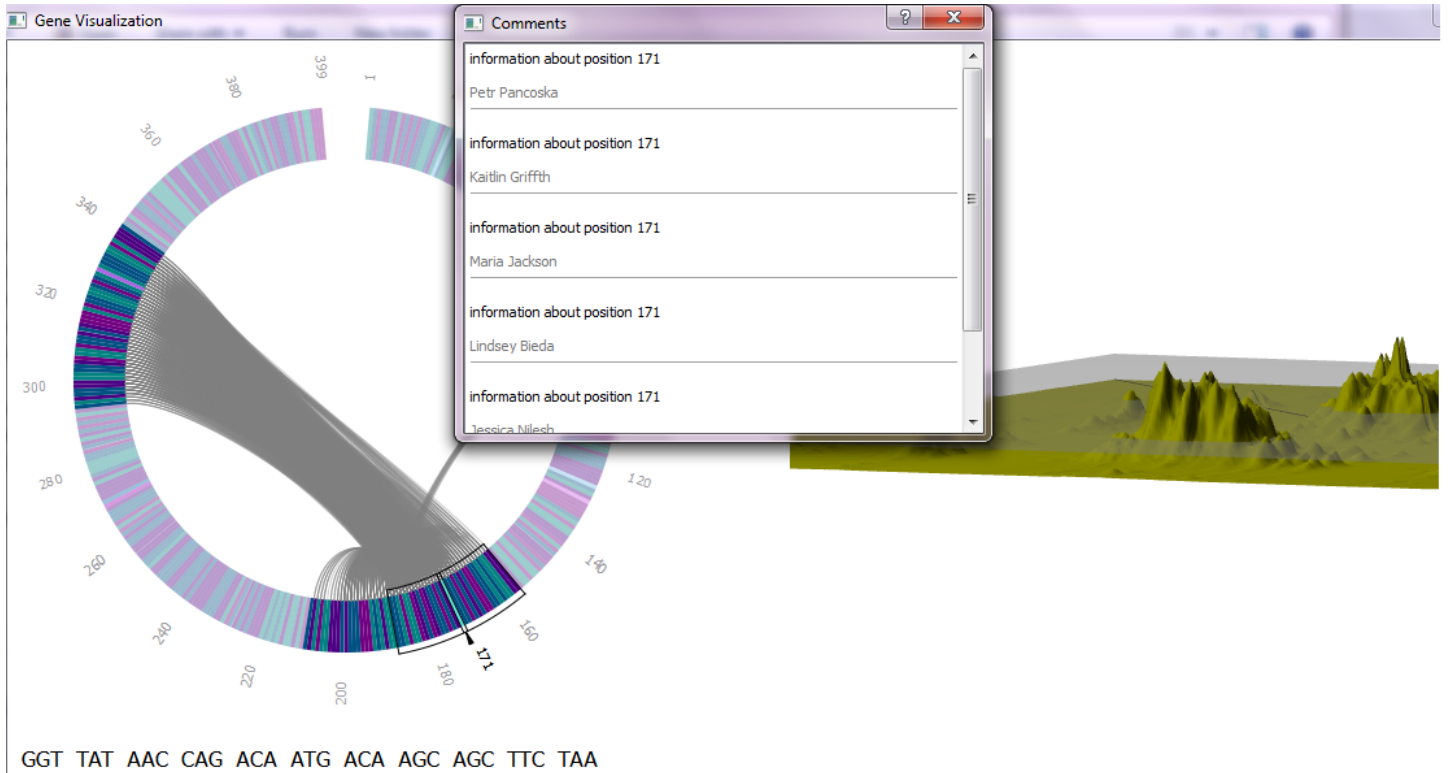


Figure 2 - Network and Resulting Information

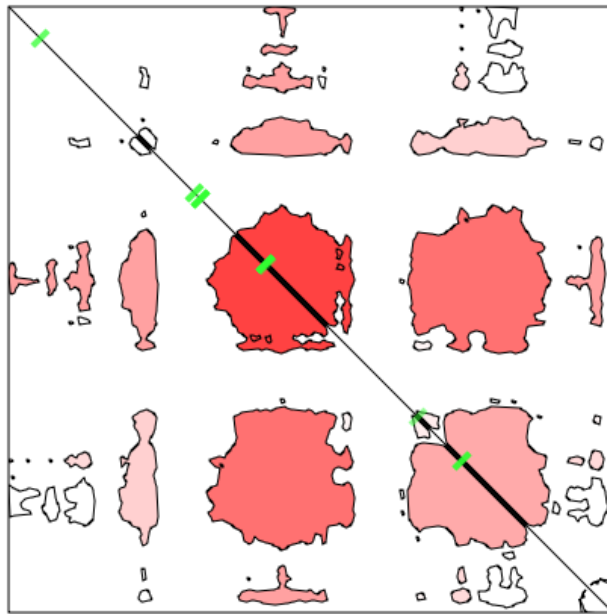


Figure 3 - 2D Layer Visualization

Query Results - Mozilla Firefox

http://entromics.ccp.pitt.edu/2DContours/

Positions	Sequence	Information	First Name	Last Name
300	ACAGTGACACTGATGAAGGTTAAGCTTCAAGAG	300	tomoko	nukui
170	GGGTTATAACCAGACAATGACAAGCAGCTTCTA		bob	branch
300	ACAGTCAGACTGATGACAGCTTCAAGCTTCAAGAG			

Stopped

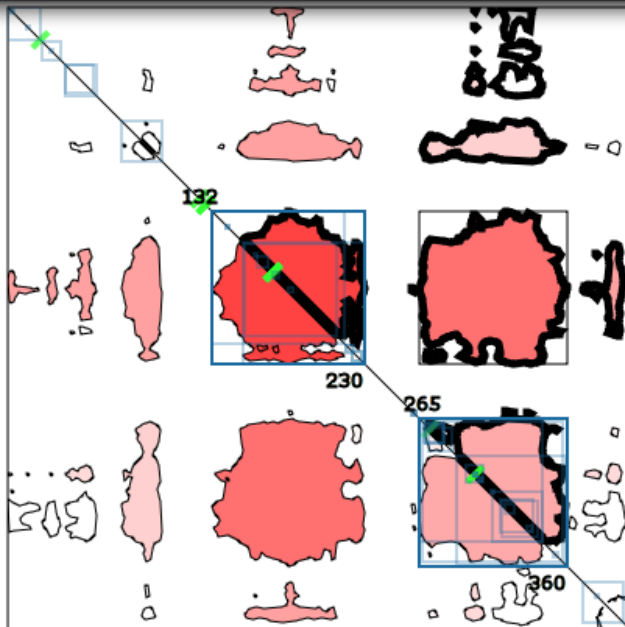


Figure 4 - Network and Resulting Information

Preliminary Survey:

What we are looking to build:

We are designing a visualization for representing a 3D matrix plot of genomic coherences. This would provide a weighted links (or network) between genome positions and most importantly non-local long-range comparisons of multiple distant genome regions. The novel genome characterization is a 3D matrix of incorporation energies. Visualization (both in 2D and 3D) will allow researchers to annotate pertinent information about all genome regions up to the single base resolution. The visualizations will (besides biological/genomic information) also represent the links/networks and places/regions along the genome sequence that has appended annotations (this will also show where most of the information is being appended relative to all other positions).

**Ratings:**

**Rate from 1 (Strongly Disagree) to 5 (Strongly Agree)**

1. I would consider myself someone who is knowledgeable with visualizations (as in a visual representation of data)

Strongly disagree = 1 2 3 4 5 = Strongly Agree

2. I would consider myself someone who is knowledgeable with interactive visualizations (as in a visual representation of data that allows you to interact, change, manipulate, or alter to learn new information from the visualization)

