



Streptococcus in Toto

The Orthologous Gene Database

User Guide

Table of contents

<i>Title</i>	<i>Page(s)</i>
Introduction	3
Methods	3-5
Using the <i>Streptococcus in Toto</i> website	5-20
A. Website location	5
B. Browsing the <i>Streptococcus in Toto</i> orthologous groups	6-8
1. By orthologous groups	6
2. By functional classes	6-7
3. By organisms	7-8
C. Searching for orthologous groups	9-14
1. Quick search	9
2. Advanced search	10-15
D. Viewing the alignments for an orthologus group	16-18
E. Resources	19
F. Genomes	20
G. Partners	21
H. References	21

Introduction

With rising numbers of fully sequenced genomes the importance of high-throughput genome annotation is constantly increasing. Although several automatic annotation software systems do exist, their quality and flexibility is still limited. Therefore, we developed the semi-automatic approach called DENTOTI (Etymology: English - dental +Latin *totus* whole, entire +English -i) to annotate all oral pathogen sequences hosted at the ORALGEN database (<http://www.oralgen.lanl.gov/>), which was designed and implemented to support the oral pathogen genetics sequence analysis, metabolic reconstruction and comparative genome analysis.

The DENTOTI approach is based on a common heuristic method of identifying orthologs using bidirectional best hits (Overbeek et al, 1999): if the most similar sequence to protein A in genome 2 is B, and if the most similar sequence to protein B in genome 1 is A, then A and B are bidirectional best hits, and are operationally considered to be orthologs. This relationship is especially strong if the blast E value is very small and if the alignment of the proteins spans a majority of each sequence. Applying the Dentoti approach to 5 closely related genomes from *Streptococcus* phylum; *Streptococcus agalactiae*, *Streptococcus mitis*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Streptococcus sanguinis* and *Streptococcus thermophilus*, we were able to construct the *Streptococcus in Toto* database and transfer gene function from previously annotated *Streptococcus* genomes to approximately 70% of the genes in *Streptococcus sanguinis* and *Streptococcus mitis*.

Of course, relying on this high-throughput bidirectional best-hit database search criterion for annotation can be problematic, as in the case of multi-domain organization of protein and non-orthologous gene displacement, this approach would incorrectly predict a functional linkage. Therefore, some manual curation is necessary to validate the results. Considering this approach for groups of closely related species within the first-pass genome annotation process, it is very useful for the initial assignment of the function to genes.

Methods

A. To isolate genes from different organisms that play identical roles in each organism.

A simple method that has been used for the prediction of orthologous proteins in two organisms is to search for a pair of sequences, Xa in organism Ga and Xb in organism Gb, such that (1) a search of the proteome of Gb with Xa finds Xb as the best hit, and (2) a search of the proteome of Ga with Xb finds Xa as the best hit. The method is called the bidirectional best-hit (BBH) method (Overbeek et al, 1999; Tatusov et al, 1997) (**figure 1**). The relationship is especially strong if the E value is very small and if an alignment of the proteins spans a majority of each sequence. In this study, an e-value less than or equal to 10^{-15} is set as cut-off score. A gene is considered strain-specific if it has no hits with an e-value 10^{-5} or less.

