# A Virtual Genome Environment (VGE) for the visualization and navigation of genomic data

Derek A. Ruths[1], Edward S. Chen[1], and Leland Ellis[2,3]

[1]HyperSoft
Rice University
Houston, Texas 77005
<druths@rice.edu, echen@cs.rice.edu>

[2]W. M. Keck Center for Informatics
Institute of Biosciences and Technology
Texas A&M University System Health Science Center
Houston, Texas 77030
<leland@xian.tamu.edu>

[3]Corresponding author

## Abstract

As the genomes of many organisms are completely sequenced, genome databases will store thousands of megabases of DNA sequence. Such data is inherently highly-dimensional, as many different data types and inter-relationships derive from its subsequent analysis. In the present study, we have developed visual metaphors to represent classes of genome data in an interactive 3D Virtual Genome Environment (VGE), which is rendered and displayed on a projection workbench.

## 1 Introduction

Genome projects are rapidly providing the complete DNA sequence for a wide range of biological organisms. These include single-celled archaea, bacteria and yeast, a variety of multi-cellular 'model' species (fly, nematode), plants (*Arabidopsis*, rice) and animals (mouse, human). These primary data sets range in size from one to thousands of megabases. However, numerous additional data types and relationships are generated by the annotation and analysis of this sequence information. This creates significant challenges for all aspects of data management (storage, analysis, and visualization). Typical user interfaces lack dimensionality, as they are often based around a Web browser, which provides access to text or numerical data stored locally or remotely in flat files or databases.

3D virtual environments have been effectively used in a variety of disciplines for the visualization of large and often quite complex data sets (e.g., astrophysics, finance, geosciences, medicine, oil and gas, space exploration and training, etc.). The effective design of the user environment requires a careful choice of visual metaphors for display, navigation, and interaction with the information [16]. Such interaction provides user access to the underlying data.

In biology, there are examples of data types for which 3D representations are quite familiar.

For example, ribbon diagrams are often used to visualize the 3D structure of proteins [9], facilitating the recognition of elements of secondary structure (e.g., α-helices, β-strands). The user can interact with this structure in a 3D environment using stereoscopic shutter glasses. Such molecules have been rendered in virtual environments.

In the present study, we use data sets typically generated early in the study of DNA sequences, and develop visual metaphors for their representation in an immersive 3D virtual environment.

## 2 System and Methods

### 2.1 Sequence Comparisons

One of the first steps in the analysis of newly derived DNA sequence is its comparison versus all other available DNA and/or protein sequences, which are curated in a number of international databases (e.g., Genbank, EMBL, Swiss-Protein, etc.) [3]. Users often simultaneously submit from one to thousands of sequence queries for such comparisons against all sequences in multiple public databases. Each database is now comprised of thousands to millions of individual sequences.

Programs such as NCBI's BLAST [1] provide voluminous output from such comparisons, including lists of 'hits', sequence annotation, and sequence alignments (of high-scoring segment pairs, HSPs). In addition, statistical parameters for each alignment are provided. One such example is the statistical Expect value - given the score of an alignment, this value is the probability that the alignment is a chance occurrence.

There are many other data types that can be derived from these reports. For example, NCBI requires that all sequences submitted to its databases include the organism (species) from which the sequence came. They organize this information into their Taxonomy Database (TaxDB). In BLAST reports, one can find the taxon of origin (i.e., *Genus species*) of each 'hit'. Thus, using TaxDB, one can use each taxon to derive the full taxonomy of the organism.

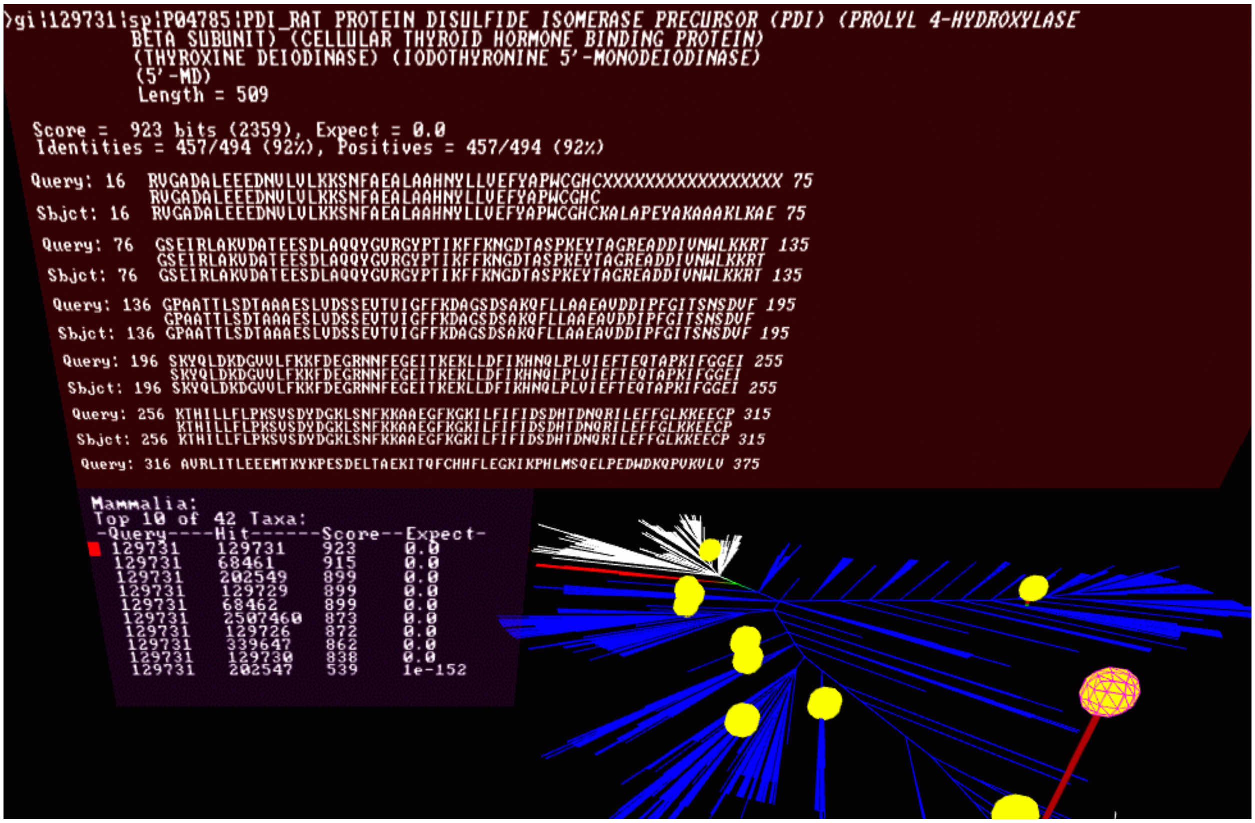
Each step in the analysis of sequences may well be comprised of extensive, voluminous data sets: sequence queries, one or more sequence databases, primary data product (BLAST reports), and secondary 'derived' data (e.g., complete taxonomic description of each 'hit').

