

## Contents

- [1 Which components have you used?](#)
- [2 Documentation](#)
- [3 Use Cases](#)
- [4 Features](#)
  - ◆ [4.1 Expansion / Clarification](#)
  - ◆ [4.2 Other High Priority Features](#)
  - ◆ [4.3 Other Medium Priority Features](#)
  - ◆ [4.4 Other Low Priority Features](#)
- [5 Other Feedback on Visualization Tools](#)
- [6 Other Feedback](#)

The 2009 GMOD Community Survey focused on genome and comparative genomics [visualization](#). The survey was open for 10 days in September 2009 and received 45 responses.

*For a broader picture of the GMOD community and how [GMOD Components](#) are used, see the [2008 GMOD Community Survey](#).*

## Which components have you used?

GMOD has a wealth of genome and comparative genomics browsers. Which of the following have you used?

Component	%
<a href="#">GBrowse</a>	87%
<a href="#">CMap</a>	31%
<a href="#">JBrowse</a>	29%
<a href="#">GBrowse_syn</a>	29%
<a href="#">Apollo's synteny viewer</a>	13%
<a href="#">BLAST Graphic Viewer</a>	13%
<a href="#">SynBrowse</a>	7%
<a href="#">GBrowse karyotype</a>	7%
<a href="#">SynView</a>	4%
<a href="#">Sybil</a>	2%

## Documentation

How satisfied are you with the documentation for these components?

**Administration Documentation** - e.g. [GBrowse Configuration HOWTO](#), [GBrowse syn Tutorial](#), ...

	%
Very	50%
Average	35%
Not at all	5%
No Opinoin	10%

**End user Documentation** - e.g. [GBrowse User Tutorial](#), [CMap Tutorial](#), ...

	%
Very	45%
Average	33%
Not at all	3%
No Opinoin	20%

**Overview Documentation** - e.g. [Comparative Genomics](#), [Overview](#), ...

	%
Very	23%
Average	40%
Not at all	0%
No Opinoin	38%

## Use Cases

The aim of this section was to determine what types of questions people want to ask with visualization software.

**What's an example question that you would like to answer using genome or comparative genomics browsing software? Leave this question blank to skip to the next section of the survey.**

**To the best of your knowledge, which if any, existing tools support this type of question?**

What is the distribution of transposable elements (TEs) relative to the gene coding regions of the genome. What is the global distribution of genes and TEs in the genome (*ie.* a heatmap based view). I would like to be able to look at heatmap distributions using multiple algorithms for partitioning the data by color (*ie.* equal interval, quantiles etc.).

Difficult to do this easily with existing tools.

Display deep sequencing paired reads in a pileup that is vertically sorted by length of the matching pair. I think the [AJAX](#) code in [Lookseq](#) could be integrated into [JBrowse](#) or

[Lookseq](#)

Gbrowse2.

Within a genome I want to be able to make it easy to visually find areas of the genome where two or more features overlap or coincide e.g. a QTL and a gene

GBrowse

Compare genomes between related species.

I would like to be able to see SNP existing between a reference genome strain and other genome strains of interest.

I think it could be possible using GBrowse

My request is actually more general then that. I'd like some way to say :

- show me all regions holding this feature (*ie* : a protein annotation) AND holding also this kind of features (*ie* : SNP, QTL, ....)

None, but I'm only very familiar with GBrowse.

We want to visualize high throughput data coming from cgh, or expression array and also from short read technologies. the organisms we study are not always well annotated. we want to aggregate annotation data, gene models from other sources.

Gbrowse2 with Bio::DB::Sam is a good start point even if the documentation about the connection between GBrowse and the perl module is too short. I intend to install MAKER for the annotation part and its connection to Gbrowse2.

in the end, the ideal would be the possibility of querying efficiently the browser on numerical data.

I know about the expression of DNA (for instance I have the mRNA or protein sequence that is expressed) and I would like to view the stretch of DNA this maps to while aligning cDNA, EST, predictor, promoter, etc. information

Apollo

To my knowledge there are a few tools that can do this, like the commercial ERGO package, Genoscope has these capabilities and the Microbial Genome Viewer as well (although MGW lacks other functionalities that make GBrowse unique)

In bacterial genomics we often look at regions of many (>20) strains and species stacked above each other, aligned at a gene of interest based on selections or homology searches. These regions may consist of inter AND intra species locations.

