G&I Genomics & Informatics

eISSN 2234-0742
Genomics Inform 2015;13(1):7-14
http://dx.doi.org/10.5808/GI.2015.13.1.7

ORIGINAL ARTICLE

An Ontology-Based GIS for Genomic Data Management of Rumen Microbes

Saber Jelokhani-Niaraki[1], Mojtaba Tahmoorespur[1]*, Zarrin Minuchehr[2], Mohammad Reza Nassiri[1]

[1]Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad 91775-1163, Iran,
[2]National Institute of Genetic Engineering and Biotechnology, Tehran 14965-161, Iran

During recent years, there has been exponential growth in biological information. With the emergence of large datasets in biology, life scientists are encountering bottlenecks in handling the biological data. This study presents an integrated geographic information system (GIS)-ontology application for handling microbial genome data. The application uses a linear referencing technique as one of the GIS functionalities to represent genes as linear events on the genome layer, where users can define/change the attributes of genes in an event table and interactively see the gene events on a genome layer. Our application adopted ontology to portray and store genomic data in a semantic framework, which facilitates data-sharing among biology domains, applications, and experts. The application was developed in two steps. In the first step, the genome annotated data were prepared and stored in a MySQL database. The second step involved the connection of the database to both ArcGIS and Protégé as the GIS engine and ontology platform, respectively. We have designed this application specifically to manage the genome-annotated data of rumen microbial populations. Such a GIS-ontology application offers powerful capabilities for visualizing, managing, reusing, sharing, and querying genome-related data.

Keywords: gene ontology, geographic information systems, linear referencing, rumen

Introduction

The advances in genomics technologies and other molecular research, together with recent developments in information technologies, have generated a great quantity of data in the molecular biology domain [1]. For example, GenBank, as a comprehensive and well-known database of National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) online resources, holds nucleotide sequences for more than 280,000 species that are available to the public [2]. Another example of large biological datasets is the 2013 edition of the annual Nucleic Acids Research database issue, which currently lists 1,512 databases, organized into 14 categories and 41 subcategories, available at http://www.oxfordjournals.org/nar/database/a/ [3].

Among organisms, rumen microbes have not been an exception to these debates. But, much less attention has been paid to integrate and manage the genomic data related to these microorganisms. Molecular research on ruminal microbes has gained attention as more genomic sequences and related genomic data are becoming available to scientific communities. The primary focus of this study was to develop an application that allows users to interactively manage the genome-annotated data of rumen microbial consortia. The rumen environment is the main part of the digestive system and consists of a complex microbial community that makes the rumen microbial ecosystem [4]. This ecosystem is a natural habitat that houses a wide spectrum of bacteria, protozoa, anaerobic fungi, and bacteriophages [5]. By having a symbiotic mutualistic relationship with ruminants, the microbes are capable of converting feed components into a source of energy for the animals (i.e., volatile fatty acids). Ruminants are able to utilize a variety of feeds owing to an extremely diversified rumen microbial ecosystem [4]. According to previous molecular studies, it has been reported that the population of rumen bacteria accounts for at least 300 to 400 phylotypes [6-8].

Received December 6, 2014; Revised January 16, 2015; Accepted January 28, 2015

***Corresponding author:** Tel: +98-511-8796845, Fax: +98-511-8796845, E-mail: tahmoores@um.ac.ir

There is no doubt that progress in biological research could not be achieved without equivalent efforts in informatics research. Since data management and analysis are facing new challenges in parallel with the age of big data, many research activities have been conducted for handling the biological data. There are several special issues that reflect the ever increasing importance of data management approaches in biology [9-13].

Geographic information system (GIS) is a computer-based system designed for displaying, storing, manipulating, and analyzing geographical data [14]. As an analytical system, GIS can serve biological systems by establishing biological databases, with the use of powerful representation and display tools, as well as providing query and analysis tools. An example of a GIS query is finding all hospitals that are equipped with an advanced technology, located in the city center, and with more than 500 beds. A number of GIS-based approaches have been used to handle biological information. A typical example in the biology domain is discussed by Kozak *et al*. [15], in which they applied GIS technologies and concepts for evolutionary biology. Their study uses the GIS to track macroevolutionary changes among species and the evolutionary process of geographic variation within species. In previous studies of GIS applications in genomics, a major effort was made by Dolan *et al*. [16]. They presented a GIS application, called Genome Spatial Information System (GenoSIS), to manage and represent genomic information.

Ontology is about terminology (domain vocabulary), all essential concepts in a domain, their classification, taxonomy, relations (including all important hierarchies and constraints), and domain axioms [17]. It represents the particular meanings of concepts by specifying the concepts and their relationships. The use of ontology enables one to structure biological data (genome/gene data) in a machine-understandable form by defining biological concepts, relationships, and instances in a semantic structure.

This paper presents an integrated GIS-ontology application for the management of microbial genome data. The main rationale behind integrating GIS and ontology is that these two distinct areas of research can complement each other for biological applications. While the GIS is commonly recognized as a powerful and integrated tool with unique capabilities for storing, manipulating, analyzing, and visualizing geographical information, ontology provides a rich collection of procedures and algorithms to meaningfully share and exchange data. Our application is implemented by using ArcGIS and Protégé to couple GIS-based biological applications with ontology. The genome-annotated data of rumen microbial flora were adopted to evaluate the performance of the application. The system can be extended and employed for handling other types of genomic data.

