Phylogenomic Analyses to Detect Prokaryote to Eukaryote Horizontal Gene Transfer

T. Martin Embley

Newcastle University

Overview

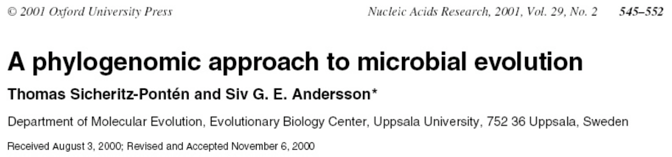
- The approach used - detection of 'recent' HGT based upon phylogenetic trees
- The computer programmes used and some of their features
- Some results from analysis of the genome of *Entamoeba histolytica*
- Future directions and challenges

Why study HGT?

- What is the role of HGT, outside of endosymbiosis, in the evolution of eukaryotes?
- How much HGT has occurred? What types of genes? Where from?
- How has HGT affected the evolution of parasitic eukaryotes? Does HGT provide drug targets?

How have we chosen to look for HGT?

- A semi-automated phylogenomics approach based upon a published computer programme called PyPhy
- Use trees to detect HGT in a two-stage process:
  - Simple parsimony or p-distance trees from edited alignments
  - Bayesian trees with more sophisticated phylogenetic models as a second stage screen
  - Distance-based bootstrapping to gain a measure of confidence in groups on trees



© 2001 Oxford University Press Nucleic Acids Research, 2001, Vol. 29, No. 2 545–552

A phylogenomic approach to microbial evolution

Thomas Sicheritz-Pontén and Siv G. E. Andersson*

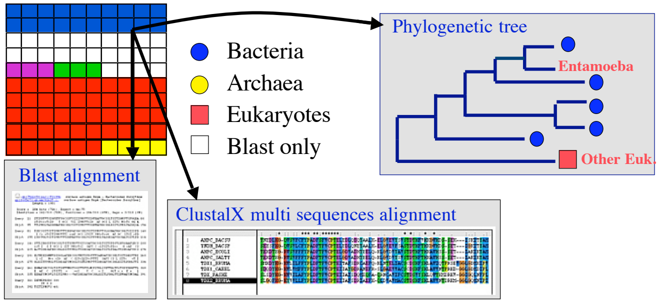
Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, 752 36 Uppsala, Sweden

Received August 3, 2000; Revised and Accepted November 6, 2000

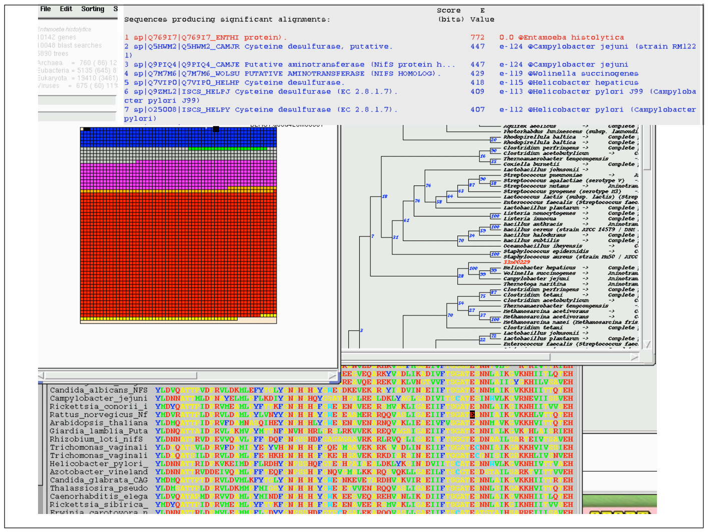
PyPhy - the Primary Screen

- The database
  - Translated (Tr)EMBL
  - Swissprot entries (for annotation)
- BlastP searches using query genome to recover homologues
  - user defined cut-off values (e.g. at least 40% similarity over >70<200% of the length of query sequence)
- Align sequences using ClustalX and trim the alignment using GBLOCKS
- Construct bootstrapped P-distance or parsimony trees
- Display results as a grid with coloured squares indicating neighbor relationships

PyPhy graphical interface



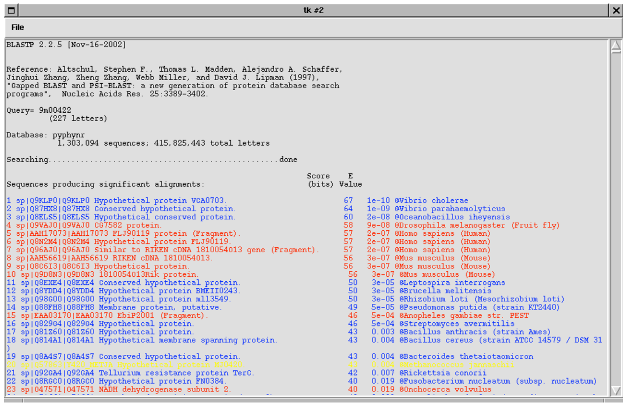
- ● Bacteria
- ● Archaea
- 🟥 Eukaryotes
- □ Blast only