### 2.2 Experimental Methods

The data illustrated in Figures 1-4 derive from a BLASTP (version 2.0.9) analysis of the NCBI non-redundant (nr) protein database (399,380 sequences) using rat protein disulfide isomerase as the query (gi 129731, sp P04785, PDI_RAT PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI), 509 residues) [4].

### 2.3 Visual Metaphors

Phylogenetic trees are familiar 2D visual metaphors for the display of relationships between sequences and/or organisms, in which node placement and branch length reflect evolutionary distance between individuals [20]. Such trees may be derived from the analysis of small subunit ribosomal RNA sequences from multiple sources (prokaryotic, eukaryotic, mitochondrial) [24]. Taxonomic data (see above) is itself also inherently hierarchical, and can also be displayed as a branched tree structure, even though the evolutionary distance between nodes may not be known (in such trees, all branches are drawn of equal and arbitrary length).



```
>gi|129731|sp|P04785|PDI_RAT PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) (PROLYL 4-HYDROXYLASE
           BETA SUBUNIT) (CELLULAR THYROID HORMONE BINDING PROTEIN)
           (THYROXINE DEIODINASE) (IODOTHYRONINE 5'-MONODEIODINASE)
           (5'-MD)
           Length = 509

 Score =  923 bits (2359), Expect = 0.0
 Identities = 457/494 (92%), Positives = 457/494 (92%)

Query: 16  RVGADALEEEDNVLVLKKSNFAEALAAHNYLLVEFYAPWCGHCXXXXXXXXXXXXXXXXX 75
            RVGADALEEEDNVLVLKKSNFAEALAAHNYLLVEFYAPWCGHC
Sbjct: 16  RVGADALEEEDNVLVLKKSNFAEALAAHNYLLVEFYAPWCGHCKALAPEYAKAAAKLKAE 75

Query: 76  GSEIRLAKVDATEESDLAQQYGVRGYPTIKFFKNGDTASPKEYTAGREADDIVNWLKKRT 135
            GSEIRLAKVDATEESDLAQQYGVRGYPTIKFFKNGDTASPKEYTAGREADDIVNWLKKRT
Sbjct: 76  GSEIRLAKVDATEESDLAQQYGVRGYPTIKFFKNGDTASPKEYTAGREADDIVNWLKKRT 135

Query: 136 GPAATTLSDTAAAESLVDSSEVTVIGFFKDAGSDSAKQFLLAAEAVDDIPFGITSNSDVF 195
            GPAATTLSDTAAAESLVDSSEVTVIGFFKDAGSDSAKQFLLAAEAVDDIPFGITSNSDVF
Sbjct: 136 GPAATTLSDTAAAESLVDSSEVTVIGFFKDAGSDSAKQFLLAAEAVDDIPFGITSNSDVF 195

Query: 196 SKYQLDKDGVVLFKKFDEGRNNFEGEITKEKLLDFIKHNQLPLVIEFTEQTAPKIFGGEI 255
            SKYQLDKDGVVLFKKFDEGRNNFEGEITKEKLLDFIKHNQLPLVIEFTEQTAPKIFGGEI
Sbjct: 196 SKYQLDKDGVVLFKKFDEGRNNFEGEITKEKLLDFIKHNQLPLVIEFTEQTAPKIFGGEI 255

Query: 256 KTHILLFLPKSVSDYDGKLSNFKKAAEGFKGKILFIFIDSDHTDNQRILEFFGLKKEECP 315
            KTHILLFLPKSVSDYDGKLSNFKKAAEGFKGKILFIFIDSDHTDNQRILEFFGLKKEECP
Sbjct: 256 KTHILLFLPKSVSDYDGKLSNFKKAAEGFKGKILFIFIDSDHTDNQRILEFFGLKKEECP 315

Query: 316 AVRLITLEEEMTKYKPESDELTAEKITQFCHHFLEGKIKPHLMSQELPEDWDKQPVKVLV 375
```

```
Mammalia:
Top 10 of 42 Taxa:
-Query----Hit------Score--Expect-
 129731   129731   923    0.0
 129731   68461    915    0.0
 129731   202549   899    0.0
 129731   129729   899    0.0
 129731   68462    899    0.0
 129731   2507460  873    0.0
 129731   129726   872    0.0
 129731   339647   862    0.0
 129731   129730   838    0.0
 129731   202547   539    1e-152
```