Methods

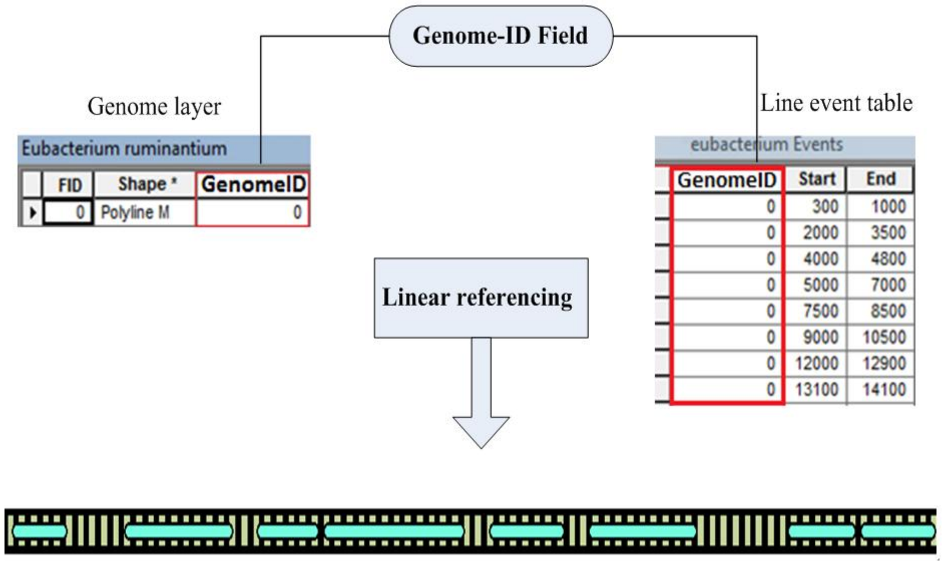
More specifically, this study adopts the functionality of linear referencing (LR) for representing genome-annotated data in a GIS environment. Our application uses ontology to semantically organize and reason the data in a semantic and machine-understandable framework. In this application, we defined a measurement unit for genomes, so that one meter was assigned per nucleotide. The users have the flexibility to interact with the genome as a layer or with different genomes as a collection of layers. The ArcGIS platform of the application supports users with interactive genomic maps and implementation of special queries.

Linear referencing

LR is one of the GIS tools representing geographic locations of linear events based on relative positions on linear features [18], where the beginning and end of the events are recorded as measurement values in event tables. Using LR, attributes are linearly referenced and linked dynamically to layers (e.g., the genome). This technique can be effectively used to represent genomes and genes as layers and linear events, respectively. By applying this technique, the genomic data initially stored in event tables can be visualized, displayed, queried, and analyzed on the genome network. The LR method provides flexibility to split a linear genome into new sets of gene segments (i.e., line events), without having to draw a new line layer for gene events. As presented in Fig. 1, the LR process includes the fields of location and attribute(s) associated with gene events. As shown, the line gene events are related to the genome layers through the values of the field “Genome-ID.” The location of line events (genes) on a given genome layer is indicated in the fields “Start” and “End.” In other words, a line event uses both “Start” and “End” measure values to describe a portion (i.e., a gene) of a genome.

Ontology

From a computer science perspective, ontology is defined as the structure representing a shared, formal, and explicit specification of knowledge in a particular domain (e.g., biology). Hendler [19] describes ontology as a set of knowledge terms, containing the terminology (domain vocabulary), all essential concepts in the domain, the semantic interconnections, and simple rules of logic and inference for some particular topics. A typical example of an ontology approach in biology is using standard concepts to classify microorganisms. There are a large number of identified microbes in nature, so that this amount has long stimulated interest to establish an anthological classification for them. Gene Ontology (GO) is another example of an



Eubacterium ruminantium

FID	Shape *	GenomeID
0	Polyline M	0

eubacterium Events

GenomeID	Start	End
0	300	1000
0	2000	3500
0	4000	4800
0	5000	7000
0	7500	8500
0	9000	10500
0	12000	12900
0	13100	14100

Fig. 1. A schematic diagram outlining the linear referencing procedure for mapping gene events on a genome linear feature. The route identifier filed "GenomeID" is presented with a red box.

ontology application in biology. GO is intended to formulize biological knowledge in orthogonal hierarchies [20]. It includes controlled vocabularies that are described for gene products with respect to their biological processes, cellular components, and molecular functions [21]. Ontology specification languages are knowledge representation languages used to encode ontologies in a formal (machine-readable and -processable) and standard format. The most common and popular ontology language is OWL from the World Wide Web Consortium (W3C). OWL concepts/classes (e.g., genome, gene, etc.) are created using formal specifications that an individual object/instance (e.g., genome and gene instances) needs to satisfy to belong to a particular class. The individuals that share similar characteristics would be members of the same concept.