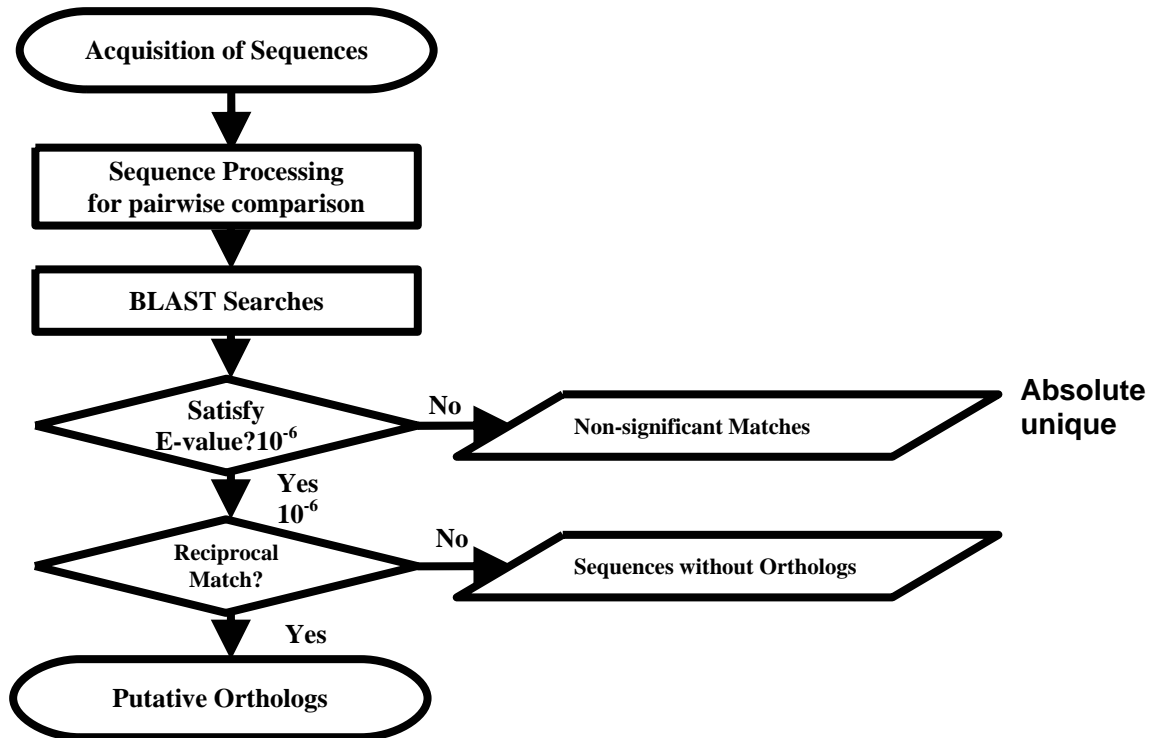


Figure 1: The flowchart of orthologs identification using Bi-directional Best Hit (BBH).

B. To cluster Orthologous Groups (OG) genes together

We begin by forming sets of genes that we call Orthologous Groups (OG). An orthologous group is a set of genes such that each gene in the set is a bidirectional best hit with at least one other gene from another organism in the set. The orthologous group cluster is connected in the sense that one could not split the group without separating two bidirectional best hits.

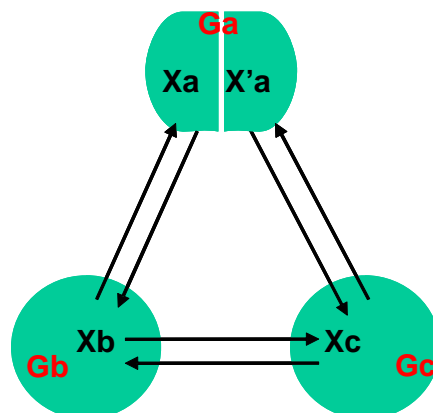


Figure 2: Illustration of the definitions of BBHs between Xa and Xb, Xb and Xc, but not Xc and Xa.

A unique id has been assigned to each orthologous group. The homologous relationships in the gene/protein universe can be represented as a network, in which nodes represent sequences and edges represent similarities between pairs of the sequences.

The orthologous group contains no pair of genes Xa and Xc from organisms Ga and Gc, respectively, such that Xa is a bidirectional best hit with Xb from Gb and Xb is a bidirectional best hit with Xc from Gc, but Xc is not Xa. (see **figure 2**) This condition is frequently seen in cases of paralogs.

C) To transfer the orthologous group annotation for those true orthologs

After the true orthologs were identified, the group annotation was transferred for those closely related species within the first-pass genome annotation process, then subsequently subjected to manual review to validate the annotation.

Using the *Streptococcus in Toto* website

A. Website location

Streptococcus in Toto website (**Figure 3**) is located at the following URL:

<http://www.oralgen.lanl.gov/toto>



Figure 3. *Streptococcus in Toto* website main page.

B. Browsing the *Streptococcus in Toto* orthologous groups

1. Browsing by orthologous groups

Clicking on the “**orthologous groups**” link on the left menu will display all orthologous group definitions. By default, the list is sorted by definition and user can sort the list using orthologous group ID by clicking on the Id link on top right corner of the table (**Figure 4**). In general, column headings are generally clickable, and clicking them will toggle the sort. The first click will sort the table in ascending order by the clicked column and clicking it again will sort the table in descending order. Clicking another column will sort in ascending order by the new column.

Orthologous Group Definition	Id
(3R)-hydroxymyristoyl-(acyl carrier protein) dehydratase	832
1,4-alpha-glucan branching enzyme	1094
1-acylglycerol-3-phosphate O-acyltransferase	295
16S rRNA processing protein	419
2',3'-cyclic-nucleotide 2'-phosphodiesterase	1156
2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	470
2-isopropylmalate synthase	638
23S rRNA (guanine-N1-)-methyltransferase	1207
3-dehydroquinate dehydratase	370
3-dehydroquinate synthase	372
3-deoxy-7-phosphoheptulonate synthase	878
3-hydroxy-3-methylglutaryl-CoA reductase	450

Figure 4. Orthologous groups Definition/Id table

2. Browsing by functional classes

Orthologous groups can be browsed by functional classes. Click on “**functional classes**” on the left menu to bring up the table as shown below (**Figure 5**). Clicking on one of the links shown in **Figure 5** will display orthologous groups belonging to a particular functional class. For example, clicking on “Amino acid biosynthesis” will bring a table of orthologous groups that have broad functional class involved in amino acid biosynthesis (**Figure 6**). Categories and subcategories of functional classes are separated by semi-colons. For example, clicking on ‘Amino Acid Biosynthesis; Aromatic amino acid family’ will provide a specific subset of the results generated by clicking on the ‘Amino Acid Biosynthesis’ link. The table shown can be sorted by definition or by orthologous group ID.