**Figure 1.** VGE's graphical space (lower right panel) and two graphical windows (upper, and lower left, panels). The data are from a BLASTP analysis of NCBI's non-redundant protein database (339,380 sequences), using rat protein disulfide isomerase [4] as a query. The colors of the taxonomic tree (lower right panel) correspond to the three Superkingdoms: Archaea (red), Proteobacteria (white), and Eukarya (blue).

### 2.4 Virtual Genome Environment (VGE)

For the design of the Virtual Genome Environment (VGE), we chose to begin with primary data derived from DNA sequence comparisons (i.e., BLAST analysis reports). A planar, 2D branched taxonomic tree, when rendered as a graphical object in a 3D environment, provides a convenient and intuitive visual metaphor for the user (see also [2]). This also serves as a framework onto which to map other 3D objects. Such objects correspond to biology - in the current implementation, they refer to information contained within BLAST reports.

Thus, prior to graphical rendering, the data flow for input into VGE consists of the following steps: (i) begin with a BLAST report of sequence comparisons; (ii) parse the full taxonomy of each sequence 'hit' from TaxDB; (iii) sum all of the 'hits' at each taxonomic node, to create a data input file (.hit) that is read by VGE at the beginning of the user session.

As a second visual metaphor, we represent the summed 'hits' at each taxonomic node as a 3D graphical push-pin, in which the height of the pin is proportional to the number of 'hits' at the node. In addition, the color of the stalk of the pin is used to indicate the number of scores that are above or below (red and green, respectively) a user-defined threshold (e.g., an Expect value of 0.01). The head of the pin is represented as a yellow sphere of uniform diameter, and provides a useful target for the user to interact with.

VGE is rendered using an SGI Onyx2 [19], and displayed on a projection workbench (Fakespace's ImmersaDesk) [5]. The user wears head-tracked stereoscopic shutter glasses, and interacts with the virtual environment using a hand-held wand (with three buttons and a joy stick). Examples of VGE displays are illustrated in Figures 1-4.

## 3 Implementation

### 3.1 Data Parsers

The primary data for input to VGE derives from sequence comparisons (e.g., BLAST reports) using from one to thousands of queries versus one of the NCBI protein or DNA databases. There are two kinds of Perl [23] data parsers: (i) one to convert taxonomic data from TaxDB into a data format used by our Arbor3D program [17] (see below) to render the tree in the graphical space of the VGE display; (ii) a second to group BLAST 'hits' into taxonomic categories corresponding to nodes of the tree. As detailed below, visual metaphors in the display graphically represent underlying data (i.e., the data formatted by the parsers).

VGE and Arbor3D use the Newick data format [7]. This format uses nested parentheses to encode the representation of phylogenetic trees, and is used by programs such as PAUP [22] and PHYLIP [6] to generate 2D displays of such trees.

However, we were unable to find any existing programs to render such trees as 3D graphical objects. In addition, while Newick data files are available from the Ribosomal Database (RDB) Project [12] for species whose small ribosomal subunits have been sequenced (prokaryotes, eukaryotes, mitochondria), there are far more species represented in the DNA sequence databases (and in TaxDB, and thus in the BLAST reports) compared to RDB. We could find no Newick files for the taxonomic data represented in TaxDB (see below).

Therefore, we developed a parser to generate an enhanced Newick format data file from the NCBI taxonomy flatfiles. The parser creates a tree where each node corresponds to a taxon of given name and level. This facilitates graphical querying and parsing of the tree in VGE (see below). This file is used by VGE to render the tree (see below).

In addition, we developed a second parser for BLAST reports, which sums the number of 'hits' at each taxonomic level.