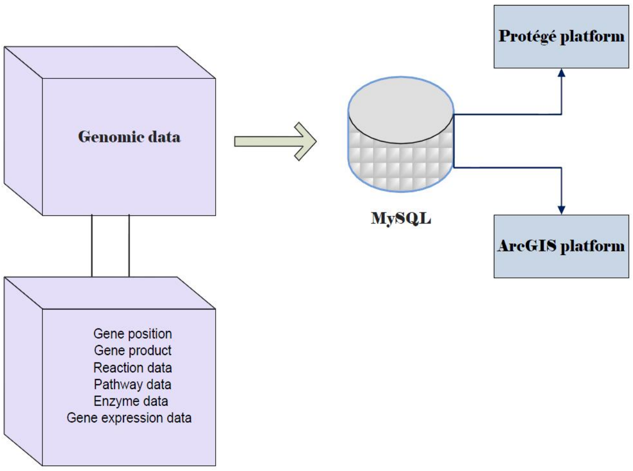


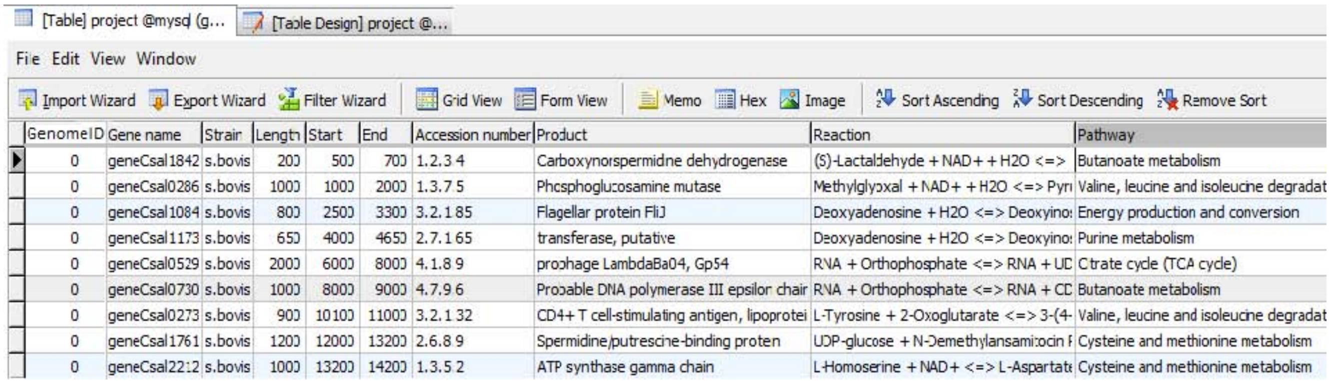
Fig. 2. The ontology-based GIS architecture for managing genomic data.

Implementation

As depicted in Fig. 2, the ontology-based GIS application was implemented using a combination of the two platforms of ArcGIS [22] and Protégé [23] and MySQL as a database management system. ArcGIS is a commercial software package containing a set of integrated applications, such as ArcMap and ArcCatalog. ArcMap is the main mapping application allowing users to generate maps, query attributes, analyze spatial relationships, and lay out final projects. The ArcCatalog application allows users to manage geographic data [24].

Several programs have been developed to edit and view ontologies, such as Ontolingua and Protégé. Both Ontolingua and Protégé are capable of generating a separate frame for every knowledge item that covers the information, such as definitions, links, and relationships [20]. In the proposed application, we applied the Protégé software as an ontology-developing environment. Protégé is a free and open-source platform that provides a set of tools for constructing domain ontologies and knowledge-based applications. We constructed the rumen microbial genome ontology to organize genome-related data, including classes (e.g., microbial genomes found in rumen and their associated genes) and their relationships/properties (e.g., gene name, product, catalyzed reaction, and accession number of a gene). In order to implement the system, we applied the genome-annotated data of several ruminal microbes deposited in NCBI.

In the first phase of implementation, the genome-annotated data were stored in a MySQL database using Navicat software (Fig. 3). Navicat, as a powerful solution for MySQL administration, offers a graphical interface for managing databases (http://navicat.com/). The second phase involved the connection of the database to both ArcGIS and Protégé as



GenomeID	Gene name	Strain	Length	Start	End	Accession number	Product	Reaction	Pathway
0	geneCsal1842	s.bovis	200	500	700	1.2.3.4	Carboxynorspermidne dehydrogenase	(S)-Lactaldehyde + NAD+ + H2O <=>	Butanoate metabolism
0	geneCsal0286	s.bovis	1000	1000	2000	1.3.7.5	Phosphoglucosamine mutase	Methylglyoxal + NAD+ + H2O <=> Pyr	Valine, leucine and isoleucine degradat
0	geneCsal1084	s.bovis	800	2500	3300	3.2.1.85	Flagellar protein FliJ	Deoxyadenosine + H2O <=> Deoxyino	Energy production and conversion
0	geneCsal1173	s.bovis	650	4000	4650	2.7.1.65	transferase, putative	Deoxyadenosine + H2O <=> Deoxyino	Purine metabolism
0	geneCsal0529	s.bovis	2000	6000	8000	4.1.8.9	prophage LambdaBa04, Gp54	RNA + Orthophosphate <=> RNA + UD	Citrate cycle (TCA cycle)
0	geneCsal0730	s.bovis	1000	8000	9000	4.7.9.6	Probable DNA polymerase III epsilon chair	RNA + Orthophosphate <=> RNA + CD	Butanoate metabolism
0	geneCsal0273	s.bovis	900	10100	11000	3.2.1.32	CD4+ T cell-stimulating antigen, lipoprotei	L-Tyrosine + 2-Oxoglutarate <=> 3-(4-	Valine, leucine and isoleucine degradat
0	geneCsal1761	s.bovis	1200	12000	13200	2.6.8.9	Spermidine/putrescine-binding proten	UDP-glucose + N-Demethylansamitocin F	Cysteine and methionine metabolism
0	geneCsal2212	s.bovis	1000	13200	14200	1.3.5.2	ATP synthase gamma chain	L-Homoserine + NAD+ <=> L-Aspartate	Cysteine and methionine metabolism

Fig. 3. The representation of genome-annotated data in a MySQL database using Navicat.

the GIS engine and ontology platform, respectively. The database was connected to the ArcCatalog application of ArcGIS. The genome layers were created as shapefiles via ArcCatalog and displayed in an ArcMap environment (Fig. 4). As shown in Fig. 5, the DataMaster plugin of Protégé was integrated with a MySQL database through a JDBC driver. DataMaster is a plugin for Protégé editor that allows the conversion of a relational database into an RDF model (resource descriptive framework) that can be processed and understood by machines [25]. After integrating DataMaster with the MySQL database, each of the tables from the database was imported into Protégé as an OWL class, along with its instances and properties.