File
Edit
Sorting
Entamoeba histolytica
10142 genes
10048 blast searches
5890 trees
Archaea = 760 (86) 12
Eubacteria = 5135 (645) 8
Eukaryota = 19410 (3461)
Viruses = 675 (60) 11%
Rhodopirellula baltica -> Complete
Clostridium perfringens -> Complete
Clostridium acetobutylicum
Thermoanaerobacter tengcongensis
Coxiella burnetii -> Complete
Lactobacillus johnsonii ->
Streptococcus pneumoniae
Streptococcus agalactiae (serotype V)
Streptococcus mutans
Streptococcus pyogenes (serotype M3)
Lactococcus lactis (subsp. lactis)
Enterococcus faecalis (Streptococcus faec
Lactobacillus plantarum -> Complete
Methanosarcina acetivorans
Methanosarcina mazei
Clostridium tetani -> Complete

```
                                                              Score    E
Sequences producing significant alignments:                   (bits) Value

1 sp|Q769I7|Q769I7_ENTHI protein).                             772   0.0 @Entamoeba histolytica
2 sp|Q5HWM2|Q5HWM2_CAMJR Cysteine desulfurase, putative.       447   e-124 @Campylobacter jejuni (strain RM122
1)
3 sp|Q9PIQ4|Q9PIQ4_CAMJE Putative aminotransferase (NifS protein h...   447   e-124 @Campylobacter jejuni
4 sp|Q7M7M6|Q7M7M6_WOLSU PUTATIVE AMINOTRANSFERASE (NIFS HOMOLOG).   429   e-119 @Wolinella succinogenes
5 sp|Q7VIP0|Q7VIP0_HELHP Cysteine desulfurase.                 418   e-115 @Helicobacter hepaticus
6 sp|Q9ZML2|ISCS_HELPJ Cysteine desulfurase (EC 2.8.1.7).      409   e-113 @Helicobacter pylori J99 (Campyloba
cter pylori J99)
7 sp|O25008|ISCS_HELPY Cysteine desulfurase (EC 2.8.1.7).      407   e-112 @Helicobacter pylori (Campylobacter
pylori)
```