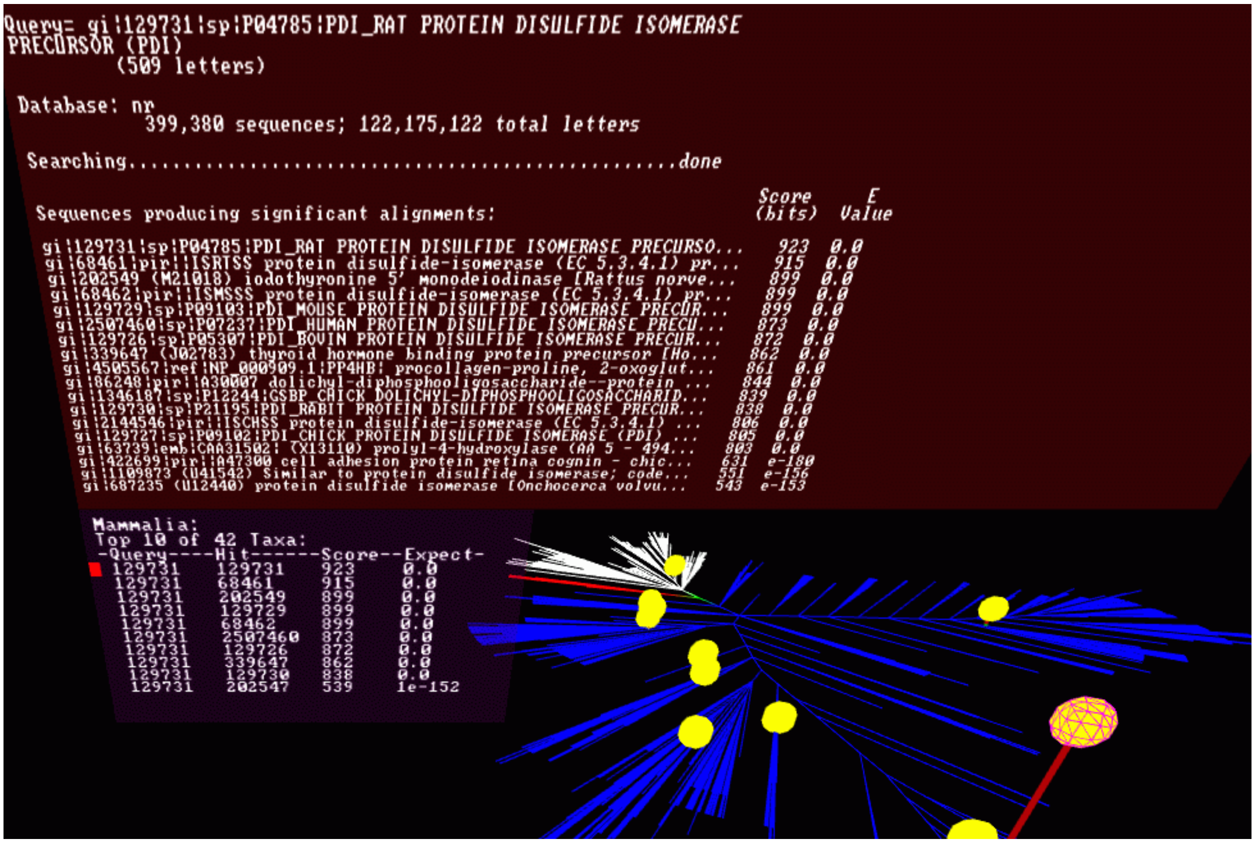
The resulting '.hit' data file is read by VGE to display the 'hits' as 3D graphical push-pins at taxonomic nodes on the tree.

### 3.2 Arbor3D - Tree Representation

Phylogenetic analysis programs often output results as 2D phylogenetic trees. For VGE, we chose a modification of Knuth's Threaded tree structure[10], which lends itself well to representing the tree in a hierarchical scene graph using SGI Performer [18], and permits real-time tree traversal and manipulation. We call this the Parent-Child-Sibling tree structure.

We placed two initial constraints upon the Arbor3D drawing algorithm: (i) it should be algorithmically efficient to traverse in order to find subnodes; (ii) the representation should lend itself to 3D graphical rendering.

The Parent-Child-Sibling data structure developed for Arbor3D makes tree traversal logi-



```
Query= gi¦129731¦sp¦P04785¦PDI_RAT PROTEIN DISULFIDE ISOMERASE
PRECURSOR (PDI)
         (509 letters)

 Database: nr
           399,380 sequences; 122,175,122 total letters

 Searching..................................................done

                                                              Score    E
 Sequences producing significant alignments:                  (bits)  Value

 gi¦129731¦sp¦P04785¦PDI_RAT PROTEIN DISULFIDE ISOMERASE PRECURSO...   923  0.0
 gi¦68461¦pir¦¦ISRTSS protein disulfide-isomerase (EC 5.3.4.1) pr...   915  0.0
 gi¦202549 (M21018) iodothyronine 5' monodeiodinase [Rattus norve...   899  0.0
 gi¦68462¦pir¦¦ISMSSS protein disulfide-isomerase (EC 5.3.4.1) pr...   899  0.0
 gi¦129729¦sp¦P09103¦PDI_MOUSE PROTEIN DISULFIDE ISOMERASE PRECUR...   899  0.0
 gi¦2507460¦sp¦P07237¦PDI_HUMAN PROTEIN DISULFIDE ISOMERASE PRECU...   873  0.0
 gi¦129726¦sp¦P05307¦PDI_BOVIN PROTEIN DISULFIDE ISOMERASE PRECUR...   872  0.0
 gi¦339647 (J02783) thyroid hormone binding protein precursor [Ho...   862  0.0
 gi¦4505567¦ref¦NP_000909.1¦PP4HB¦ procollagen-proline, 2-oxoglut...   861  0.0
 gi¦86248¦pir¦¦A30007 dolichyl-diphosphooligosaccharide--protein ...   844  0.0
 gi¦1346187¦sp¦P12244¦GSBP_CHICK DOLICHYL-DIPHOSPHOOLIGOSACCHARID...   839  0.0
 gi¦129730¦sp¦P21195¦PDI_RABIT PROTEIN DISULFIDE ISOMERASE PRECUR...   838  0.0
 gi¦2144546¦pir¦¦ISCHSS protein disulfide-isomerase (EC 5.3.4.1) ...   806  0.0
 gi¦129727¦sp¦P09102¦PDI_CHICK PROTEIN DISULFIDE ISOMERASE (PDI) ...   805  0.0
 gi¦63739¦emb¦CAA31502¦ (X13110) prolyl-4-hydroxylase (AA 5 - 494...   803  0.0
 gi¦422699¦pir¦¦A47300 cell adhesion protein retina cognin - chic...   631  e-180
 gi¦1109873 (U41542) Similar to protein disulfide isomerase; code...   551  e-156
 gi¦687235 (U12440) protein disulfide isomerase [Onchocerca volvu...   543  e-153
```

```
Mammalia:
Top 10 of 42 Taxa:
-Query----Hit------Score--Expect-
 129731   129731    923    0.0
 129731   68461     915    0.0
 129731   202549    899    0.0
 129731   129729    899    0.0
 129731   68462     899    0.0
 129731   2507460   873    0.0
 129731   129726    872    0.0
 129731   339647    862    0.0
 129731   129730    838    0.0
 129731   202547    539    1e-152
```