How can we view repetitive regions using GBrowse\_syn?

Ability to easily leverage data from CMap and GBrowse for comparative genomics

Gbrowse Sythn (GBrowse\_syn?)

Good multiple genome support

SynBrowse, but not very well

I want to find lots of annotation and context for genes of interest from new papers or new data discoveries.

UCSC Genome Browser, GBrowse (varies by species), Ensembl, Map Viewer

I have a few.

None that I know of

- I'd like to be able to move a back and forth between a phylogenetic tree viewer where I can select a clade (which may include paralogs) and then see the selected members of this clade in a genomic context. And vice versa.
- I'd also like to be able to see a protein multiple sequence alignment and select a region or set of and then see where these regions are overlaid onto the 3-D structure

of the protein. (i.e. where each of the proteins have been threaded into a 3-D protein rendering of this family).

- I'd like to have some visual clues indicating common/shared functions for genes in a given syntenic region. Do they appear in the same order?

What insight does the coverage of next-generation sequence data give regarding repetitive elements within the genome?

- Genome visualisation (very specialised tracks)
- NGS support
- Comparative maps support

I would like to integrate the information from physical and genetic maps as well as the genome sequence. Including BACs, markers, BAC-ends, unigenes, etc.

Scalable view of genomes side by side, linked by markers, locations, or features (genes)

I like to easily visualize microarray and next generation sequencing data from chromatin IP experiments across the genome and in relation to genomic features. Adjusting data plot parameters, order, and graph type spontaneously is important.

I would like to compare on-going sequencing projects (*e.g.* 'an incomplete genome, chromosome or plasmid) with closely related finished and annotated sequences (*e.g.* a finished genome). It would be great to see the reference genome with annotation and the pieces of the unfinished project together, to manually infer putative genes, for example.

If possible, I'd like a genome browser that can show sequence similarity (i.e. multiple sequence alignments) between portions of genomes.

Is the region that is transcribed according to tiling array data associated with published data (such as siRNAs, DNA methylation, chromatin modifications etc)?

I'd like to map all types of information onto the genome in the browser, eg Microarray data. But would also like to see it mapped to multiple genomes with comparative displays Much like ACT (from Sanger) can do, but in a web interface

We would like to visualize where EST contigs sequenced from a novel genome align to the genome of a model organism.

Given a class of transposable elements, how does the distribution of these elements differ between two genomes. This class could be very broad such as DNA transposons, or it could be a superfamily of DNA TEs or an individual family.

This sort of data is quite complicated to view using GBrowse currently.

GBrowse, CMap

Chado, GBrowse and CMap.

GBrowse

Multiple competing genome browsers are capable of this function, but I am most familiar with GBrowse.

May be with some of the synteny-aimed software, but this is not their scope.

I'm not aware.

GBrowse

ACT, although this is a standalone tool

I would assume that any genome browser should be able to do this relatively well. While I have used GBrowse, JBrowse, UCSC Genome Browser, and Apollo, I have used only GBrowse extensively.

GBrowse\_syn

With multiple genomes I want to make it easy to find regions where there has been some sort of rearrangement (indel or inversion or translocation)

I would like to be able to enter a AA sequence and then get all tblastn (not really a blast search allowing gaps, but more a typical substring search) hits and have all those hits displayed within their neighborhood.

I want to quickly find SNPs in regions of interest, and have them color coded based on my own criteria.

Ability to add arbitrary number of genome elements/features, choices to make them permanently or privately part of a database

Where are the start codons, according to the conservation in a clade?

I mean, to be able to determine mispredicted start codons based solely in previous annotations and the conservation of all the potential start codons (off course, assuming this is a good criterion based on the previous knowledge of the protein(s) analysed).

I'd like to show members of one gene family in different genomes as a stack on a synteny viewer. With the functional domains (in the sequence) highlighted.