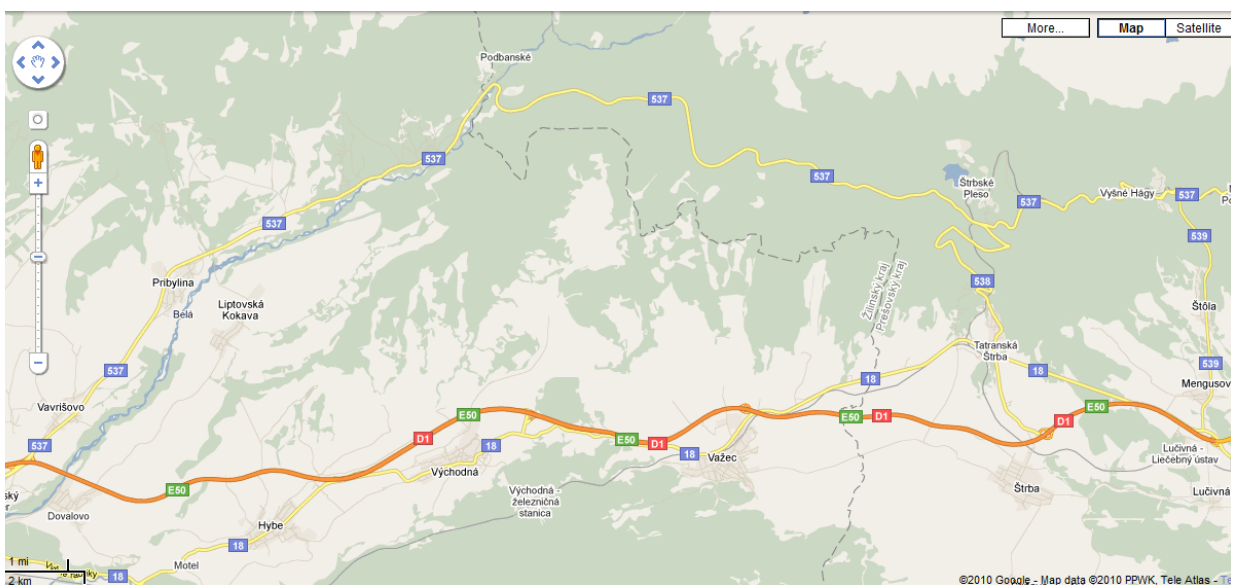
Strongly disagree = 1 2 3 4 5 = Strongly Agree

3. I would consider myself someone who is knowledgeable with collaborative visualizations (as in a visual representation of data that allows you to annotate and search for information embedded in the visualization for as well as with other collaborators, researchers or colleagues)

Strongly disagree = 1 2 3 4 5 = Strongly Agree

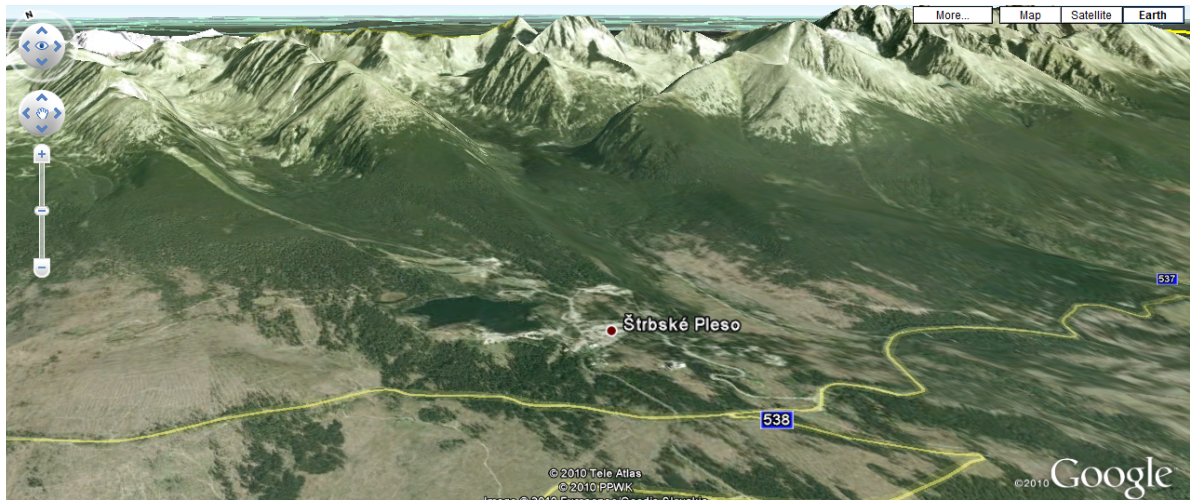
4. I would consider myself someone who is knowledgeable in 2D spatial representations of spatial data (as in Google Maps, MapQuest, Bing Maps, etc.) [please see figure below for example]

