

COG08

Organism	Genome File	Protein	Low	High	Strand
Treponema pallidum	AE000520.txt	TP0673	739667	739845	-
Chlamydia trachomatis	AE001273.txt	CT445	516123	516395	-
Chlamydophila pneumoniae CWL029	AE001363.txt	CPn0560	645612	645868	-
Bacillus halodurans	BA000004.txt	BH0109	142624	142707	+
Bacillus subtilis	AL009126.txt	Bsu0092	110953	111044	+
Escherichia coli O157:H7 EDL933	AE005174.txt	Z3665	3319047	3319823	-
Escherichia coli K12	U00096.txt	b2400	2518692	2519613	-
Pasteurella multocida	AE004439.txt	PM1115	1313817	1314570	-
Haemophilus influenzae Rd	L42023.txt	HI0274	308877	309627	-

Parameters

Command line: /homes/gws/blanchem/FootPrinterWeb/FootPrinter_test
 /cse/abstract/blanchem/www/FootPrinterWeb/__webquery__.fasta0.376473090274050.5
 96245328648351
 /cse/abstract/blanchem/www/FootPrinterWeb/__webquery__0.376473090274050.596245
 328648351.tree -details -triplet_filtering -post_filtering -size 12 -max_mutations 3 -
 max_mutations_per_branch 2 -losses
 /homes/gws/blanchem/FootPrinterWeb/universal12.config -compute_branch_lengths -
 loss_cost 1 -sequence_type upstream -subregion_size 100 -position_change_cost 1

Number of mutations of interest

0 mutations over a tree of size at least 4

1 mutations over a tree of size at least 8

2 mutations over a tree of size at least 12

3 mutations over a tree of size at least 16

The phylogenetic tree is



(((((GI|AE005174|3319048-3319347:0,GI|U00096|2518693-2518992:0.15628):5.15281,(GI|AE004439|1313818-1314117:1.54,GI|L42023|308878-309177:1.64983):0.59069):0,GI|AE000520|739668-739844:9.24557):0.79445,((GI|AE001273|516124-516394:7.51414,GI|AE001363|645613-645867:5.04256):1.43954,(GI|BA000004|142625-142706:4.99725,GI|AL009126|110954-111043:4.50504):2.22446):0):0;
The total tree size is 44.8386

Results

http://abstract.cs.washington.edu/~blanchem/FootPrinterWeb/webquery_fasta0.376473090274050.596245328648351.main.html

Comments

This set produced the most interesting results. The most significant motifs were of length 12 and spanned about half of the sequences in the tree.

COG13

Organism	Genome File	Protein	Low	High	Strand
Treponema pallidum	AE000520.txt	TP1017	1109754	1109926	-
Bacillus halodurans	BA000004.txt	BH1267	1362781	1363112	+
Bacillus subtilis	AL009126.txt	Bsu2736	2800080	2800411	-
Pseudomonas aeruginosa PA01	AE004091.txt	PA0903	986696	986818	+
Lactococcus lactis subsp. lactis	AE005176.txt	L0343	1780987	1781253	-
Neisseria meningitidis serogroup A strain Z2491	AL157959.txt	NMA1788	1729809	1730014	+
Neisseria meningitidis serogroup B strain MC58	AE002098.txt	NMB1595	1656508	1656713	+
Pasteurella multocida	AE004439.txt	PM1287	1480912	1481152	+
Synechocystis sp. PCC 6803	AB001339.txt	slI0362	2144246	2144427	-

Parameters

[1] 25266 Command line: /homes/gws/blanchem/FootPrinterWeb/FootPrinter_test /cse/abstract/blanchem/www/FootPrinterWeb/__webquery__.fasta0.5406578128577910.184573360005729 /cse/abstract/blanchem/www/FootPrinterWeb/__webquery__0.5406578128577910.184573360005729.tree -details -triplet_filtering -post_filtering -size 6 -max_mutations 3 -max_mutations_per_branch 2 -losses /homes/gws/blanchem/FootPrinterWeb/universal6.config -compute_branch_lengths -loss_cost 1 -sequence_type upstream -subregion_size 100 -position_change_cost 1 FootPrinter is running...