**Figure 2.** VGE display of high ranking sequence similarity scores for the query (i.e., 'top hits'). the user can toggle between this view in the top panel and that illustrated in Figure 1.

cally recursive, and allows easy transition from raw data to the visual realm.

### 3.3 Program Architecture

For the implementation of VGE, we developed an object-oriented mapping between the SGI Performer [18] scene graph (graphics) and the underlying sequence similarity data (parsed primary data).

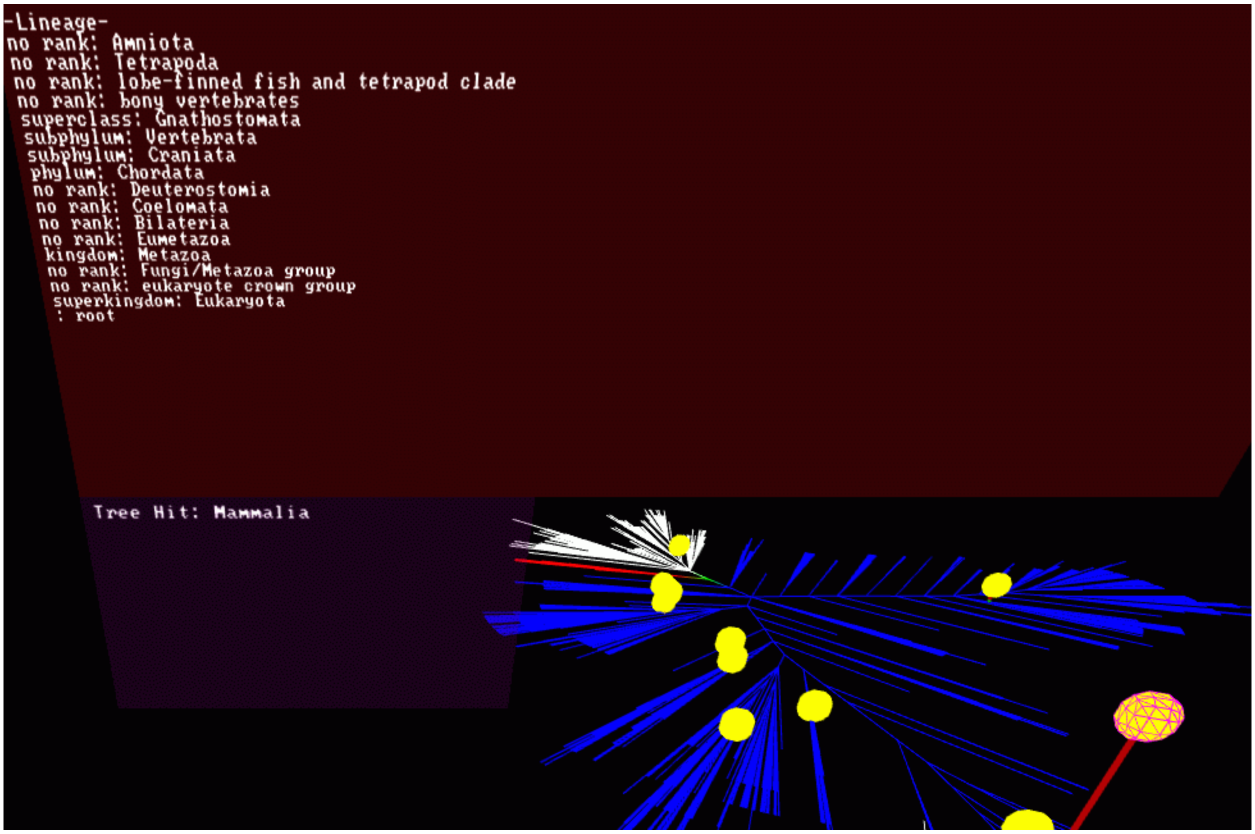
The tree provides the visual framework for the representation and display of the underlying data, as well as the basis for user interaction and navigation. The colors used for the three main divisions of the tree correspond to the three Superkingdoms: Archaea (red), Proteobacteria (white) and Eukarya (blue).

Tree nodes can be selected and their lineage displayed. In addition, push-pins denote the taxonomic location of 'hits', and two additional graphical windows (upper and lower) display the underlying primary data based on user interactions (Figure 1).