Strongly disagree = 1 2 3 4 5 = Strongly Agree



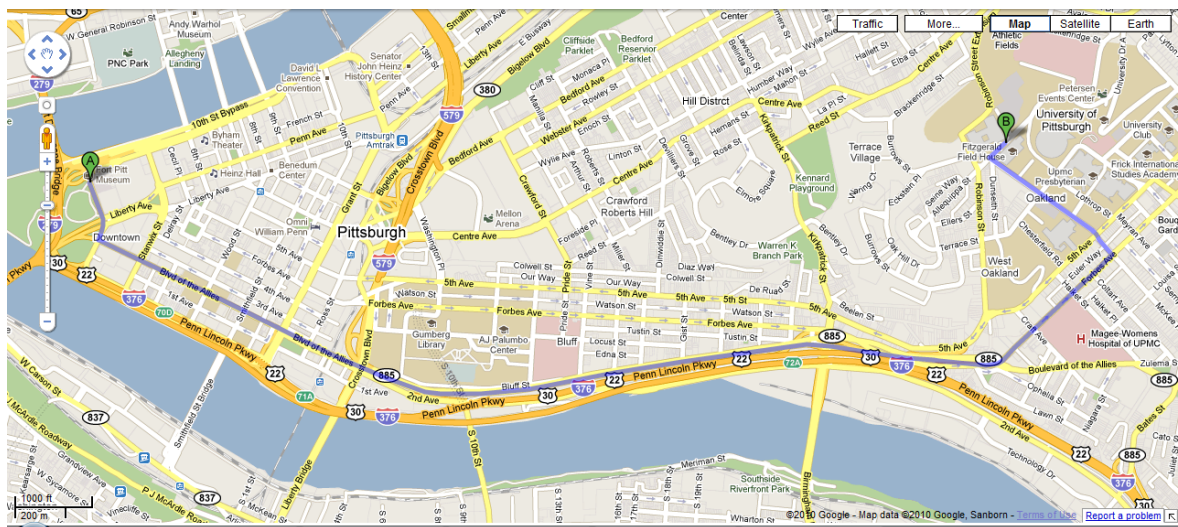
5. I would consider myself someone who is knowledgeable in 3D spatial representations of spatial data (as in Google Earth) [please see figure below for example]

Strongly disagree = 1 2 3 4 5 = Strongly Agree



6. If I had to view a map route, I would prefer using a 2D spatial representation (as in, Google Maps, MapQuest, Bing Maps, etc.) [please see figure below for example]

Strongly disagree = 1 2 3 4 5 = Strongly Agree

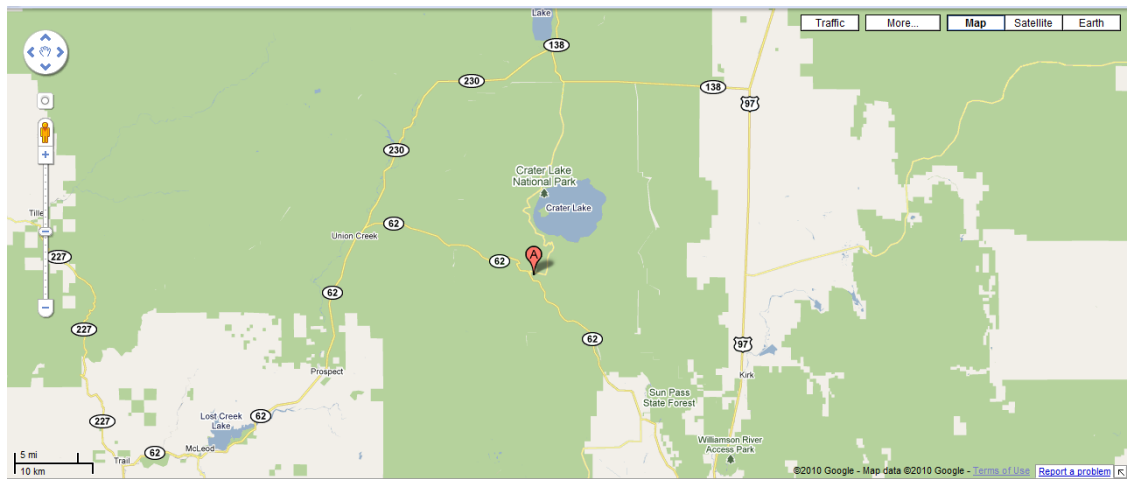


7. If I had to view a map route, I would prefer using a 3D spatial representation (as in Google Earth) [please see figure below for example]