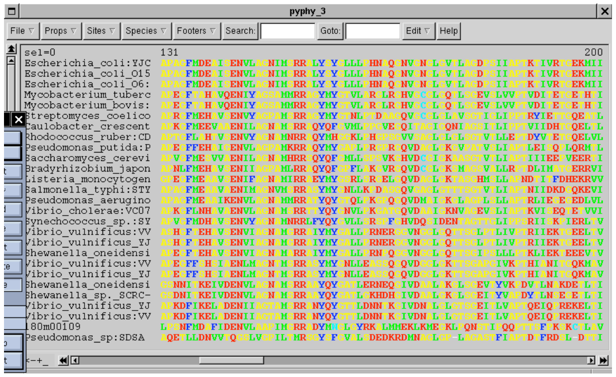
```
Candida_albicans_NFS  YLDVQATTPVDPRVLDKMLEFYTGLYGNPHSHAYGWETDKEVEKARTYIADVINAEIIFTSGATETNNMAIKGVKKHIITTQTEH
Campylobacter_jejuni  YLDNNATTMLDPNAYELMLPFLKDIYGNPNSHQYGSATHPALREALDKLYAGLGADIVITSCATESINWVLKGVRNEVIISSVEH
Rickettsia_conorii_i  YMDYQATTPIDPRVMEAMLPYFTTKFGNPHSHSFGWEAENAVEEARSMVAKLIGAEIIFTSGATESNNLAIKGIKNHIITVVSEH
Rattus_norvegicus_Nf  YMDVRATTPLDPRVLDAMLPYLVNYYGNPHSHAYGWESEAAMERARQQVASLIGAEIIFTSGATESNNIAIKGVKKHLVTTQTEH
Arabidopsis_thaliana  YLDMQATTPIDPRVFDAMNASQIHEYGNPHSHLYGWEAENAVENARNQVAKLIEAEIVFVSGATEANNMAVKGVKKHVITTQTEH
Giardia_lamblia_Puta  YLDNQATTAIDPRVLAKMVPYMTTNFANVHSHRLGRSLRKVVEKAREQVAGAIGAEIIFTSGATESNNIALKGVKPHLAISRIEH
Rhizobium_loti_nifS   YLDNNATTRVDPEVVQAVLPFFTDQFGNPSSHDFGASAGAAVRKARLRVQTLIGAEITFTSGGTESDNAAILSARTEIVTSAVEH
Trichomonas_vaginali  YLDTQATSVLDPRVFDTMIPYETYVHGNAHSHCFGQEAMAAVEKARKSVADLINAEIIFTSGATECNNIAIKGAKKHVIVSSIEH
Trichomonas_vaginali  YLDAQATSILDPRVLDAMLPFETHKHGNPHSHCFGKEAHGAVEKARKDIARAINAEIIFTGGATECCNISIKGSKKHLIVSNVEH
Helicobacter_pylori_  YLDNNATTRIDPKVKEIMDPFLRDHYGNPSSHQFGTETHPAIAEALDKLYKGINADVIITSCATESNNWVLKGVKNHIVTTVAEH
Azotobacter_vineland  YLDNNATTRVDDEIVQAMLPFFTEQFGNPSSHSFGNQVGMALKKARQSVQKLLGAEILFTSCGTESDSTAILSARKTVITTVVEH
Candida_glabrata_CAG  YMDMQATTPTDPRVLDVMLKFYTGLYGNPHSHAYGWETNKEVETARDHVAKVIRAEIIFTSGATESNNLAIKGVKKHIITTRTEH
Thalassiosira_pseudo  YLDMSATTPLDPRVLDKMMPFMIGSYGNPHSHSYGWESETVVENARQQVASLIGAEIIFTSGATESNNLSIKGAKKHVITTQTEH
Caenorhabditis_elega  YLDVQATAPMDPRVVDAMLPYMINDFGNPHSHSYGWKAEEGVEQAREHVANLIKADIIFTSGATESNNLAIKGVKNHIITLQTEH
Rickettsia_sibirica_  YMDYQATTPIDPRVMEAMLPYFTTKFGNPHSHSFGWEAENAVEEARSMVAKLIGAEIIFTSGATESNNLAIKGIKNHIITVVSEH
```

BlastP search



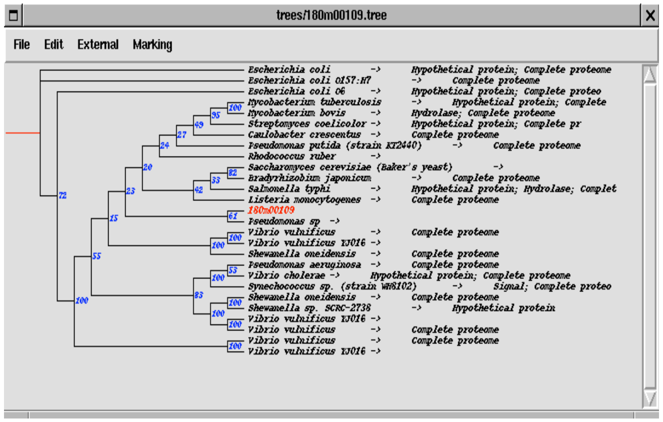
```
BLASTP 2.2.5 [Nov-16-2002]


Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.

Query= 9m00422
         (227 letters)

Database: pyphynr
           1,303,094 sequences; 415,825,443 total letters

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

                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

1 sp|Q9KLP0|Q9KLP0 Hypothetical protein VCA0703.                    67   1e-10 @Vibrio cholerae
2 sp|Q87HX8|Q87HX8 Conserved hypothetical protein.                  64   1e-09 @Vibrio parahaemolyticus