## 4 User Interaction

### 4.1 VGE's Graphical Windows

Users interact with VGE by using the handheld wand to select either a graphical object (i.e., a push-pin), or a branch of the tree, for interrogation. Selection of a push-pin highlights the object (in red cross-hatch), and presents a list of the underlying 'hits' in the lower graphical window (Figure 1). The list includes several pieces of primary data: the NCBI Geninfo Identifier (Gi number) for the Query and



```
-Lineage-
no rank: Amniota
 no rank: Tetrapoda
  no rank: lobe-finned fish and tetrapod clade
   no rank: bony vertebrates
    superclass: Gnathostomata
     subphylum: Vertebrata
      subphylum: Craniata
       phylum: Chordata
        no rank: Deuterostomia
         no rank: Coelomata
          no rank: Bilateria
           no rank: Eumetazoa
            kingdom: Metazoa
             no rank: Fungi/Metazoa group
              no rank: eukaryote crown group
               superkingdom: Eukaryota
                : root
```

```
Tree Hit: Mammalia
```

**Figure 3.** VGE display of the complete taxonomy of the selected node of the tree (in red cross-hatch).

Hit, and the numerical Score and Expect values for the alignment. Users can scroll through this list, which displays the selected alignment in the upper graphical window. Users of BLAST will recognize this display as the corresponding section of the BLAST report.

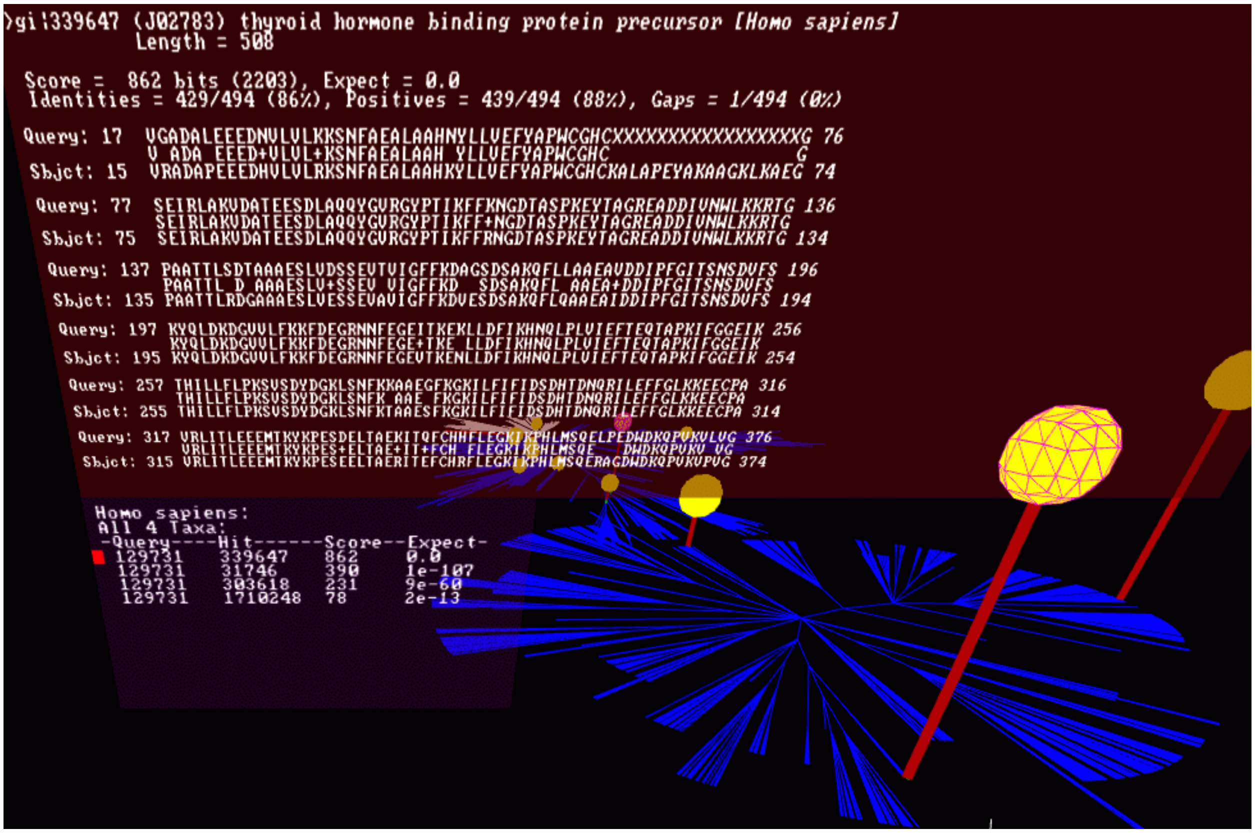
In addition, users can toggle the display in the upper graphical window between the selected alignment (Figure 1), a list of the top similarity 'hits' (Figure 2), or the taxonomy of the 'hit' (Figure 3).

### 4.2 Hierarchical Tree Navigation

Given that VGE often displays push-pins at multiple nodes of the tree, the user can create a spanned (zoomed) view for closer inspection by selecting a node and toggling the right wand button (Figure 4). Tree spanning is a fully recursive process, trees can be spanned many levels deep, and the user functionality at any sub-branch is equivalent to the base tree.