Right now I am working on integrating GBrowse system to my gene prediction program. So user will be able to visualize the complete genome picture according to their input raw sequence.

Would like to get guidance for the annotation of a novel sequence by showing similarity to a well-annotated sequence

none- Typically I have to come up with the coordinates of genome hits myself and then get it displayed somehow.

UCSC Genome Browser

GBrowse

Any with comparative capabilities, but I ask because is great!

I'm still exploring GBrowse\_syn so maybe it can be customized to do this.

I was very happy with previous work and I did not look for alternative.

Apollo?

## Features

This section asked participants to prioritize features in browsers.

For each feature, please indicate that feature's importance to you. Please try to classify no more than 1/3 of the features as high importance.

Features are listed in the order they appeared in the survey. *You can resort the table by clicking on a column header.*

**Key: High Medium Low Not at all No opinion**

### Feature

Feature	High	Medium	Low	Not at all	No opinion
<b>Browser response time (speed!)</b>	71%	22%	4%	2%	0%
<b>Data loading speed.</b>	31%	33%	29%	2%	4%
How long it takes to process and load data into backing databases.					
<b>Browser install and setup Script</b>	20%	31%	29%	4%	16%

The GBrowse NetInstaller is an example.

**Graphical user interface for administering the browser.**

24% 24% 28% 7% 7%

Update configuration, add tracks, load data, ... through a GUI.

**Configuration file checker with helpful error messages.**

31% 51% 9% 0% 9%

**User management**

27% 29% 24% 4% 16%

Allow users to login. This would enable other functionality.

**Community annotation**

38% 40% 18% 2% 2%

Support users adding annotation to individual features and/or uploading features or tracks for sharing with others.

**Package browser software within a ready-to-install virtual machine that includes several other commonly used GMOD components.**

16% 33% 24% 11% 16%

For example, see the [community Annotation System](#).

**Make browser instance metadata available via web services**

13% 36% 20% 7% 24%

See [this page](#) for an explanation of how this might be done in GBrowse.

**A public repository of browser-ready reference genomes, including example annotations such as gene models, NGS data, quantitative data (wiggle), ...**

29% 44% 20% 0% 7%

[GBrowse.org](#) is a step in this direction.

**Extensibility**

27% 38% 24% 0% 11%

Support for plugins and user defined glyphs

**Individual feature display customization**

Allow browser admin to write their own code to adjust how a feature is shown (height, color, border, ...), based on the feature's attributes. (This is done with Perl callbacks in [GBrowse](#).)

42% 36% 16% 2% 4%

**Individual base display customization**

Allow browser admin to write their own code to adjust how an individual base is shown (height, color, border, ...) at run time, based on the base's attributes. This could show the alignment quality, or coverage, or ... for next generation sequencing data.

22% 36% 27% 4% 11%

**Admin control of browser layout.**

The browser admin configuration of what sections (e.g., search box, instructions, etc.) appear and where, what text appears and where, and so on. GBrowse allows admins to control some aspects of the layout.