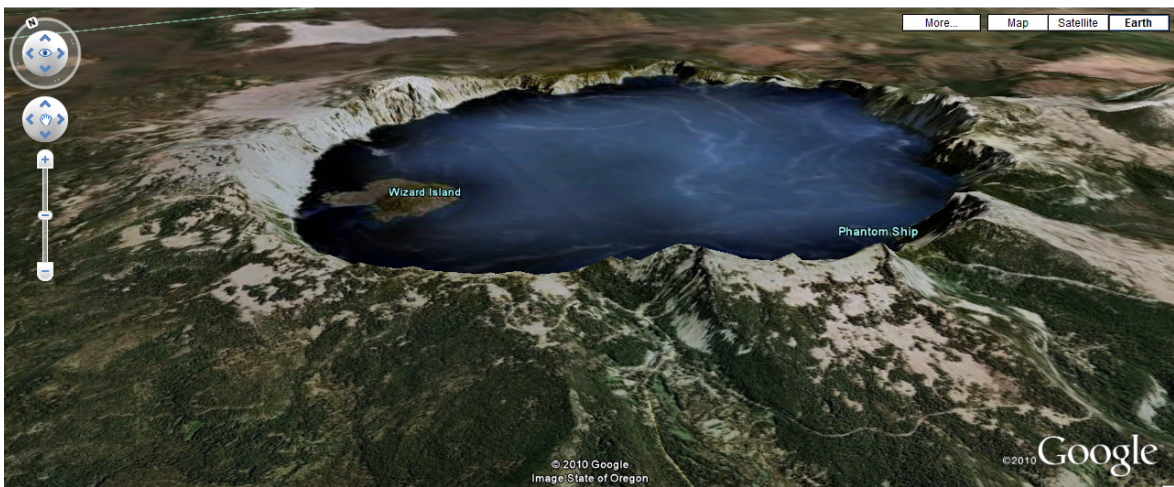
Strongly disagree = 1 2 3 4 5 = Strongly Agree



8. If I had to find points of interests, I would prefer using a 2D spatial representation (as in Google Maps, MapQuest, Bing Maps, etc.) [please see figure below for example]  
 Strongly disagree = 1 2 3 4 5 = Strongly Agree

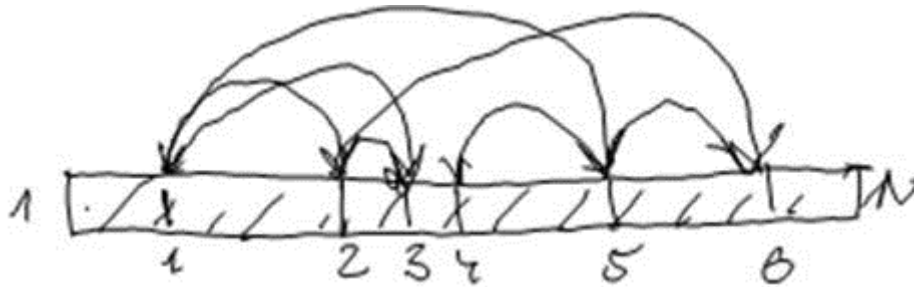


9. If I had to find points of interests, I would prefer using a 3D spatial representation (as in Google Earth)  
 Strongly disagree = 1 2 3 4 5 = Strongly Agree



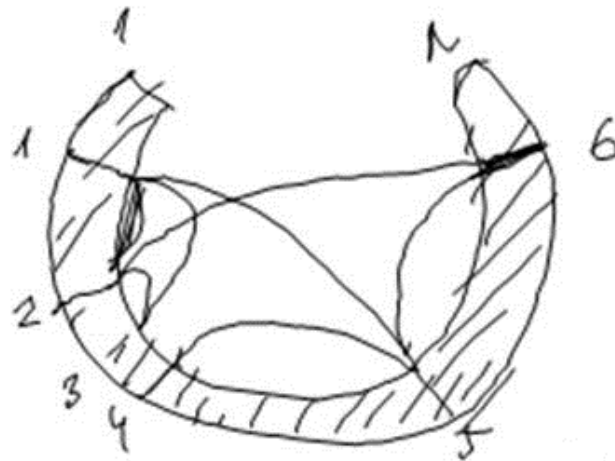


10. If I had to view connections between various points of a sequence, I would prefer a linear representation [please see figure below for example]