3 sp|Q8ELS5|Q8ELS5 Hypothetical conserved protein.                  60   2e-08 @Oceanobacillus iheyensis
4 sp|Q9VAJ0|Q9VAJ0 CG7582 protein.                                  58   9e-08 @Drosophila melanogaster (Fruit fly)
5 sp|AAH17073|AAH17073 FLJ90119 protein (Fragment).                 57   2e-07 @Homo sapiens (Human)
6 sp|Q8N2M4|Q8N2M4 Hypothetical protein FLJ90119.                   57   2e-07 @Homo sapiens (Human)
7 sp|Q96AJ0|Q96AJ0 Similar to RIKEN cDNA 1810054013 gene (Fragment).  57   2e-07 @Homo sapiens (Human)
8 sp|AAH56619|AAH56619 RIKEN cDNA 1810054013.                       56   3e-07 @Mus musculus (Mouse)
9 sp|Q8C6I3|Q8C6I3 Hypothetical protein.                            56   3e-07 @Mus musculus (Mouse)
10 sp|Q9D8N3|Q9D8N3 1810054013Rik protein.                          56   3e-07 @Mus musculus (Mouse)
11 sp|Q8EXE4|Q8EXE4 Conserved hypothetical protein.                 50   3e-05 @Leptospira interrogans
12 sp|Q8YDD4|Q8YDD4 Hypothetical protein BMEII0243.                 50   3e-05 @Brucella melitensis
13 sp|Q98G00|Q98G00 Hypothetical protein mll3549.                   50   3e-05 @Rhizobium loti (Mesorhizobium loti)
14 sp|Q88FH8|Q88FH8 Membrane protein, putative.                     49   5e-05 @Pseudomonas putida (strain KT2440)
15 sp|EAA03170|EAA03170 EbiP2001 (Fragment).                        46   5e-04 @Anopheles gambiae str. PEST
16 sp|Q829G4|Q829G4 Hypothetical protein.                           46   5e-04 @Streptomyces avermitilis
17 sp|Q81Z60|Q81Z60 Hypothetical protein.                           43   0.003 @Bacillus anthracis (strain Ames)
18 sp|Q814A1|Q814A1 Hypothetical membrane spanning protein.         43   0.004 @Bacillus cereus (strain ATCC 14579 / DSM 31
)
19 sp|Q8A4S7|Q8A4S7 Conserved hypothetical protein.                 43   0.004 @Bacteroides thetaiotaomicron
20 sp|Q57863|Y420_METJA Hypothetical protein MJ0420.                43   0.004 @Methanococcus jannaschii
21 sp|Q92GA4|Q92GA4 Tellurium resistance protein TerC.              42   0.007 @Rickettsia conorii
22 sp|Q8RGC0|Q8RGC0 Hypothetical protein FN0384.                    40   0.019 @Fusobacterium nucleatum (subsp. nucleatum)
23 sp|O47571|O47571 NADH dehydrogenase subunit 2.                   40   0.019 @Onchocerca volvulus
```