20% 42% 24% 4% 9%

**Hierarchical listing of available tracks.**

31% 29% 22% 0% 18%

<u>GBrowse</u> already supports this.					
<b>Show multiple regions simultaneously</b>	29%	40%	18%	0%	13%
Select and then show multiple regions of the genome.					
<b>Comparing two or more genomes.</b>	49%	24%	7%	2%	18%
<u>GBrowse_syn</u> , for example, does this.					
<b>Whole genome/chromosome browsing</b>	22%	38%	13%	4%	22%
<i>e.g.</i> , <u>GBrowse karyotype</u>					
<b>Browsing on mobile devices</b>	4%	7%	16%	44%	29%
<b>Semantic zooming</b>	22%	27%	27%	0%	24%
<b>Autocompletion of Search Terms</b>	18%	40%	18%	11%	13%
<b>Popup Balloons</b>	20%	36%	24%	7%	13%
<b>Rubber Band Selection</b>	27%	40%	16%	0%	18%
<b>Linkage disequilibrium tracks</b>	20%	18%	18%	4%	40%
<b>Support <u>next generation sequencing individual reads</u>.</b>					
Visualize the NGS short reads themselves, showing items like read quality for individual bases.	53%	24%	11%	0%	11%
<b>Display markers from <u>CMap</u> in the genome browser.</b>					
See <a href="#">this page</a> for an explanation of how this might be done in <u>GBrowse</u> .	22%	18%	20%	2%	38%
<b>Quantitative data shown with color intensity</b>					
<i>i.e.</i> , wiggle_density tracks	53%	20%	11%	2%	13%
<b>Quantitative data shown on an x-y graph</b>					
<i>i.e.</i> , wiggle_xyplot tracks	49%	22%	11%	2%	16%
<b>Log scaling for quantitative data</b>	36%	20%	16%	2%	27%
<b>Show multiple datasets in a single quantitative track.</b>					
Data in color intensity tracks (wiggle_density) could be stacked; data in x-y tracks (wiggle-xyplot) could be superimposed (may be require multiple scales), or stacked.	42%	18%	20%	2%	18%
<b>Aggregation functions for quantitative data</b>					
<i>e.g.</i> , show mean, max, min, across all data or sliding windows.	44%	9%	22%	2%	22%
<b>Alignment tracks.</b>					
Showing insertions, deletions, ...	33%	33%	9%	4%	20%
<b>Geolocation data</b>	11%	11%	29%	13%	36%
how things like genotype and allele frequencies phenotypes,					

environment, ... by geolocation.

**Key: High Medium Low Not at all No opinion**

## Expansion / Clarification

If you want to explain/expand any of your answers above, please do so here.

- Geolocation within genomic space would be interesting but not so interested in ecological applications quite yet.
- I think basic elements like speed to load a region in GBrowse and documentation of track configuration and additional data type adding are core requirements that could be made better. Although I haven't used GBrowse 2 it sounds like the performance element is being addressed.
- User GBrowse v2 Tutorial?
- part of my job takes place on an academic expression platform. I have to deal with expression microarray analyzes coming from 2 colors arrays and also from RNA-Seq and ChIP-Seq experiments. We use to give our users text files (in the time of small format arrays) but now, with tiling arrays, our users, who are big fans of bioinformatics skills, they cannot open their files with conventional text tools. we intend to give them GFF3 files so that they can visualize the data on GBrowse locally or on our platform. The problem is that we have to ensure the confidentiality of their data. That's the reason I am very interested in a better User management.
- Perl callbacks are not difficult to use when well documented. It's the major problem of bioperl to me. There are many modules but not enough documentation with concrete examples. So a big YES for ""Allow browser admin to write their own code to adjust how a feature is shown"" with better doc.
- It would be great to merge two of the questions: the comparison feature but with single NGS reads. It might be something like AMOS' hawkeye, but by web! I think the current engine can be adapted to support this, and it would be great to be able to manually curate some sequence with weird features (for example, an unexpected stop codon) based in both the summarised quality (already supported by quantitative tracks) and the original reads.
- no

## Other High Priority Features

Are there other high priority features you would like to see that are not in the list above?

- Ability to curate features that are not genes. The majority of genomes are not composed of genes, and the weakness of most browsers is their inability to display features that are not genes. I would especially like better support for the display and curation of of Transposable Elements.
- Support for displaying/summarizing alignments in BAM format.
- About 'support next generation sequencing individual reads', it is beautiful if SAM (or other format) has got an 'attributes' field.
- bridging interface between the different tools, (common data format?)
- Improve the chado examples for different use cases.
- More customization of data tracks by end user, not just by the admin. For example, in GBrowse (v1.69) XY data plots, generic track attributes may be adjusted by the end user, but not specific XY plot attributes, such as min and max values, etc.
- The previous answer also counts for this one. *The previous answer was:*
  - ◆ It would be great to merge two of the questions: the comparison feature but with single NGS reads. It might be something like AMOS' hawkeye, but by web! I think the current engine can be adapted to support this, and it would be great to be able to manually curate some sequence with weird

features (for example, an unexpected stop codon) based in both the summarised quality (already supported by quantitative tracks) and the original reads.