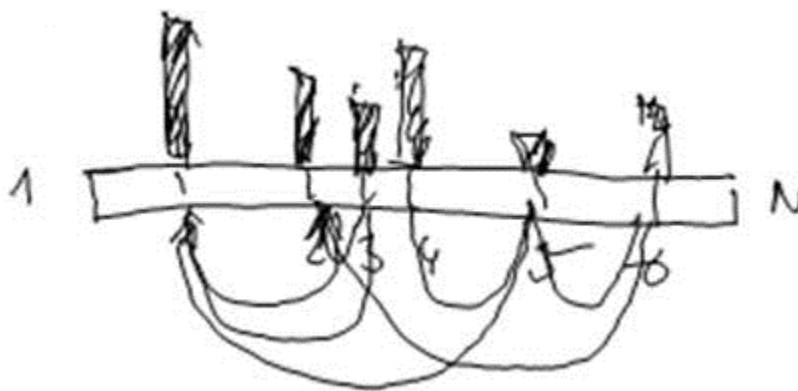
Strongly disagree = 1 2 3 4 5 = Strongly Agree

11. If I had to view connections between various points of a sequence, I would prefer a circular representation [please see figure below for example]



Strongly disagree = 1 2 3 4 5 = Strongly Agree

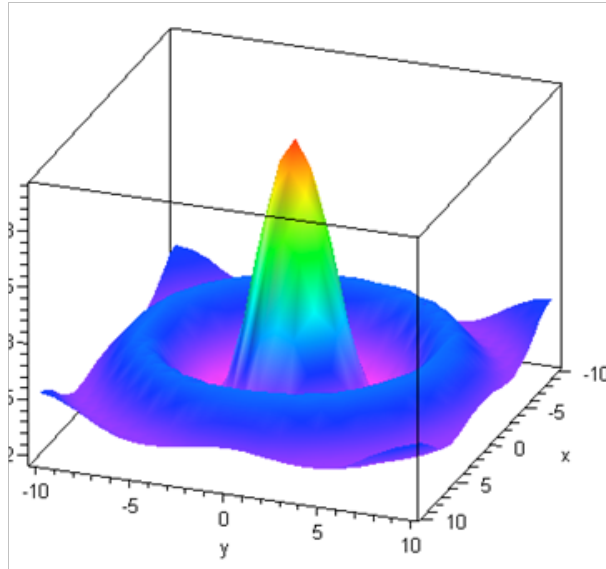
12. If I had to view connections between various points of a sequence, I would prefer a discrete representation [please see figure below for example]



Strongly disagree = 1 2 3 4 5 = Strongly Agree

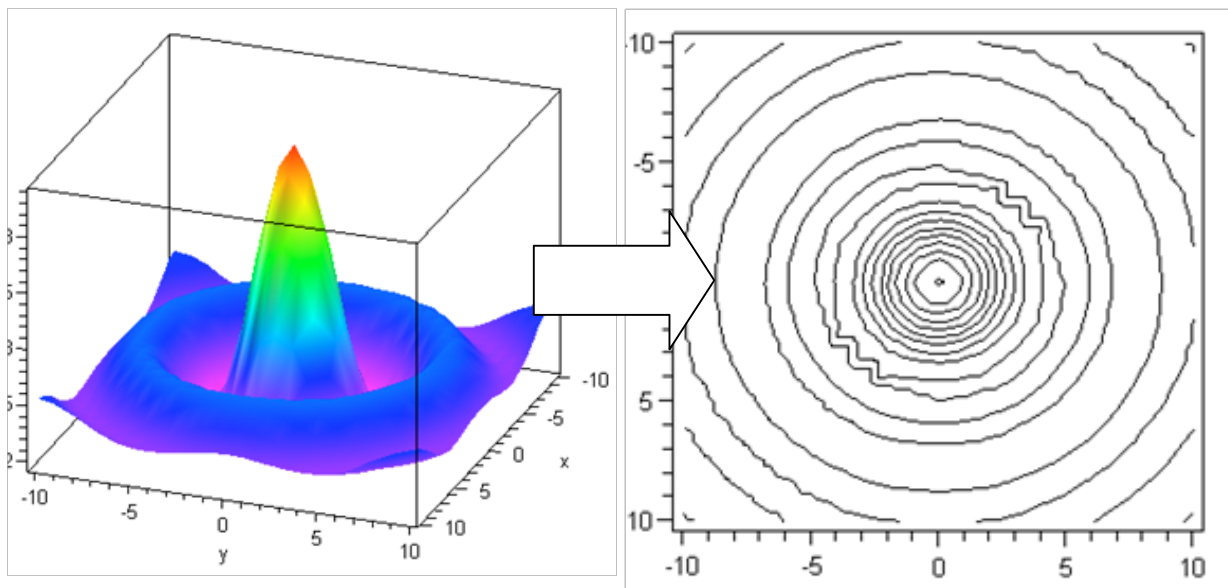
13. If I had to view 3D plot, I would prefer to view this in its original 3D representation [please see figure below for example]