G-blocked Alignment



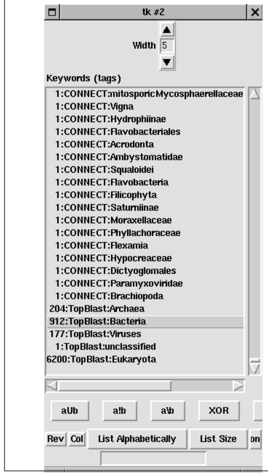
```
sel=0                131                                                                  200
Escherichia_coli:YJC APAGFMDEAISENVLAGNIMSRRALYSYGLLLPHNAQGNVGNGLGVTLAGDPSIIAPTKTIVRTGEKMII
Escherichia_coli_O15 APAGFMDEAISENVLAGNIMSRRALYSYGLLLPHNAQGNVGNGLGVTLAGDPSIIAPTKTIVRTGEKMII
Escherichia_coli_O6: APAGFMDEAISENVLAGNIMSRRALYSYGLLLPHNAQGNVGNGLGVTLAGDPSIIAPTKTIVRTGEKMII
Mycobacterium_tuberc APEGFTAHAVQENIYAGSAMMRRAGYMYGTVLARGLRGHVGCGLGQTLSGEVSLVVPTVDITETGETHTI
Mycobacterium_bovis: APEGFTAHAVQENIYAGSAMMRRAGYMYGTVLARGLRGHVGCGLGQTLSGEVSLVVPTVDITETGETHTI
Streptomyces_coelico APRGFMEHAVSENVYAGPAMSRRAGYMYGTNLPTDAAGQVGCGLGLAVSGTIGLIPPTRYIETTGQEATL
Caulobacter_crescent APKGFMEEVAAENILAGNAMSRRAQYQFGVMLPPGVEGQITAGIGQNIAGSITLIPPTVIIDHTGQELTL
Rhodococcus_ruber:CD APTGFLAHTVSENVYAGNAMNRRAQYMHGGKLPHGPSGVVSAGLGLGLSGTVTLLEPTDYVTETGQELVL
Pseudomonas_putida:P APEGFFEHAIGENVLAGPAMKRRAQYMYGAPLPRGPRGQVDAGLGKGVPATVSLIAPTLEISQPLQRMTL
Saccharomyces_cerevi APVGFMESVVAENILAGNAMHRRSQYQFGMLLSPSVKGHVDCGIGKAASGTVTLIAPTIIIEEPVEERTI
Bradyrhizobium_japon APNLFMEHAVSENIIAGPAMLRRAQYQFGPFLAKGVRGQVDCGLGKSMAGSVALLRPTDLIMATGERRVI
Listeria_monocytogen GPEGFMEAAVSENIFAGNAMIRRAEYMYGSRLSRGELGQVDAGLGKTASGHMSLLAPNDTITFDHEKRVV
Salmonella_typhi:STY APAGFMEAAVAENIMAGNVMSRRASYMYGNLLKPDASGQVGAGLGTTTSGTVTLIAPTNIIDKDGQKEVI
Pseudomonas_aerugino APAGFMEAAIKENVLAGNAMMRRATYQYGTQLPKGPQGQVDMAIGKGLAGPLSLLAPTRLIEGEGEDLVL
Vibrio_cholerae:VC07 APKGFLNHAVAENVLAGNAMSRRTTYQYGNVLPKGATGQVDAAIGKNVAGEVSLIAPTKVISEQTETVVI
Synechococcus_sp.:SY APVGFMDHAVSENVYAGNAMNRRLFYQYGVLLSRSPFGHVDQSIDENTAGTTGLIPPSRIISKPIERLTV
Vibrio_vulnificus:VV ASHGFTEHAVSENVIAGNAMGRRAIYMYGALLPRNERGGVNGGLGQTTSGLPTLIVPTRIIEKTGEELTV
Vibrio_vulnificus_YJ ASHGFTEHAVSENVIAGNAMGRRAIYMYGALLPRNERGGVNGGLGQTTSGLPTLIVPTRIIEKTGEELTV
Shewanella_oneidensi APEGFTEHAVSENVIAGNAMARRAIYMYGALLPRNPQGGVNGGLGQTTSGIPSLLLPTKLIEKTEEEVTV
Vibrio_vulnificus:VV APEGFFSHSIAENLMAGNTMSRRASYMYGNLLEASSQGQVDGGLGKTTSGAPGIVKPTHIANITGQKMAV
Vibrio_vulnificus_YJ APEGFFSHSIAENLMAGNTMSRRASYMYGNLLEASSQGQVDGGLGKTTSGAPGIVKPTHIANITGQKMAV
Shewanella_oneidensi GSNNITKEIVDENVLAGNAMSRRAAYQYGATLERNEQGIVDAALAKGLSGEVTYVKPDVTLNAKDETLTI
Shewanella_sp._SCRC- GSDNITKEIVDENVLAGNAMSRRAAYQYGATLGKHDHGIVDAALGKGLSGEITYVAPDYTLNSEGETLTI
Vibrio_vulnificus_YJ APKDFIKELADENIIAGTAMSRRANYQYGTTLDNNTKGIVDNALGLGTSGEITLVAPTQEIQPREKELTI
Vibrio_vulnificus:VV APKDFIKELADENIIAGTAMSRRANYQYGTTLDNNTKGIVDNALGLGTSGEITLVAPTQEIQPREKELTI
180m00109            LPSNFMDAFIDENVLAAPIMGRRADYMWGLGYRKALMMEKLKMEGKLPQNSTIPQQPTTSFPKSKCTLAV
Pseudomonas_sp:SDSA  AQETLLDNVVTQGSLVGPILTMRSGYSFGVALSDEDKRDMNAGLGP-LAGASTFIAPTDTFRDSL-DTTI
```

Bootstrapped P-distance trees



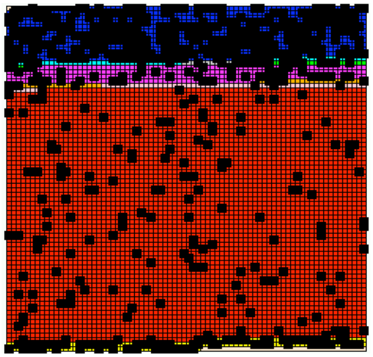
```
Escherichia coli           ->     Hypothetical protein; Complete proteome
Escherichia coli O157:H7          ->     Complete proteome
Escherichia coli O6        ->     Hypothetical protein; Complete proteo
Mycobacterium tuberculosis        ->     Hypothetical protein; Complete
Mycobacterium bovis        ->     Hydrolase; Complete proteome
Streptomyces coelicolor ->        Hypothetical protein; Complete pr
Caulobacter crescentus  ->        Complete proteome
Pseudomonas putida (strain KT2440)      ->     Complete proteome
Rhodococcus ruber          ->
Saccharomyces cerevisiae (Baker's yeast)       ->
Bradyrhizobium japonicum          ->     Complete proteome
Salmonella typhi           ->     Hypothetical protein; Hydrolase; Complet
Listeria monocytogenes  ->        Complete proteome
180m00109
Pseudomonas sp  ->
Vibrio vulnificus          ->     Complete proteome
Vibrio vulnificus YJ016 ->
Shewanella oneidensis   ->        Complete proteome
Pseudomonas aeruginosa  ->        Complete proteome
Vibrio cholerae ->     Hypothetical protein; Complete proteome
Synechococcus sp. (strain WH8102)       ->     Signal; Complete proteo
Shewanella oneidensis   ->        Complete proteome
Shewanella sp. SCRC-2738          ->     Hypothetical protein
Vibrio vulnificus YJ016 ->
Vibrio vulnificus          ->     Complete proteome
Vibrio vulnificus          ->     Complete proteome
Vibrio vulnificus YJ016 ->
```