### 4.3 Tree Branch Interrogation

In addition to interacting with the push-pins, users can interrogate the tree itself (i.e., VGE functions as a taxonomic tree browser) by clicking (left button) on a branch of the tree (data not shown). The lower graphical window displays the tree node selected, while the upper graphical window displays the full taxonomic hierarchy for the node. The Newick file format is fully annotated with respect to taxonomic name and level at each node.



```
>gi|339647 (J02783) thyroid hormone binding protein precursor [Homo sapiens]
          Length = 508

 Score =  862 bits (2203), Expect = 0.0
 Identities = 429/494 (86%), Positives = 439/494 (88%), Gaps = 1/494 (0%)

Query: 17  VGADALEEEDNVLVLKKSNFAEALAAHNYLLVEFYAPWCGHCXXXXXXXXXXXXXXXXXG 76
            V ADA EEED+VLVL+KSNFAEALAAH YLLVEFYAPWCGHC                 G
Sbjct: 15  VRADAPEEEDHVLVLRKSNFAEALAAHKYLLVEFYAPWCGHCKALAPEYAKAAGKLKAEG 74

Query: 77  SEIRLAKVDATEESDLAQQYGVRGYPTIKFFKNGDTASPKEYTAGREADDIVNWLKKRTG 136
            SEIRLAKVDATEESDLAQQYGVRGYPTIKFF+NGDTASPKEYTAGREADDIVNWLKKRTG
Sbjct: 75  SEIRLAKVDATEESDLAQQYGVRGYPTIKFFRNGDTASPKEYTAGREADDIVNWLKKRTG 134

Query: 137 PAATTLSDTAAAESLVDSSEVTVIGFFKDAGSDSAKQFLLAAEAVDDIPFGITSNSDVFS 196
            PAATTL D AAAESLV+SSEV VIGFFKD  SDSAKQFL AAEA+DDIPFGITSNSDVFS
Sbjct: 135 PAATTLRDGAAAESLVESSEVAVIGFFKDVESDSAKQFLQAAEAIDDIPFGITSNSDVFS 194

Query: 197 KYQLDKDGVVLFKKFDEGRNNFEGEITKEKLLDFIKHNQLPLVIEFTEQTAPKIFGGEIK 256
            KYQLDKDGVVLFKKFDEGRNNFEGE+TKE LLDFIKHNQLPLVIEFTEQTAPKIFGGEIK
Sbjct: 195 KYQLDKDGVVLFKKFDEGRNNFEGEVTKENLLDFIKHNQLPLVIEFTEQTAPKIFGGEIK 254

Query: 257 THILLFLPKSVSDYDGKLSNFKKAAEGFKGKILFIFIDSDHTDNQRILEFFGLKKEECPA 316
            THILLFLPKSVSDYDGKLSNFK AAE FKGKILFIFIDSDHTDNQRILEFFGLKKEECPA
Sbjct: 255 THILLFLPKSVSDYDGKLSNFKTAAESFKGKILFIFIDSDHTDNQRILEFFGLKKEECPA 314

Query: 317 VRLITLEEEMTKYKPESDELTAEKITQFCHHFLEGKIKPHLMSQELPEDWDKQPVKVLVG 376
            VRLITLEEEMTKYKPES+ELTAE+IT+FCH FLEGKIKPHLMSQE   DWDKQPVKV VG
Sbjct: 315 VRLITLEEEMTKYKPESEELTAERITEFCHRFLEGKIKPHLMSQERAGDWDKQPVKVPVG 374
```

```
Homo sapiens:
All 4 Taxa:
-Query----Hit------Score--Expect-
■ 129731   339647    862    0.0
 129731   31746     390    1e-107
 129731   303618    231    9e-60
 129731   1710248   78     2e-13
```

**Figure 4.** VGE display of a spanned (zoomed) tree.

## 5 Future Work

### 5.1 Additional Visual Metaphors

The two graphical windows of VGE are used for the display of underlying data, all of which is currently text. The addition of other data types and visual metaphors to the VGE display is clearly of interest. Other data types which are relevant to the current approach might include multiple sequence alignments, sequence motifs, protein classifications (functional families), 3D protein structures, and genome maps (e.g., genetic, physical). In addition, the use of implicit surfaces to display the results of queries as 3D graphical 'glyphs' offers a potentially useful layer of abstraction between the user and primary textual data [15].

### 5.2 Display of Multiple Data Sets

We have focused on the use of a single search program (BLAST). We would like to be able to compare results using multiple similarity search algorithms (e.g., BLAST versus FASTA [14], which uses a heuristic hashing algorithm), as well as multiple data sets simultaneously, so that a user can directly compare the results of multiple queries over time (e.g., as databases are updated). We plan to explore the use of transparency methods (e.g., OpenGL's a channel) [25] to compare two or more overlapping trees.

### 5.3 Modules

The requirements outlined in Sections 5.1 and 5.2 underscore the need for flexibility to extend VGE functionality. This can be accomplished programmatically via use of the modules or plug-ins. We are currently investigating extending VGE in this manner.

### 5.4 Distributed VGE

Both real-time and asynchronous interactions among multiple participants at different institutions are common aspects of collaboration over genomics data. We plan to experiment with several systems to share virtual collaborative environments in real-time, including CAVERNsoft [11] and VEGA [13].

### 5.5 Non-immersive VGE applications

Facilities for immersive environments are still relatively uncommon, and are not currently used in the genomics community. We are experimenting with several different strategies to implement non-immersive applications based on VGE, including VEGA (Multigen/Paradigm) [13], SGI Performer [18], and Java 3D [21,8]. We expect that such applications will ultimately provide a means for more formal evaluation of the utility of 2D versus 3D, and immersive versus non-immersive environments in this application domain.