Results

One of the advantages of a GIS application in the genomic area is that users can easily query and analyze genome-related data. In addition, the users are able to select locations on a genome map and display their records in the attribute table. GIS has a special functionality to support users with simple and complex queries for the data represented in each layer. As shown in Fig. 6, complicated queries were made based on the attributes of genomic data and a given gene location. Fig. 6 presents two query examples: one is for a special gene product (phosphoglucosamine mutase) in the annotated genome of *Streptococcus bovis*, and another is a query based on a specific accession number belonging to the gene (genes). The results of both query examples are presented on the genome, such that they are highlighted with a red rectangle on the genome layer. ArcGIS allows one to use "Identify tool." Applying this tool allows users to be able to have information about a gene displayed in ArcMap. It has been used as the fastest way to find out the characteristics of a location on a map. On the gene event layer, users are able to identify genes by clicking on a particular location (Fig. 7). This will present the attribute data of the selected location.

The genome ontology of ruminal microbes was developed using OWL/RDF language in the Protégé environment. Fig. 8 shows a screenshot of Protégé's class tab, indicated by a red rectangle. Via this tab, classes, along with their subclasses for the ontology, are displayed in the blue rectangle, and all instances and their attributes are shown in yellow and green rectangles, respectively. Additional tabs for object properties and data type properties are available in the editor, which has the same view as the class tab.

In conclusion, the proposed application provides users with the ability to manage genomic data related to the microbial community of the rumen. This application will be particularly helpful to researchers and experts who are working in fields, such as biology, animal sciences, and genomics. Two central aspects of designing and using such an interactive application are data querying and data sharing. The application allows users (1) to access, visualize, store, query, and analyze the genomic data and (2) to share such data among biology users and applications by providing a semantic structure of data. The data stored in the ontology is machine-processable and can be shared and reused by biology applications.

Discussion

With the emergence of large datasets in biology, life scientists are encountering bottlenecks in handling biological data. The increase of genomic data caused by next-generation sequencing technologies is leading to a continuous decrease in the effectiveness of conventional genomic browsers [26]. Most genomic browsers are principally used for visualizing the genomic features and a user's genomic data [26]. In order to visualize genomic features, different genomic browsers are available on the web, such as Ensembl [27] and University of California, Santa Cruz (UCSC) [28]. Some genomic browsers represent gene annotations as horizontal data tracks. For example, the UCSC genomic browser

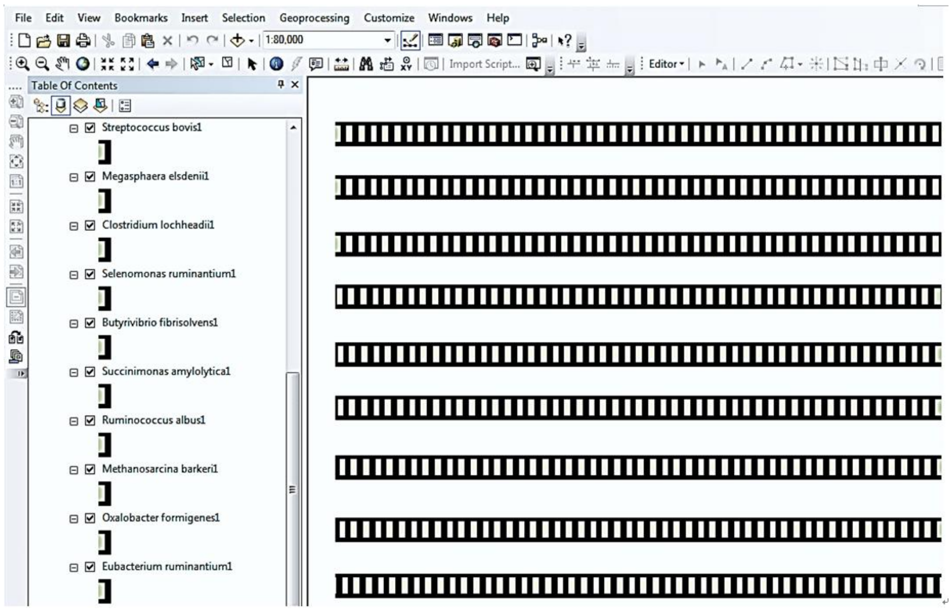
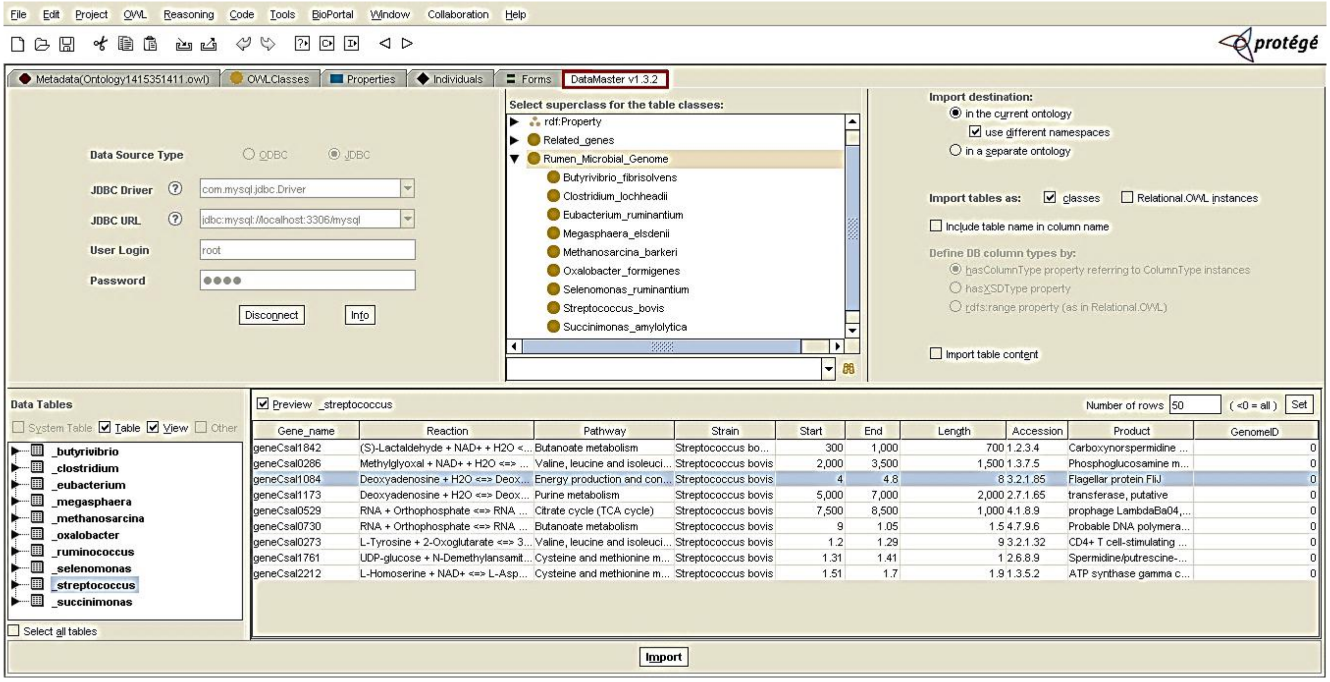