Number of mutations of interest

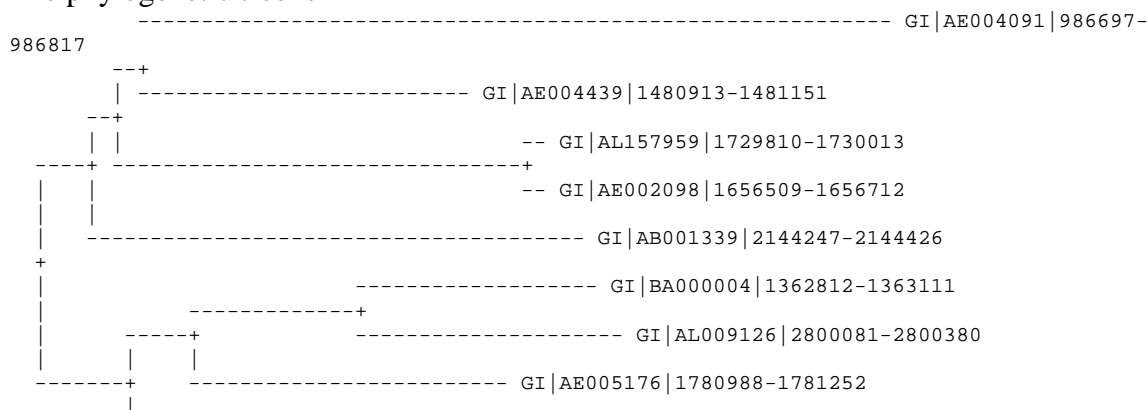
0 mutations over a tree of size at least 18

1 mutations over a tree of size at least 30

2 mutations over a tree of size at least 42

3 mutations over a tree of size at least 54

The phylogenetic tree is



----- GI|AE000520|1109755-1109925

((((GI|AE004091|986697-986817:9.83519,GI|AE004439|1480913-1481151:4.25531):0,(GI|AL157959|1729810-1730013:0,GI|AE002098|1656509-1656712:0):5.15184):0,GI|AB001339|2144247-2144426:6.3348):0.40167,(((GI|BA000004|1362812-1363111:3.06678,GI|AL009126|2800081-2800380:3.31119):1.95047,GI|AE005176|1780988-1781252:3.97479):0.68331,GI|AE000520|1109755-1109925:5.97556):0.97763):0;
The total tree size is 45.9078

Results

http://abstract.cs.washington.edu/~blanchem/FootPrinterWeb/_webquery_.fasta0.5406578128577910.184573360005729.main.html

Comments

Not very interesting. No motifs of significance appearing on more than 2 or 3 sequences.

COG50

Organism	Genome File	Protein	Low	High	Strand
Treponema pallidum	AE000520.txt	TP0187	203018	203082	+
Escherichia coli O157:H7 EDL933	AE005174.txt	Z4697	4258457	4258528	-
Escherichia coli K12	U00096.txt	b3339	3468966	3469037	-
Escherichia coli O157:H7 EDL933	AE005174.txt	Z5553	5053131	5053707	+
Neisseria meningitidis serogroup A strain Z2491	AL157959.txt	NMA0149	136055	136472	-
Neisseria meningitidis serogroup A strain Z2492	AL157959.txt	NMA0134	118729	118817	-
Neisseria meningitidis serogroup B strain MC58	AE002098.txt	NMB0124	131847	132264	+
Haemophilus influenzae Rd	L42023.txt	HI0632	671059	671764	+

Parameters

[1] 30979 Command line: /homes/gws/blanchem/FootPrinterWeb/FootPrinter_test /cse/abstract/blanchem/www/FootPrinterWeb/__webquery__.fasta0.3627390867450980.217966628013514 /cse/abstract/blanchem/www/FootPrinterWeb/__webquery__0.3627390867450980.217966628013514.tree -details -triplet_filtering -post_filtering -size 6 -max_mutations 3 -max_mutations_per_branch 2 -sequence_type upstream -subregion_size 100 -position_change_cost 1

FootPrinter is running...

Warning: Your input tree was not a binary tree. It has been arbitrarily binarized. The phylogenetic tree is

```

----- GI|AE005174|4258458-4258527
|-----+
|----- GI|U00096|3468967-3469036
+
|----- GI|AE005174|5053407-5053706
|-----+
|----- GI|L42023|671464-671763
|-----+
GI|AL157959|136056-136355
|-----+
|-----+
GI|AL157959|118730-118816
|-----+
|----- GI|AE002098|131964-132263
|-----+
|----- GI|AE000520|203019-203081

```

Results

http://abstract.cs.washington.edu/~blanchem/FootPrinterWeb/_webquery_.fasta0.3627390867450980.217966628013514.main.html

Comments

Not very interesting. Couldn't find any motifs when allowing for losses across branches. Disallowing loss necessitated shortening the motif length to a trivial degree (length 6).