## 6 Summary

We have developed the Virtual Genome Environment (VGE) as an initial prototype of an immersive 3D application for visualization and navigation of genomic data. We will continue to enhance the functionality of VGE, including the addition of new data types and visual metaphors to the program, as well as explore the utility of distributed VGE to enhance collaboration over large, multi-dimensional genomic datasets.

With the introduction of VGE, we hope to engage the input of the VR community to develop new visual metaphors for the application domain of genomics. VGE's object-oriented API provides a modular, extensible architecture by which new application functionality and display design can be rapidly prototyped.

## 7 Acknowledgments

We are grateful to Joseph Felsenstein (University of Washington) and Mitch Sogin (Marine Biological Laboratory) for their input on phylogenetic trees; Bowen Loftin (University of Houston) for his advice and many helpful discussions during the course of this work; Geetha Vasudevan (Institute of Biosciences and Technology) for her comments on the manuscript; the NSF Center for Research on Parallel Computation at Rice University for the use of their Onyx2 and ImmersaDesk; and the W. M. Keck Foundation for their support (to LE).

## References

[1] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25:3389-3402, 1997.

[2] M. Bailey, J. Humphries, and D. Nadeau. Phylogenetic visualization.
http://www.sdsc.edu/phylo/, 1996.

[3] A. Baxevanis and B. F. F. Ouellette, editors. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley & Sons, 4th edition, 1998.

[4] J. C. Edman, L. Ellis, R. W. Blacher, Richard A. Roth, and W. J. Rutter. Sequence of protein disulfide isomerase and implications of its relationship to thioredoxin. Nature (London), 317:267-270, 1985.

[5] Fakespace. Fakespace home page.
http://www.fakespace.com/, 1999.

[6] J. Felsenstein. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol., 17:368-376, 1981.

[7] J. Felsenstein. Phylip, a free package of programs for inferring phylogenies.
http://evolution.genetics.washington.edu/phylip.html, 1999.

[8] W. Hibbard. An example of Unidata's future in new software: The VisAD component architecture for collaborative data analysis and visualization.
http://www.ssec.wisc.edu/~billh/visad.html, 1999.

[9] S.A. Hubbard, L. Wei, L. Ellis, and W.A. Hendrickson. Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature (London), 372:746-753, 1994.

[10] D.E. Knuth. The Art of Computer Programming, volume 1. Fundamental Algorithms. Addison-Wesley, Reading, Massachusetts, 3rd edition, 1997.

[11] Electronic Visualization Laboratory. The CAVE research network.
http://www.evl.uic.edu/cavern/vrserver.html, 1999.

[12] B. L. Maidak, J. R. Cole, C. T. Parker, G. M. Garrity, N. Larsen, B. Li, T. G. Lilburn, M. J. McCaughey, G. J. Olsen, R. Overbeek, S. Pramanik, T. M. Schmidt, J. M. Tiedje, and C. R. Woese. A new version of the RDP (Ribosomal Database Project). Nucleic Acids Res., 27:171-173, 1999.

[13] Multigen-Paradigm. Vega software environment for real-time simulation, virtual reality, and visualization.
http://www.multigen.com/, 1999.

[14] W.R. Pearson. Flexible sequence similarity searching with the FASTA3 program package.
http://www.people.Virginia.EDU/~wrp/papers/mmol98f.pdf, 1999.

[15] R. M. Rohrer, J. L. Sibert, and D. S. Ebert. A shape-based visual interface for text retrieval. IEEE Computer Graphics and Applications, 19(5):40-46, Sept.-Oct., 1999.

[16] L. Rosenblum, P. Astheimer, and D. Teichmann, editors. IEEE Virtual Reality '99, Houston, Texas, March 13-17, 1999. Technical Committee on Visualization and Graphics, IEEE Computer Society, Los Alamitos, California.

[17] D. A. Ruths, E. S. Chen, and L. Ellis. Arbor3D: An interactive environment for examining phylogenetic and taxonomic trees in multiple dimensions. Manuscript in revision, 2000.

[18] SGI. IRIS Performer.
http://www.sgi.com/software/performer/, 1999.

[19] SGI. Silicon Graphics Onyx2.
http://www.sgi.com/onyx2/, 1999.

[20] M. Sogin. Sogin lab research summary.
http://evol2.mbl.edu/researchsum.html, 1999.

[21] H. Sowizral, K. Rushforth, and M. Deering. The Java 3D API Specification. Addison-Wesley, Reading, Massachusetts, 1st edition, 1997.

[22] D.L. Swofford. Phylogenetic analysis using parsimony (PAUP).
http://www.lms.si.edu/PAUP/, 1999.

[23] L. Wall, T. Christiansen, and Schwartz R. L. Programming Perl. O'Reilly & Associates, 2nd edition, 1996.

[24] C.R. Woese, O. Kandler, and M.L. Wheelis. Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. Proc. Natl. Acad. Sci. USA, 87:4576-4579, 1990.

[25] M. Woo, J. Neider, T. David, Shriner D., OpenGL Architecture Review Board, T. Davis, and D. Shreiner. OpenGL 1.2 Programming Guide. Addison-Wesley, Reading, Massachusetts, 3rd edition, 1999.