Functional Class
Amino acid biosynthesis
Amino acid biosynthesis; Aromatic amino acid family
Amino acid biosynthesis; Aspartate family
Amino acid biosynthesis; Glutamate family
Amino acid biosynthesis; Histidine family
Amino acid biosynthesis; Pyruvate family
Amino acid biosynthesis; Serine family
Biosynthesis of cofactors, prosthetic groups, and carriers
Biosynthesis of cofactors, prosthetic groups, and carriers; Biotin
Biosynthesis of cofactors, prosthetic groups, and carriers; Folic acid
Biosynthesis of cofactors, prosthetic groups, and carriers; Glutathione and analogs

Figure 5. Orthologous groups grouped into functional classes

Orthologous groups for functional class 'Amino acid biosynthesis'	
Orthologous Group Definition	Id
2-isopropylmalate synthase	638
3-dehydroquinate dehydratase	370
3-dehydroquinate synthase	372
3-deoxy-7-phosphoheptulonate synthase	878
3-isopropylmalate dehydratase, small subunit	635
3-isopropylmalate dehydrogenase	637
3-phosphoshikimate 1-carboxyvinyltransferase (5-enolpyruvylshikimate-3-phosphate synthase)	376
5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	425
acetolactate synthase, large subunit (AHAS)	89
acetolactate synthase, small subunit	90
acetylglutamate kinase	315
alpha-isopropylmalate isomerase large subunit	636
anthranilate synthase, component I	246
anthranilate synthase, component II	247

Figure 6. Table of orthologous groups belonging to 'amino acid biosynthesis' functional class

3. Browsing by organisms

This option allows users to browse orthologous groups specific to a particular organism (**Figure 7**). Click on “**organisms**” on left menu to display the table of organisms available. Currently, there are 11 streptococci genomes available on the *Streptococcus in Toto* website. Clicking on a link to one of these organisms will display a table of all orthologous groups present in this organism (**Figure 8**). Again, this table can be sorted by definition or by orthologous group ID.

Organism
Streptococcus mutans UA159
Streptococcus thermophilus CNRZ1066
Streptococcus sanguinis SK36
Streptococcus agalactiae 2603V/R
Streptococcus mitis NCTC 12261
Streptococcus pneumoniae TIGR4
Streptococcus mutans GS5
Streptococcus mutans UA140
Streptococcus mutans UA787
Streptococcus mutans CH43
Streptococcus mutans T8

Figure 7. List of organisms available on *Streptococcus in Toto* website

Orthologous groups for organism ' <i>Streptococcus mutans</i> UA159'	
Orthologous Group Definition	Id
(3R)-hydroxymyristoyl-(acyl carrier protein) dehydratase	832
1,4-alpha-glucan branching enzyme	1094
1-acylglycerol-3-phosphate O-acyltransferase	295
16S rRNA processing protein	419
2-amino-4-hydroxy-6-hydroxymethylidihydropteridine pyrophosphokinase	470
2-isopropylmalate synthase	638
3-dehydroquinate dehydratase	370
3-dehydroquinate synthase	372
3-deoxy-7-phosphoheptulonate synthase	878
3-hydroxy-3-methylglutaryl-CoA reductase	450
3-isopropylmalate dehydratase, small subunit	635
3-isopropylmalate dehydrogenase	637
3-methyladenine DNA glycosylase I	1020
3-oxoacyl-(acyl-carrier-protein) synthase	834
3-oxoacyl-acyl-carrier-protein reductase / 3-ketoacyl-acyl carrier protein reductase	835
3-oxoacyl-[acyl-carrier-protein] synthase III	839

Figure 8. Table of orthologous groups present in *Streptococcus mutans* UA159

C. Searching for orthologous groups

Users have several options to search the orthologous groups of interest:

1. Quick Search
2. Advanced Search

1. Quick Search

Orthologous groups can be searched by simply entering a search query in the text box provided on top right corner of the page as shown in **Figure 9**. This search enables users to search several fields: gene name, orthologous group definition, and functional class.

kinase Search

[Advanced Search](#)

news

[Oralgen News](#)

Figure 9. Quick search box

For example, entering a search term “kinase” will display a table of orthologous group members that have the term “kinase” in their gene definitions (**Figure 10**). This table will include orthologous group members from the organisms on *Streptococcus in Toto* website. The table can be sorted by group definition or ID.

Search returned 46 orthologous groups

Orthologous Group Definition	Id
2-amino-4-hydroxy-6-hydroxymethylidihydropteridine pyrophosphokinase	470
6-phosphofructokinase	578
acetate kinase	945
acetylglutamate kinase	315
adenylate kinase (ATP-AMP transphosphorylase) (superoxide-inducible protein 16)	969
aspartate kinase / homoserine dehydrogenase	842
bifunctional protein: riboflavin kinase/flavin adenine dinucleotide (FAD) synthase	563
carbamate kinase	1152
cytidylate kinase	331

Figure 10. Quick search results of the term “kinase”

2. Advanced Search

There is also advanced search functionality to limit the results returned. Click on the “Advanced Search” link right below the quick search box (**Figure 9**) to bring up the advanced search page (**Figure 10**).