Fig. 4. Microbial genomes displayed in ArcMap as different map layers.


File Edit Project OWL Reasoning Code Tools BioPortal Window Collaboration Help
protégé
Metadata(Ontology1415351411.owl) OWLClasses Properties Individuals Forms DataMaster v1.3.2
Disconnect
Info
Import

Data Source Type ○ ODBC ● JDBC

JDBC Driver com.mysql.jdbc.Driver

JDBC URL jdbc:mysql://localhost:3306/mysql

User Login root

Password ●●●●

Select superclass for the table classes:
- rdf:Property
- Related_genes
- Rumen_Microbial_Genome
  - Butyrivibrio_fibrisolvens
  - Clostridium_lochheadii
  - Eubacterium_ruminantium
  - Megasphaera_elsdenii
  - Methanosarcina_barkeri
  - Oxalobacter_formigenes
  - Selenomonas_ruminantium
  - Streptococcus_bovis
  - Succinimonas_amylolytica

Import destination:
● in the current ontology
☑ use different namespaces
○ in a separate ontology

Import tables as: ☑ classes □ Relational.OWL instances

□ Include table name in column name

Define DB column types by:
● hasColumnType property referring to ColumnType instances
○ hasXSDType property
○ rdfs:range property (as in Relational.OWL)

□ Import table content

Data Tables
□ System Table ☑ Table ☑ View □ Other
- _butyrivibrio
- _clostridium
- _eubacterium
- _megasphaera
- _methanosarcina
- _oxalobacter
- _ruminococcus
- _selenomonas
- _streptococcus
- _succinimonas