Strongly disagree = 1 2 3 4 5 = Strongly Agree



14. If I had to view 3D plot, I would prefer to view this as 2D plot representation of the 3D plot [please see figure below for example]

Strongly disagree = 1 2 3 4 5 = Strongly Agree



15. If I had to find a point of interest on a 3D plot, I would prefer to find this point on its original 3D contour form

Strongly disagree = 1 2 3 4 5 = Strongly Agree

16. If I had to find a point of interest on a 3D plot, I would prefer to find this point on 2D plot representation of the 3D plot

Strongly disagree = 1 2 3 4 5 = Strongly Agree

17. When I am using a new application or web application, I read the instructions before using to get my desired results

Strongly disagree = 1 2 3 4 5 = Strongly Agree

18. When I am using a new application or web application, I just use the application or web application to learn how to get my desired results

Strongly disagree = 1 2 3 4 5 = Strongly Agree

19. When I am using a new application or web application, I usually have a desired goal in mind before attempting to use the application or web application

Strongly disagree = 1 2 3 4 5 = Strongly Agree

20. When I am using a new application or web application, I usually try the application or web application before determining a desired goal

Strongly disagree = 1 2 3 4 5 = Strongly Agree

Pick one:

21. If I had a choice in using either a 2D or 3D spatial representation for finding a route from point A to point B, I would prefer which?

22. If I had a choice in using either a 2D or 3D spatial representation for finding points of interests I would prefer which?

Specific Biological Question

23. What type information are you looking to append to the visualization (example: text, papers, ratings, etc.)?

24. If only given the option to annotate information about genome by position number, would this be sufficient?

Yes or No

If yes, please explain in which context you are referring to:

25. If only given the option to annotate information by genome sequence (ATTATCATAT), would this be sufficient?

Yes or No

If yes, please explain in which context you are referring to:

26. If you were provided a platform to retrieve annotations about a genome sequence, would you prefer information that strictly tailored to your research or any information available related to that sequence?

Specific Information or Any information about the sequence

27. Are you currently using an application or applications that allow you to view genome sequences visualization (not just strictly text)?

Yes or No

If yes, please let us know the name of this application or applications:

28. Are you currently using an application or applications that allow you to view and manipulate a genome sequences visually (not just strictly text)?

Yes or No

If yes, please let us know the name of this application or applications:

29. Are you currently using an application or applications that allow you to annotate information about a specific single nucleotide in a collaborative manner (with researchers other than yourself)?

Yes or No

If yes, please let us know the name of this application or applications:

**Post-Prototype Usability Survey**

**Ratings:**

**Rate from 1 (Strongly Disagree) to 5 (Strongly Agree)**

- 1. I found the 2D Layer approach intuitive and easy to use  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
- 2. I found the 3D Contour approach intuitive and easy to use  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
- 3. I was able to find information easily using the 2D Layers

- Strongly disagree = 1 2 3 4 5 = Strongly Agree
4. I was able to find information easily using the 3D Contour  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
5. I would use the 2D Layer visualization again for annotating information about a genome sequence  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
6. I would use the 3D Contour visualization again for annotating information about a genome sequence  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
7. I preferred the 2D Layer visualization over the 3D Contour visualization in terms of appearance  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
8. I preferred the 2D Layer visualization over the 3D Contour visualization in terms of annotating information  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
9. I preferred the 2D Layer visualization over the 3D Contour visualization in terms of finding information  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
10. I found that the 2D Layer visualization was a good way to understand the network connection between various parts of the sequence  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
11. I found that the 3D Contour visualization was a good way to understand the network connection between various parts of the sequence  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
12. I found that the 2D Layer visualization's tasks were easy to complete  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
13. I found that the 3D Contour visualization's tasks were easy to complete  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
14. I found that annotating information by position number easy to understand  
Strongly disagree = 1 2 3 4 5 = Strongly Agree
15. I found that annotating information by genome sequence easy to understand  
Strongly disagree = 1 2 3 4 5 = Strongly Agree