Search for orthologous groups:	
<p>Select from the following organisms:</p> <ul style="list-style-type: none"> Streptococcus mutans UA159 Streptococcus thermophilus CNRZ1066 Streptococcus sanguinis SK36 Streptococcus agalactiae 2603V/R Streptococcus mitis NCTC 12261 Streptococcus pneumoniae TIGR4 Streptococcus mutans GS5 Streptococcus mutans UA140 Streptococcus mutans UA787 Streptococcus mutans CH43 Streptococcus mutans T8 	<div> <div>Require >></div> <div>Exclude >></div> <div>Remove <<</div> </div> <div> <p>Require organism(s):</p> <input type="text"/> </div> <div> <p>Exclude organism(s):</p> <input type="text"/> </div>
Functional class:	<input type="text"/>
Definition <input type="button" value="contains"/>	<input type="text"/>
Gene name <input type="button" value="contains"/>	<input type="text"/>
<input type="button" value="Reset"/>	<input type="button" value="Submit"/>

Figure 11. Advanced search page.

This search page allows users to delimit the criteria when searching for orthologous groups in streptococci genomes. Users can specify options to include or exclude one or more organisms to be considered when performing a search and also limit the search results by the following values: Functional class, Definition, and Gene name. If no organisms are required or excluded, the results will not be restricted by organism.

There are a few things that users will need to take note of.

- Searches are case **insensitive**
- Organism selection:
 - **OPTIONAL:** If no organisms are required or excluded, the results will not be restricted by organism.
 - To select consecutive organisms, click the first item, press and hold down SHIFT, and then click the last item.
 - To select nonconsecutive organisms, press and hold down CTRL, and then click each item.
 - Clicking or CTRL clicking a selected item will deselect that item.
 - Once selected, click '**Require >>**' or '**Exclude >>**' to move selected organism(s) to the appropriate list.
 - You can also remove selected organisms from the 'Require organism(s)' or 'Exclude organism(s)' lists by clicking '**Remove <<**'.
- Definition and Gene Name searches can be restricted by choosing a search type from the drop-down next to the label:
 - **contains:** will display all matches, regardless of where in the gene name or definition it is found.
 - **starts with:** will only display matches for which gene name or definition begins with the provided text.

- **ends with:** will only display matches for which gene name or definition ends with the provided text.
- **is:** will display exact matches. The full gene name or orthologous gene definition should be provided in the text.
- The '**Reset**' button will clear all fields in the form.

The '**Submit**' button will execute your search query and display the list of orthologous groups that match your criteria. In the event that no matches are found, you will see a message indicating no groups were found, along with a list of criteria used in the search.

Below is an example of a search scenario to help the users familiarized with the advanced search function.

Advanced Search Example 1: To look for orthologous groups containing genes from *Streptococcus mutans* UA159, *Streptococcus thermophilus* CNRZ1066, and *Streptococcus sanguinis* SK36 **ONLY**.

1. Select the three organism names in the selection list:
 - Click on the first item,
 - CTRL click the other two **or** SHIFT click the last one.
2. Click '**Require >>**' to move the selected items to the 'Require organism(s)' list. (Figure 12)

Figure 12. Advanced search page with 3 organisms included and the rest excluded.

3. Select the remaining organisms in the selection list:
 - Click the first item,
 - Scroll to the bottom and SHIFT click the last item.
4. Click '**Exclude >>**' to move the selected items to the 'Exclude organism(s)' list. (Figure 12)

- Click 'Submit' to execute the query and display the orthologous groups. You will see a list of orthologous groups that contain genes from only your selected organisms (**Figure 13**).

Search returned 29 orthologous groups

Orthologous Group Definition	Id
5-nucleotidase precursor	588
ABC transporter ATP-binding protein	396
ABC transporter, ATP-binding protein, possible molybdenum transport system	796
acetylglutamate kinase	315
aconitate hydratase; aconitase A	320
acyl-CoA dehydrogenase	464
amino acid amidohydrolase (hippurate amidohydrolase)	67
citrate synthase	321
conserved hypothetical protein	112
conserved hypothetical protein	128
conserved hypothetical protein	298
conserved hypothetical protein	459
conserved hypothetical protein	512
conserved hypothetical protein	568
conserved hypothetical protein	683
conserved hypothetical protein (possible hydrolase)	236
conserved hypothetical protein (probable integral membrane protein)	1008
histidine kinase	312

Figure 13. Orthologous groups belonging to 3 organisms selected but absent in the rest of the organisms.

Advanced Search Example 2: To look for orthologous groups belonging to functional class “Amino acid biosynthesis” from *Streptococcus mutans* UA159, *Streptococcus thermophilus* CNRZ1066, and *Streptococcus sanguinis* SK36 **ONLY**.

- Follow steps 1 to 4 of Advanced Search Example 1.
- Select “Amino acid biosynthesis” from the drop-down menu for functional class (**Figure 14**) and click submit. Results will show only orthologous groups belonging to the functional class of “Amino acid biosynthesis” (**Figure 15**).

Search for orthologous groups:

Select from the following organisms:

Require >>

Exclude >>

Remove <<

Require organism(s):

Streptococcus mutans UA159
Streptococcus thermophilus CNRZ1066
Streptococcus sanguinis SK36

Exclude organism(s):

Streptococcus agalactiae 2603V/R
Streptococcus mitis NCTC 12261
Streptococcus pneumoniae TIGR4
Streptococcus mutans GS5
Streptococcus mutans UA140

Functional class: Amino acid biosynthesis

Definition contains

Gene name contains

Reset Submit

Figure 14. Advanced search page showing 3 organisms selected and the rest excluded, plus the result limited only to those from functional class of “Amino acid biosynthesis”.