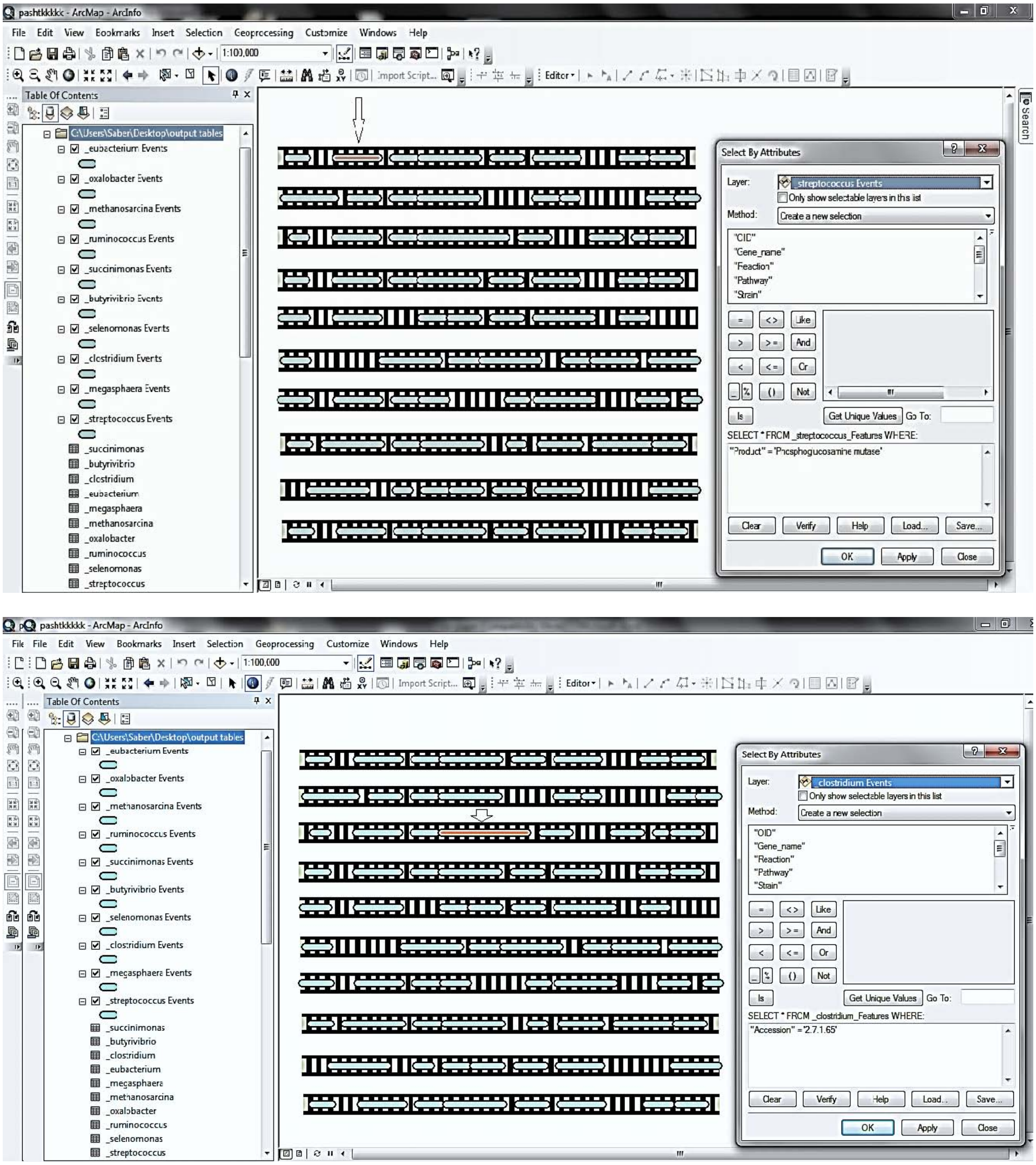
□ Select all tables

☑ Preview _streptococcus   Number of rows 50 (<0 = all) Set

Gene_name	Reaction	Pathway	Strain	Start	End	Length	Accession	Product	GenomeID
geneCsal1842	(S)-Lactaldehyde + NAD+ + H2O <...	Butanoate metabolism	Streptococcus bo...	300	1,000	700	1.2.3.4	Carboxynorspermidine ...	0
geneCsal0286	Methylglyoxal + NAD+ + H2O <=> ...	Valine, leucine and isoleuci...	Streptococcus bovis	2,000	3,500	1,500	1.3.7.5	Phosphoglucosamine m...	0
geneCsal1084	Deoxyadenosine + H2O <=> Deox...	Energy production and con...	Streptococcus bovis	4	4.8	8	3.2.1.85	Flagellar protein FliJ	0
geneCsal1173	Deoxyadenosine + H2O <=> Deox...	Purine metabolism	Streptococcus bovis	5,000	7,000	2,000	2.7.1.65	transferase, putative	0
geneCsal0529	RNA + Orthophosphate <=> RNA ...	Citrate cycle (TCA cycle)	Streptococcus bovis	7,500	8,500	1,000	4.1.8.9	prophage LambdaBa04,...	0
geneCsal0730	RNA + Orthophosphate <=> RNA ...	Butanoate metabolism	Streptococcus bovis	9	1.05	1.5	4.7.9.6	Probable DNA polymera...	0
geneCsal0273	L-Tyrosine + 2-Oxoglutarate <=> 3...	Valine, leucine and isoleuci...	Streptococcus bovis	1.2	1.29	9	3.2.1.32	CD4+ T cell-stimulating ...	0
geneCsal1761	UDP-glucose + N-Demethylansamit...	Cysteine and methionine m...	Streptococcus bovis	1.31	1.41	1	2.6.8.9	Spermidine/putrescine-...	0
geneCsal2212	L-Homoserine + NAD+ <=> L-Asp...	Cysteine and methionine m...	Streptococcus bovis	1.51	1.7	1.9	1.3.5.2	ATP synthase gamma c...	0

Fig. 5. Diagram of how a MySQL database connects to the Protégé platform. The DataMaster plugin is shown by a red small rectangle.

represents annotations as a series of horizontal tracks over the genomic sequence. In this browser, each track presents a specific type of annotation [28]. In comparison to the paradigm of horizontal data tracks applied by genomic browsers, the GIS application uses the concepts of a LR technique to present genomes in a dynamic manner, where users can define/change attributes of genes in a table and interactively see the gene events on genome layers. A previous study of a GIS application in genomic data management employed the concepts and tools that were defined for displaying the geographic data to develop GenoSIS. GenoSIS was proposed to spatially represent genomes and interact with genome-related attribute data [16]. However, this system is based on a fixed linear data model, which presents the genome as a linear feature with associated attributes. The system is fixed in the sense that defining new or changing gene attributes requires users to graphically draw new lines and change the GIS file (i.e.,

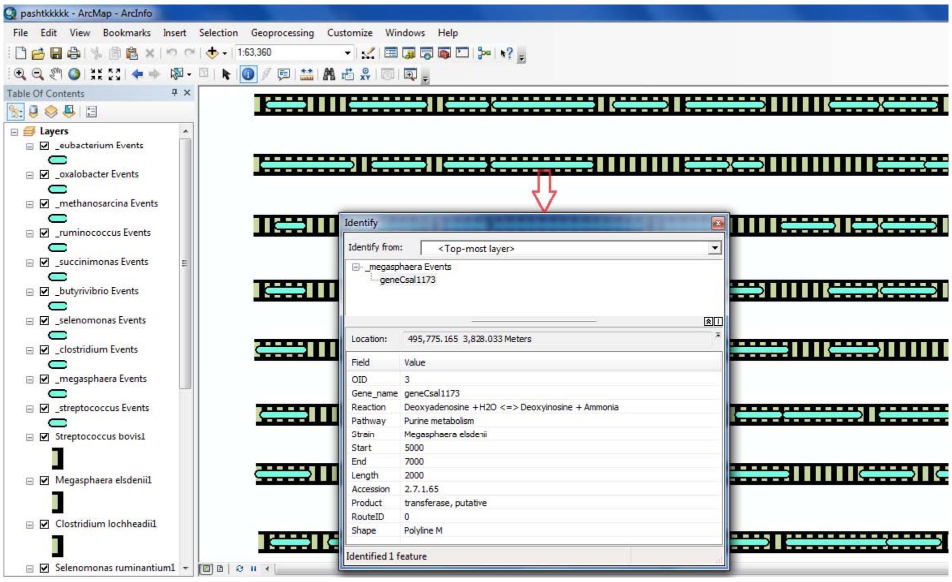


Layer: streptococcus Events
☐ Only show selectable layers in this list
Method: Create a new selection
SELECT * FROM _streptococcus_Features WHERE:
"Product" = 'Phosphoglucosamine mutase'