- not really

## Other Medium Priority Features

Are there other medium priority features you would like to see that are not in the list above?

- Ability to generate heatmap color bin from a set of different algorithms (equal interval, quantiles ..).
- Native Chado database connection instead of adapter type chado connection (although the latter may be more flexible, speed is an issue)
- A much more difficult one, but nice enough to mention: sequence and/or annotation editing capabilities.
- not really

## Other Low Priority Features

Are there other Low priority features you would like to see that are not in the list above?

- Conquer the world, but with very low priority, first science.
- not really

## Other Feedback on Visualization Tools

Do you have any other feedback on any of these tools?

Of the tools you have used, were they useful and why (or why not)? Did you try to use any of them, but couldn't get them to work?

- I have always enjoyed working with the GMOD tools
- Gbrowse has been very useful for our visualization needs. JBrowse looks promising too and I hope to work with it more.
- Installation of Gbrowse was relatively easy, BioPerl was not that straightforward and I resorted to installing one module at a time. Some errors I couldn't figure out but they aren't relevant to GBrowse. It would be nice if the BioPerl core install could be made simpler like BioRuby (I know GMOD doesn't produce BioPerl but they are intimately tied).
- used CMap in the past and found it useful. currently use Gbrowse and find it very useful
- Gbrowse is overall a useful visualization tool, although not a replacement for other tools which offer a "higher resolution" view of a genomic region (as Apollo does, although it is an editor). We've found Apollo to be very useful and is part of our core annotation work.
- I have used Gbrowse and Gbrowse\_syn, they are just great, but proper tutorial for gbrowse\_syn should be there having all the minor details which may cause problem for beginners.
- It's great, I only tried a couple (or three?) and I always find great admin documentation. I hope to be in a GMOD meeting to unlock the complete power of GMOD for my lab.
- We still have problems installing and configuring the tools (*i.e.* CMap and Gbrowse) documentation could be improved.
- I would like JBrowse to be better supported and developed... I think this will be the future way of browser, although it looks a little clunky at the moment...
- Help desk seems to be very active. I am very happy with that. Thanks for your time and consideration.

## Other Feedback

If you have any additional feedback, questions, or information you would like to provide, please tell us here.

- Please better documentation and support for additional glyphs in Apollo.
- The GBrowse mailing list has been very helpful and questions are usually answered very quickly, thanks!
- Well... I really wish there was better integration between BioMart and GBrowse.
  1. I'd like to use BioMart to filter a set of features which are displayed in a GBrowse track. What I envisage is a BioMart plugin with a configuration button which gives the full BioMart web interface Filter page. So I could have a GBrowse track showing *e.g.* Affy probesets with a certain  $p$  value in a particular statistical analysis. All the statistical data would be stored in a mart and the cutoff could be changed easily in the plugin configuration. Currently I store such things in a GFF database and have a fixed cutoff for display which I can't (easily) change.
  2. I'd like to be able to jump from a GBrowse panel to a BioMart query in which the genomic location has automatically been added as filters. (Actually this is not too hard to implement but maybe there's a neat way to do it)
- CMap installation instructions sucks.
- I love the fact that GMOD cares - for end users and programmers. Whenever I have had problems I have experienced very helpful advice and quick response times. Just the fact that you \*do\* this survey shows that you guys care. I would love to see more funding for generic programming in the bioinformatics sector.
- GBrowse prerequisites should be in form of compact package.
- Having only recently come on board in using GBrowse, I have found it really difficult to set up in particular, but also to use in correctly visualising data in GFF3 format. I have found that it appears to misinterpret a few key aspects of the GFF3 format, specifically I have found that it's reading the different fields of the GFF3 file is interrupted by spaces where it doesn't appear in the GFF3 documentation that this should, as well as not being able to read fields separated by more than one tab. These features should be fairly easy to write into the parser in the back-end and would make the user experience far easier in my opinion.
- Most of the GMOD-based packages are not really cute (very useful, very easy to use, but not exactly cute), why?
- not really, but CMap is brilliant!
- Thank you VERY MUCH for GBrowse!!! It is a fantastic tool that helped us making important discoveries!