Search returned 4 orthologous groups

Orthologous Group Definition	Id
acetylglutamate kinase	315
N-acetyl-gamma-glutamyl-phosphate reductase (N-acetyl-glutamate-gamma-semialdehyde dehydrogenase)	313
N-acetylornithine aminotransferase	316
ornithine acetyltransferase / N-acetylglutamate synthase	314

Figure 15. Orthologous groups belonging to only 3 organisms and from functional class “Amino acid biosynthesis”.

Advanced Search Example 3: To look for orthologous groups belonging to functional class “Amino acid biosynthesis” and whose definition containing the word “kinase” from *Streptococcus mutans* UA159, *Streptococcus thermophilus* CNRZ1066, and *Streptococcus sanguinis* SK36 **ONLY**.

1. Follow the steps 1 to 4 from Advanced Search Example 1.
2. Select “Amino acid biosynthesis” from the drop-down menu for functional class (**Figure 14**).
3. Type in the word “kinase” in the search box next to “Definition contains” (**Figure 16**) and click submit. Results will return fewer numbers of orthologous groups compared to previous example (**Figure 17**).

Search for orthologous groups:

Select from the following organisms: <div style="border: 1px solid #ccc; height: 100px; width: 100%; margin-top: 5px;"></div>	<input type="button" value="Require >>"/> <input type="button" value="Exclude >>"/> <input type="button" value="Remove <<"/>	<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 5px;"> Require organism(s): Streptococcus mutans UA159 Streptococcus thermophilus CNRZ1066 Streptococcus sanguinis SK36 </div> <div style="border: 1px solid #ccc; padding: 5px;"> Exclude organism(s): Streptococcus agalactiae 2603V/R Streptococcus mitis NCTC 12261 Streptococcus pneumoniae TIGR4 Streptococcus mutans GS5 Streptococcus mutans UA140 </div>
Functional class: Amino acid biosynthesis		
Definition contains kinase		
Gene name contains 		
<input type="button" value="Reset"/>		<input type="button" value="Submit"/>

Figure 16. Advanced search page showing 3 organisms selected and the rest excluded, plus the result limited only to those from functional class of “Amino acid biosynthesis” and containing the word “kinase” in definition.

Search returned 1 orthologous groups

Orthologous Group Definition	Id
acetylglutamate kinase	315

Figure 17. An orthologous group belonging to 3 organisms selected, from functional class “Amino acid biosynthesis”, and whose definition contains the word “kinase”.

D. Viewing the alignments for an orthologous group

Orthologous groups can be viewed by clicking on an orthologous group definition or Id. For example, taking a look at the scenario from the example 1 (**Figure 15**), 4 members of an orthologous group are returned by the advanced search. One can click the definition from the first result, “acetylglutamate kinase” or the orthologous group Id, “315” to view the alignment of members of orthologous group (**Figure 16**). Orthologous group members are shown in **red**. Color legend is shown immediate below of orthologous group alignment. Neighboring genes are colored according to functional classes. Note that the alignment displays only 3 members of the orthologous group since 3 organisms were selected. This particular orthologous group is also only present in these 3 organisms and not the rest of streptococci on the Toto website as specified by the advanced search criteria.