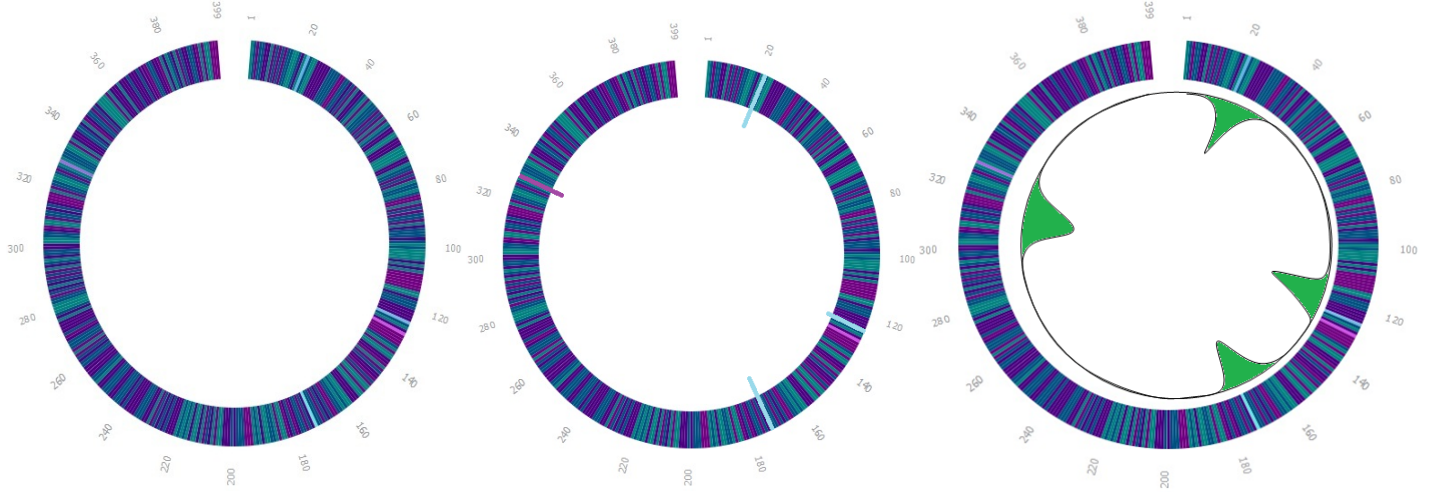
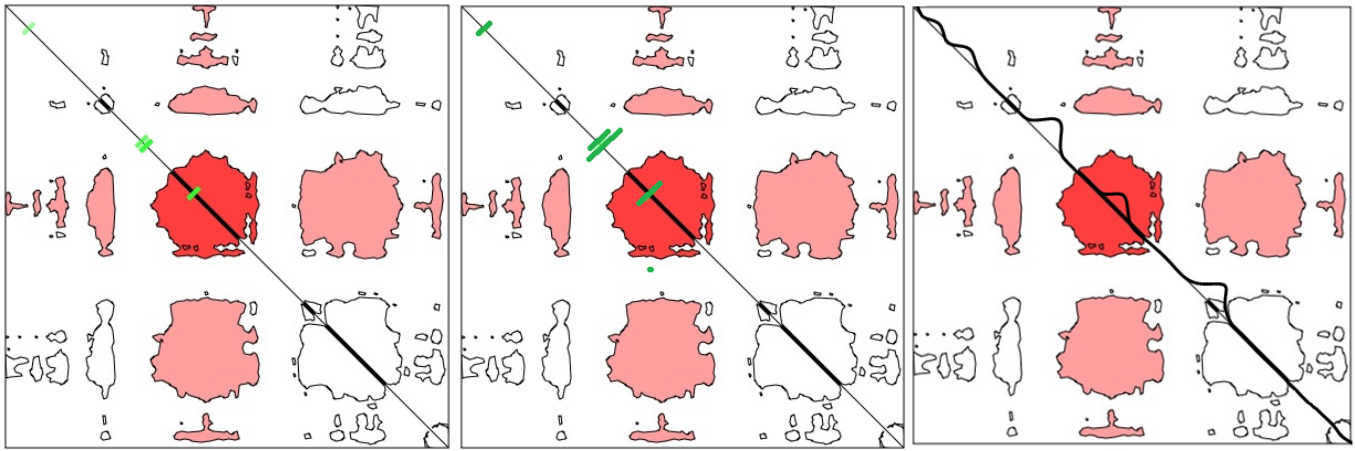


Figure 5 – Optional Display Choices