Layer: clostridium Events
☐ Only show selectable layers in this list
Method: Create a new selection
SELECT * FROM _clostridium_Features WHERE:
"Accession" = '2.7.1.65'

Fig. 6. Querying gene attributes using ArcMap. The query window is presented in the figure.

genome shapefile and geodatabase file). Moreover, this system lacks the ability to semantically represent genome-related information, where information can meaningfully be interpreted by software agents and can be shared between different biological applications. One of the most interesting capabilities of GIS compared to existing genomic browsers is its ability to make complex attribute and spatial queries. A GIS query is a user-defined request that examines genome/gene layer or tabular attributes based on user-selected criteria and displays only those features or records that satisfy the



Identify from: <Top-most layer>

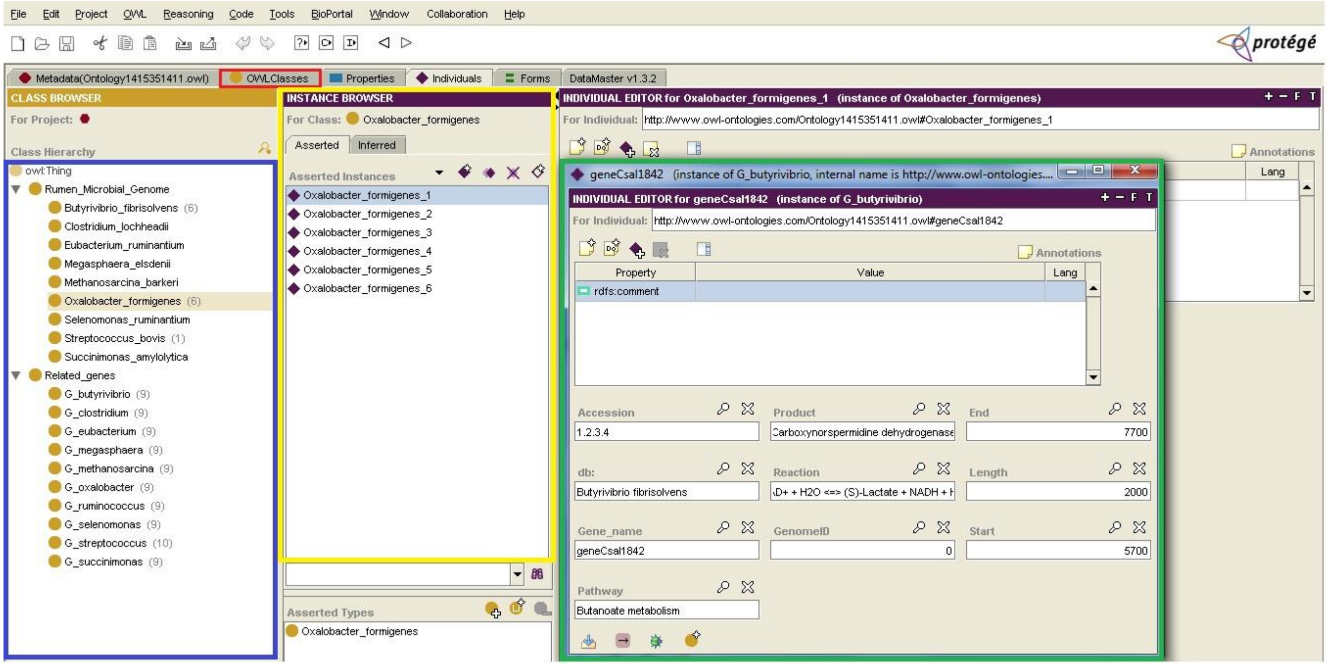
_megasphaera Events
geneCsal1173

Location: 495,775.165 3,828.033 Meters

Field	Value
OID	3
Gene_name	geneCsal1173
Reaction	Deoxyadenosine +H2O <=> Deoxyinosine + Ammonia
Pathway	Purine metabolism
Strain	Megasphaera elsdenii
Start	5000
End	7000
Length	2000
Accession	2.7.1.65
Product	transferase, putative
RouteID	0
Shape	Polyline M

Identified 1 feature

Fig. 7. Identifying features with the identify tool. The gene attributes are demonstrated in the identify window.



INDIVIDUAL EDITOR for Oxalobacter_formigenes_1 (instance of Oxalobacter_formigenes)

For Individual: http://www.owl-ontologies.com/Ontology1415351411.owl#Oxalobacter_formigenes_1

geneCsal1842 (instance of G_butyrivibrio, internal name is http://www.owl-ontologies....

INDIVIDUAL EDITOR for geneCsal1842 (instance of G_butyrivibrio)

For Individual: http://www.owl-ontologies.com/Ontology1415351411.owl#geneCsal1842

Property	Value	Lang
rdfs:comment		

Accession: 1.2.3.4

Product: Carboxynorspermidine dehydrogenase

End: 7700

db: Butyrivibrio fibrisolvens

Reaction: D+ + H2O <=> (S)-Lactate + NADH + H

Length: 2000

Gene_name: geneCsal1842

GenomeID: 0

Start: 5700

Pathway: Butanoate metabolism

Fig. 8. Representation of a Protégé screen, as an ontology-developing environment.

criteria. The proposed application has the potential to store, represent, visualize, and analyze genome-annotated data in a dynamic manner. In addition, our application adopted ontology to organize genomic data in a semantic framework, which facilitates data sharing among biology domains, applications, and experts. During recent years, there have been many web-based developments that use ontology as their knowledge base, such as GO [21], Plant Ontology [29], and Microbial Ontology [30], etc. Ontologies define domain concepts and the relationships between them and therefore provide a domain language that is meaningful to both machines and humans [30]. In the present study, we used Protégé, which contains a query interface known as the SPARQL query panel, through which users are able to query particular information from the ontology [30].