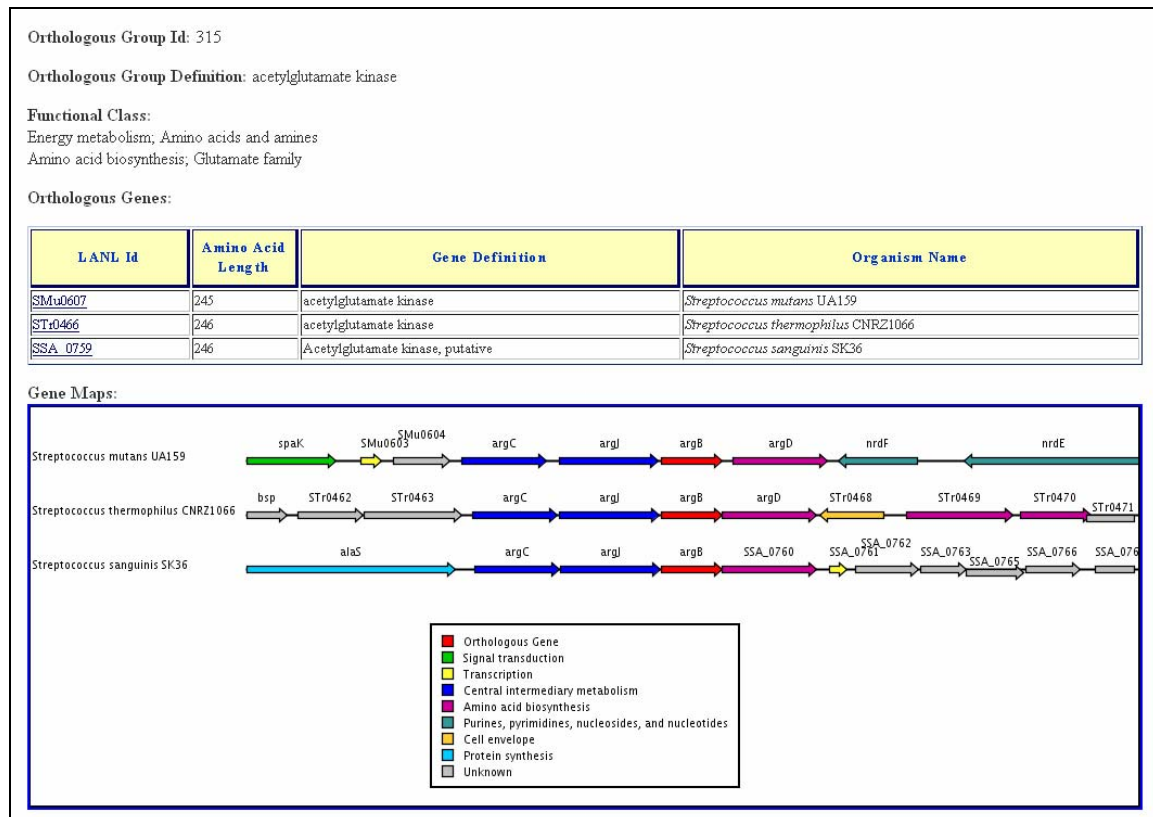


Figure 16. Alignment view of an orthologous group specific to 3 organisms and belonging to functional class “Amino acid biosynthesis”.

From **Figure 16**, users can view quite a number of things. Orthologous Group Id, Orthologous Group Definition, and Functional Class are displayed on top of the page and Orthologous Genes of these 3 organisms are displayed in tabular format. Users can view the LANL Id, Amino Acid Length, and Definition of genes and also the organisms selected in this search. Clicking on the LANL Id of the orthologous groups will bring up

the detail information associated with the gene. For example, clicking on SMu0607 from the table shown in **Figure 16** will bring up the detail page as shown in **Figure 17**.

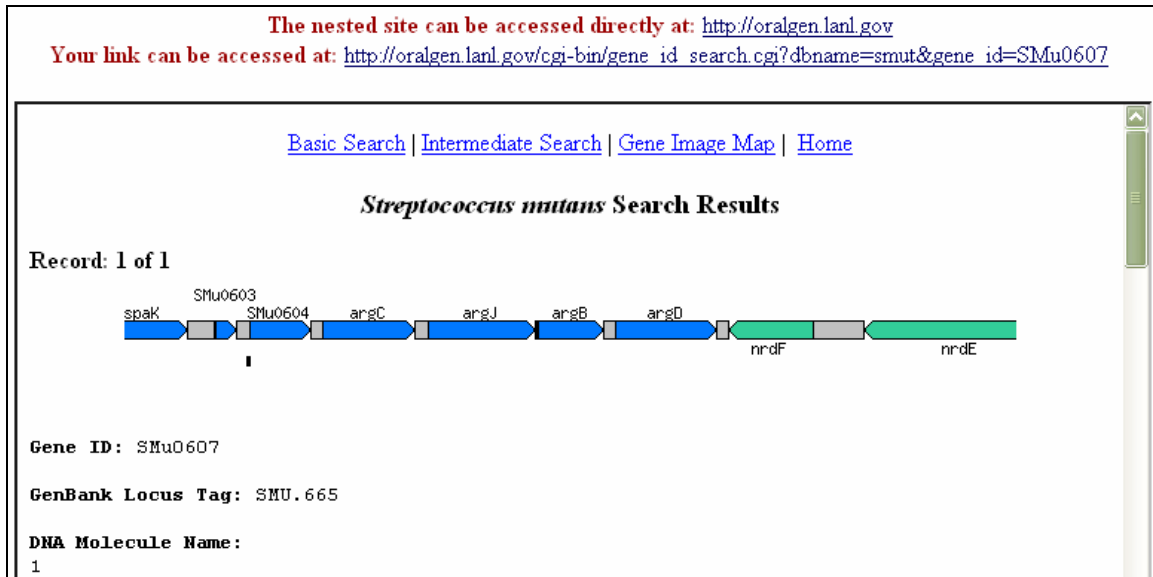


Figure 17. Detailed information about SMu0607.

This page is a nested page originating from Oralgen website (<http://oralgen.lanl.gov>) and links are provided on top of the page for users who want to view the page directly instead of being shown as a nested page. Clicking on the red arrow representing the orthologous group (**Figure 16**) has the same effect as clicking on the LANL Id of this particular orthologous group. Before following the link, users can 'Mouse over' the orthologous group (in red arrow) or neighboring genes in various colors to get a glimpse of what they are (**Figure 18**).