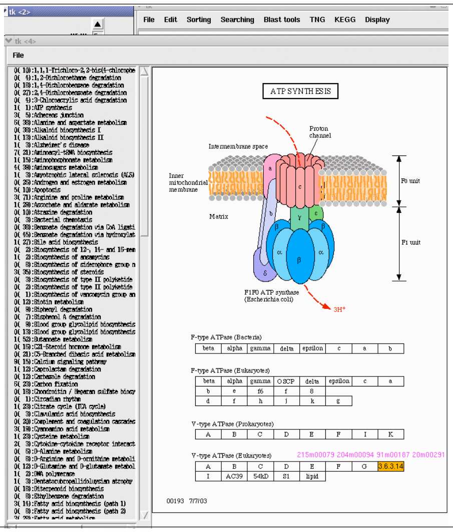
Proteins where the top BlastP hit is a bacterium



Keywords (tags)

```
   1:CONNECT:mitosporicMycosphaerellaceae
   1:CONNECT:Vigna
   1:CONNECT:Hydrophiinae
   1:CONNECT:Flavobacteriales
   1:CONNECT:Acrodonta
   1:CONNECT:Ambystomatidae
   1:CONNECT:Squaloidei
   1:CONNECT:Flavobacteria
   1:CONNECT:Filicophyta
   1:CONNECT:Saturniinae
   1:CONNECT:Moraxellaceae
   1:CONNECT:Phyllachoraceae
   1:CONNECT:Flexamia
   1:CONNECT:Hypocreaceae
   1:CONNECT:Dictyoglomales
   1:CONNECT:Paramyxoviridae
   1:CONNECT:Brachiopoda
 204:TopBlast:Archaea
 912:TopBlast:Bacteria
 177:TopBlast:Viruses
   1:TopBlast:unclassified
6200:TopBlast:Eukaryota
```





```
0( 10):1,1,1-Trichloro-2,2-bis(4-chlorophe
0(  4):1,2-Dichloroethane degradation
0( 18):1,4-Dichlorobenzene degradation
0( 27):2,4-Dichlorobenzoate degradation
0(  4):3-Chloroacrylic acid degradation
1(  1):ATP synthesis
3(  5):Adherens junction
6( 38):Alanine and aspartate metabolism
0( 38):Alkaloid biosynthesis I
1( 13):Alkaloid biosynthesis II
1(  3):Alzheimer's disease
7( 21):Aminoacyl-tRNA biosynthesis
1( 15):Aminophosphonate metabolism
4( 39):Aminosugars metabolism
1(  3):Amyotrophic lateral sclerosis (ALS)
0( 26):Androgen and estrogen metabolism
5( 10):Apoptosis
3( 71):Arginine and proline metabolism
1( 29):Ascorbate and aldarate metabolism
0( 10):Atrazine degradation
0(  3):Bacterial chemotaxis
0( 38):Benzoate degradation via CoA ligati
0( 45):Benzoate degradation via hydroxylat
1( 27):Bile acid biosynthesis
0(  2):Biosynthesis of 12-, 14- and 16-mem
1(  2):Biosynthesis of ansamycins
0(  8):Biosynthesis of siderophore group n
3( 35):Biosynthesis of steroids
0(  3):Biosynthesis of type II polyketide
0(  2):Biosynthesis of type II polyketide
0(  1):Biosynthesis of vancomycin group an
0( 12):Biotin metabolism
0(  9):Biphenyl degradation
0(  7):Bisphenol A degradation
0(  9):Blood group glycolipid biosynthesis
0( 13):Blood group glycolipid biosynthesis
1( 52):Butanoate metabolism
0( 16):C21-Steroid hormone metabolism
0( 21):C5-Branched dibasic acid metabolism
9( 15):Calcium signaling pathway
1( 12):Caprolactam degradation
0( 12):Carbazole degradation
6( 23):Carbon fixation
0( 18):Chondroitin / Heparan sulfate biosy
0(  1):Circadian rhythm
1( 23):Citrate cycle (TCA cycle)
0(  3):Clavulanic acid biosynthesis
0( 20):Complement and coagulation cascades
3( 19):Cyanoamino acid metabolism
1( 23):Cysteine metabolism
2(  3):Cytokine-cytokine receptor interact
0(  6):D-Alanine metabolism
0(  8):D-Arginine and D-ornithine metaboli
0( 12):D-Glutamine and D-glutamate metabol
1(  2):DNA polymerase
1(  3):Dentatorubropallidoluysian atrophy
0( 18):Diterpenoid biosynthesis
0(  8):Ethylbenzene degradation
3( 14):Fatty acid biosynthesis (path 1)
0(  8):Fatty acid biosynthesis (path 2)
3( 28):Fatty acid metabolism
```