However, the proposed application in this study is subject to some limitations. One of the limitations is that the application is not able to represent all genomic data (e.g., sequences) and can not make a query on several genome event layers simultaneously. Another limitation is related to the lack of powerful inference engines to reason genomic data for the purpose of data sharing. Effective inference engines are required to semantically interpret data and exchange them with their intended meanings. One source of limitation is also related to the fact that the application is available on desktop; future research is required to develop an online version of the application and a data server consisting of microbial genome databases for browsing genome-annotated data.

Our work continues toward future plans, including the extension of this application to develop an ontology-based GIS interactive database for rumen microbial communities. Effort is already under way for developing such a comprehensive database to contain a wide spectrum of rumen microbial genomes.

Acknowledgments

This work was conducted as part of a PhD thesis at Ferdowsi University of Mashhad and the National Institute of Genetic Engineering and Biotechnology, Bioinformatics Lab. The authors wish to thank Dr. Mohammadreza Jelokhani-Niaraki for his effort and collaboration as a technical advisor.

References

1. Selvi M. Bioinformatics: an information explosion arena: an overview. *J Adv Libr Inform Sci* 2012;1:192-196.
2. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2014;42:D32-D37.
3. Fernandez-Suarez XM, Galperin MY. The 2013 Nucleic Acids Research Database Issue and the online molecular biology database collection. *Nucleic Acids Res* 2013;41:D1-D7.
4. Kamra DN. Rumen microbial ecosystem. *Curr Sci* 2005;89: 124-135.
5. Hobson PN. *The Rumen Microbial Eco-system*. London: Elsevier Applied Science, 1989.
6. Edwards JE, McEwan NR, Travis AJ, Wallace RJ. 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie Van Leeuwenhoek* 2004;86:263-281.
7. Larue R, Yu Z, Parisi VA, Egan AR, Morrison M. Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and rrs gene sequencing. *Environ Microbiol* 2005;7: 530-543.
8. Yu Z, Yu M, Morrison M. Improved serial analysis of V1 ribosomal sequence tags (SARST-V1) provides a rapid, comprehensive, sequence-based characterization of bacterial diversity and community composition. *Environ Microbiol* 2006;8: 603-611.
9. Altman RB. Building successful biological databases. *Brief Bioinform* 2004;5:4-5.
10. Philippi S. Data and knowledge integration in the life sciences. *Brief Bioinform* 2008;9:451.
11. Olken F, Jagadish HV. Data management for integrative biology. *OMICS* 2003;7:1.
12. Field D, Sansone SA. A special issue on data standards. *OMICS* 2006;10:84-93.
13. Field D, Sansone SA, Garrity GM. Foreword to the special issue on the Fifth Genomic Standards Consortium Workshop. *OMICS* 2008;12:99.
14. Fischer MM, Nijkamp P. *Geographic Information Systems, Spatial Modeling, and Policy Evaluation*. Berlin: Springer-Verlag, 1993.
15. Kozak KH, Graham CH, Wiens JJ. Integrating GIS-based environmental data into evolutionary biology. *Trends Ecol Evol* 2008;23:141-148.
16. Dolan ME, Holden CC, Beard MK, Bult CJ. Genomes as geography: using GIS technology to build interactive genome feature maps. *BMC Bioinformatics* 2006;7:416.
17. Gaševic D, Djuric D, Devedžic V. *Model Driven Architecture and Ontology Development*. Berlin: Springer-Verlag, 2006.
18. Environmental Systems Research Institute. *Network Analysis*. Redlands: ESRI Press, 1995.
19. Hendler J. Agents and the semantic web. *IEEE Intell Syst* 2001;16:30-37.
20. Bard JB, Rhee SY. Ontologies in biology: design, applications and future challenges. *Nat Rev Genet* 2004;5:213-222.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25-29.
22. Environmental Systems Research Institute. *Arcgis10*. Redlands: ESRI Press, 2010.
23. SCBIR. Protégé 3.4.8. Accessed 2015 Jan 2. Available from: http://smi-protegestanfordedu/repos/protege/owl/trunk.
24. University of Maryland Libraries. Introduction to GIS using ArcGIS Desktop 10. College Park: University of Maryland Libraries, 2012. Accessed 2015 Jan 2. Available from: http://www.lib.umd.edu/gov/.
25. Sarajlic S. *Contaminant Hydrogeology Knowledge Base (CHKb) of Georgia, USA*. Atlanta: Georgia State University, 2013.
26. Medina I, Salavert F, Sanchez R, de Maria A, Alonso R, Escobar P, *et al*. Genome Maps, a new generation genome browser. *Nucleic Acids Res* 2013;41:W41-W46.
27. Stalker J, Gibbins B, Meidl P, Smith J, Spooner W, Hotz HR, *et al*. The Ensembl Web site: mechanics of a genome browser. *Genome Res* 2004;14:951-955.
28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, *et al*. The human genome browser at UCSC. *Genome Res* 2002;12:996-1006.
29. Avraham S, Tung CW, Ilic K, Jaiswal P, Kellogg EA, McCouch S, *et al*. The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res* 2008;36:D449-D454.
30. Biswas S, Marwaha S, Malhotra PK, Wahi SD, Dhar DW, Singh R. Building and querying microbial ontology. *Procedia Technol* 2013;10:13-19.