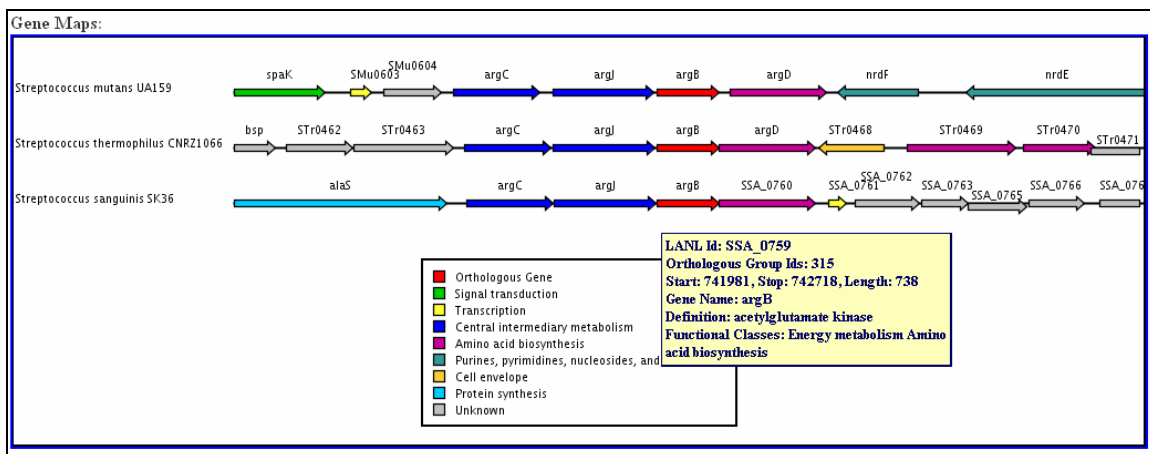


Figure 18. Pop-up menu of an orthologous group member from *S. sanguinis* SK 36.

Hovering the mouse cursor over the arrows will bring up a pop-up menu showing some details associated with a particular gene. For example, in **Figure 18**, placing the mouse over the red arrow belonging to *Streptococcus sanguinis* SK 36 will bring up the pop-up menu showing LANL Id, Orthologous Group Id, Start and Stop coordinates, Length, Gene Name, Definition, and Functional Classes.

Scrolling down the page, user can view the multiple-sequence-alignment of the members of an orthologous group belonging to the 3 organisms selected: *Streptococcus mutans* UA159, *Streptococcus thermophilus* CNRZ1066, and *Streptococcus sanguinis* SK36 (**Figure 19**). This helps users to visually inspect regions of amino acid sequence by conservation. Sequences or residues that are 100% conserved are colored orange and those that are 50% or more conserved are colored blue.

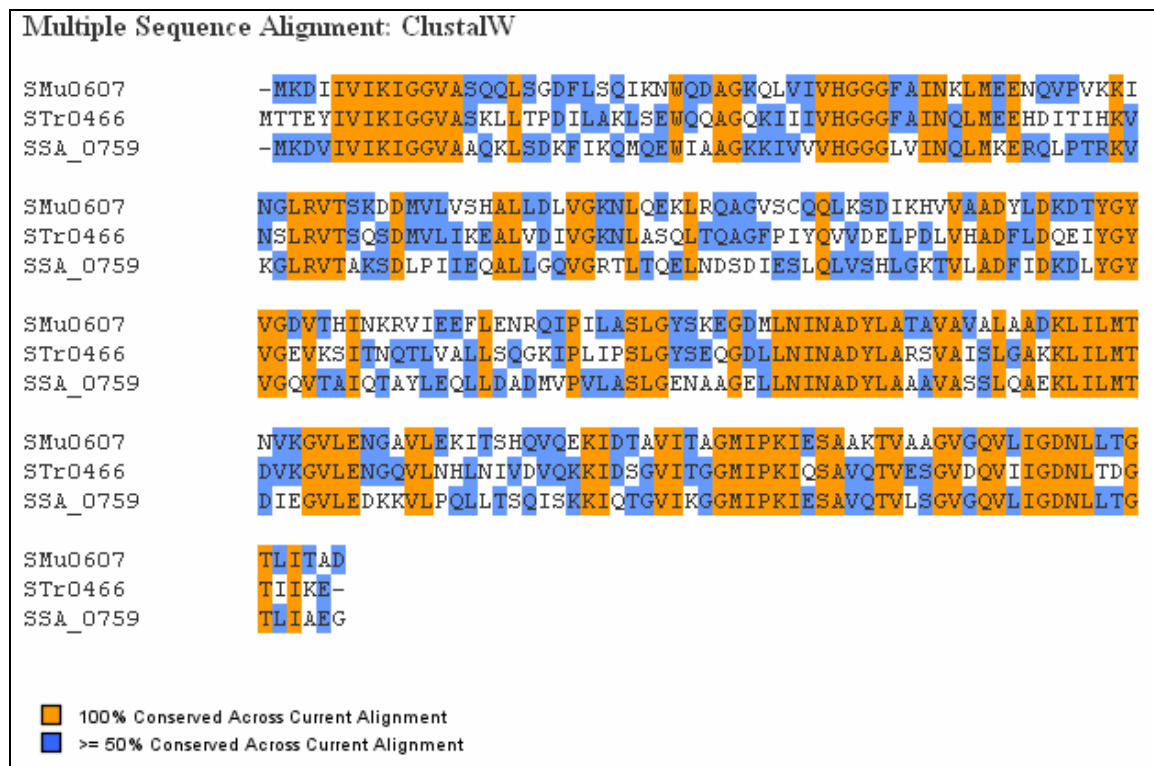


Figure 19. Multiple-sequence-alignment of members of an orthologous group from *S. mutans* UA159, *S. thermophilus* CNRZ1066, and *S. sanguinis* SK36.

E. Resources

References

This link provides molecular subset of recent literature (2003 to 2007) pertaining to the organisms presented on *Streptococcus in Toto* website. Please let us know of any pertinent references that may have been omitted.