F-type ATPase (Bacteria)

beta	alpha	gamma	delta	epsilon	c	a	b

F-type ATPase (Eukaryotes)

beta	alpha	gamma	OSCP	delta	epsilon	c	a
b	e	f6	f	8			
d	f	h	j	k	g		

V-type ATPase (Prokaryotes)

A	B	C	D	E	F	I	K

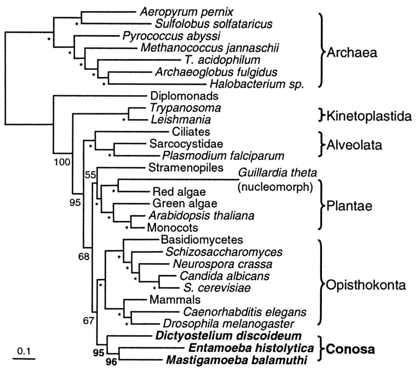
V-type ATPase (Eukaryotes) 215m00079 204m00094 91m00187 20m00291

A	B	C	D	E	F	G	3.6.3.14
I	AC39	54kD	S1	lipid			

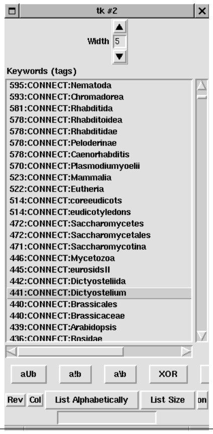
00193 7/7/03

The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*

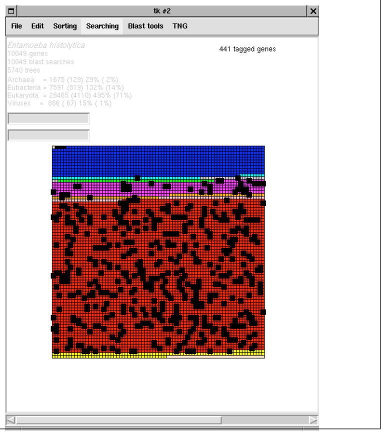
Eric Bapteste*, Henner Brinkmann†, Jennifer A. Lee‡, Dorothy V. Moore‡, Christoph W. Sensen§, Paul Gordon¶, Laure Duruflé*, Terry Gaasterland‡, Philippe Lopez*, Miklós Müller‡, and Hervé Philippe*ǀ



(Entamoeba + Dictyostelium) connections

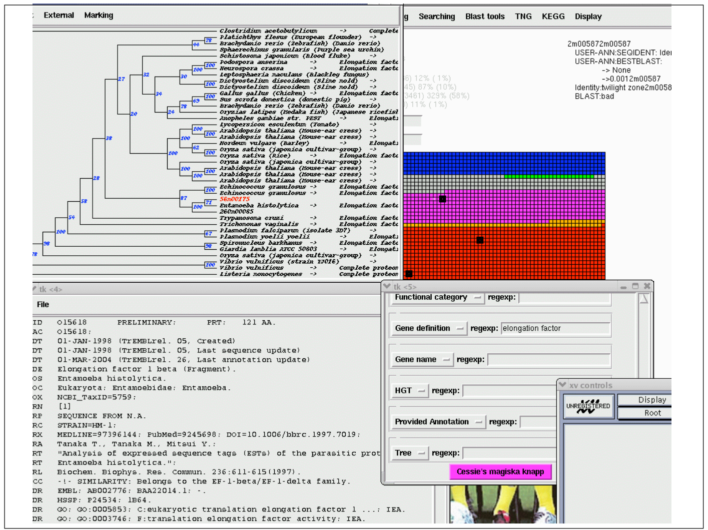


595:CONNECT:Nematoda
593:CONNECT:Chromadorea
581:CONNECT:Rhabditida
578:CONNECT:Rhabditoidea
578:CONNECT:Rhabditidae
578:CONNECT:Peloderinae
578:CONNECT:Caenorhabditis
570:CONNECT:Plasmodiumyoelii
523:CONNECT:Mammalia
522:CONNECT:Eutheria
514:CONNECT:coreeudicots
514:CONNECT:eudicotyledons
472:CONNECT:Saccharomycetes
472:CONNECT:Saccharomycetales
471:CONNECT:Saccharomycotina
446:CONNECT:Mycetozoa
445:CONNECT:eurosidsII
442:CONNECT:Dictyosteliida
441:CONNECT:Dictyostelium
440:CONNECT:Brassicales
440:CONNECT:Brassicaceae
439:CONNECT:Arabidopsis
436:CONNECT:Rosidae