2003-2007 Molecular Subset of References
<p>Banas JA, Vickerman MM. Glucan-binding proteins of the oral streptococci. Crit Rev Oral Biol Med. 2003;14(2):89-99. Review. PMID: 12764072</p>
<p>Black C, Allan I, Ford SK, Wilson M, McNab R. Biofilm-specific surface properties and protein expression in oral Streptococcus sanguis. Arch Oral Biol. 2004 Apr;49(4):295-304. PMID: 15003548</p>
<p>Deng H., Ding Y., Fu MD., Xiao XR., Liu J., Zhou T. Purification and characterization of sanguicin ? a Bacteriocin produced by Streptococcus sanguis Sichuan Da Xue Xue Bao Yi Xue Ban 2004 35(4) 555-8 PMID: 15291127</p>
<p>Kerrigan SW, Douglas I, Wray A, Heath J, Byrne MF, Fitzgerald D, Cox D. A role for glycoprotein Ib in Streptococcus sanguis-induced platelet aggregation. Blood. 2002 Jul 15;100(2):509-16. PMID: 12091342</p>
<p>Kreth J, Merritt J, Shi W, Qi F. Competition and coexistence between Streptococcus mutans and Streptococcus sanguinis in the dental biofilm. J Bacteriol. 2005 Nov;187(21):7193-203. PMID: 16237003</p>

Figure 20. Molecular subset of references.

F. Genomes

Streptococcus agalactiae

Streptococcus mitis

Streptococcus mutans

Streptococcus pneumoniae

Streptococcus sanguinis

Streptococcus thermophilus

These links help users to navigate individual genome pages hosted on the Oralgen website (<http://www.oralgen.lanl.gov/>). By clicking on these links, users can access the various components of genome analysis for a particular *Streptococcus* of interest (**Figure 10**).

These pages hosted on Oralgen website are nested inside of *Streptococcus in Toto* website and users can navigate the Oralgen website directly by clicking on one of these links.

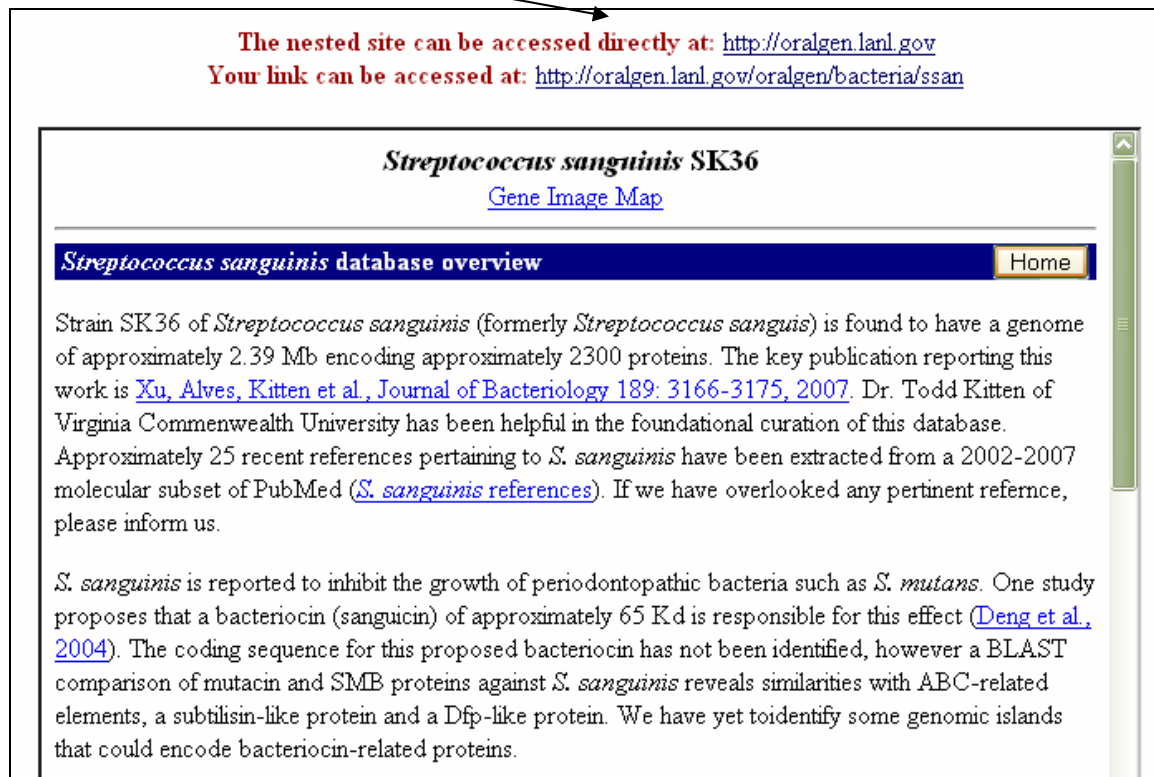


Figure 21. Main page of *Streptococcus sanguinis* SK36.

G. Partners

NIDCR – National Institute of Dental and Craniofacial Research

NIH – National Institute of Health

LANL – Los Alamos National Laboratory

These are URLs to allow users to visit the partner agencies involved in planning and making this website a reality.

H. References

Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 2896-2901

Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science (New York, NY)* **278**: 631-637