Entamoeba histolytica
10049 genes
10049 blast searches
5740 trees
Archaea = 1675 (129) 29% (2%)
Eubacteria = 7591 (819) 132% (14%)
Eukaryota = 28465 (4110) 495% (71%)
Viruses = 886 (67) 15% (1%)

441 tagged genes


External
Marking
Searching
Blast tools
TNG
KEGG
Display
Clostridium acetobutylicum -> Complete
Platichthys flesus (European flounder) ->
Brachydanio rerio (Zebrafish) (Danio rerio)
Sphaerechinus granularis (Purple sea urchin)
Schistosoma japonicum (Blood fluke) ->
Podospora anserina -> Elongation facto
Neurospora crassa -> Elongation facto
Leptosphaeria maculans (Blackleg fungus)
Dictyostelium discoideum (Slime mold) ->
Dictyostelium discoideum (Slime mold) ->
Gallus gallus (Chicken) -> Elongation facto
Sus scrofa domestica (domestic pig) ->
Brachydanio rerio (Zebrafish) (Danio rerio)
Oryzias latipes (Medaka fish) (Japanese ricefish
Anopheles gambiae str. PEST -> Elongati
Lycopersicon esculentum (Tomato) ->
Arabidopsis thaliana (Mouse-ear cress) ->
Arabidopsis thaliana (Mouse-ear cress) ->
Hordeum vulgare (Barley) -> Elongati
Oryza sativa (japonica cultivar-group) ->
Oryza sativa (Rice) -> Elongation facto
Oryza sativa (japonica cultivar-group) ->
Arabidopsis thaliana (Mouse-ear cress) ->
Arabidopsis thaliana (Mouse-ear cress) ->
Arabidopsis thaliana (Mouse-ear cress) ->
Echinococcus granulosus -> Elongation facto
Echinococcus granulosus -> Elongation facto
56m00175
Entamoeba histolytica -> Elongation facto
260m00085
Trypanosoma cruzi -> Elongation facto
Trichomonas vaginalis -> Elongation facto
Plasmodium falciparum (isolate 3D7) ->
Plasmodium yoelii yoelii -> Elongati
Spironucleus barkhanus -> Elongation facto
Giardia lamblia ATCC 50803 -> Elongati
Oryza sativa (japonica cultivar-group) ->
Vibrio vulnificus (strain YJ016) ->
Vibrio vulnificus -> Complete proteom
Listeria monocytogenes -> Complete proteom
78
44
100
100
30
100
32
27
100
69
78
24
34
20
100
100
100
62
100
100
100
100
38
28
100
87
71
100
58
54
100
67
98
98
78
100
tk <4>
File
tk <5>
regexp:
xv controls
UNREGISTERED
Display
Root

2m005872m00587
USER-ANN:SEQIDENT: Ide
USER-ANN:BESTBLAST:
-> None
->0.0012m00587
Identity:twilight zone2m0058
BLAST:bad

6) 12% (1%)
45) 87% (10%)
3461) 329% (58%)
) 11% (1%)

```
ID   O15618      PRELIMINARY:     PRT:   121 AA.
AC   O15618;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Elongation factor 1 beta (Fragment).
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=5759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM-1;
RX   MEDLINE=97396144; PubMed=9245698; DOI=10.1006/bbrc.1997.7019;
RA   Tanaka T., Tanaka M., Mitsui Y.;
RT   "Analysis of expressed sequence tags (ESTs) of the parasitic prot
RT   Entamoeba histolytica.";
RL   Biochem. Biophys. Res. Commun. 236:611-615(1997).
CC   -!- SIMILARITY: Belongs to the EF-1-beta/EF-1-delta family.
DR   EMBL; AB002776; BAA22014.1; -.
DR   HSSP; P24534; 1B64.
DR   GO; GO:0005853; C:eukaryotic translation elongation factor 1 ...; IEA.
DR   GO; GO:0003746; F:translation elongation factor activity; IEA.
```

Functional category regexp:
Gene definition regexp: elongation factor
Gene name regexp:
HGT regexp:
Provided Annotation regexp:
Tree regexp:

Cessie's magiska knapp

PyPhy - the secondary screen I

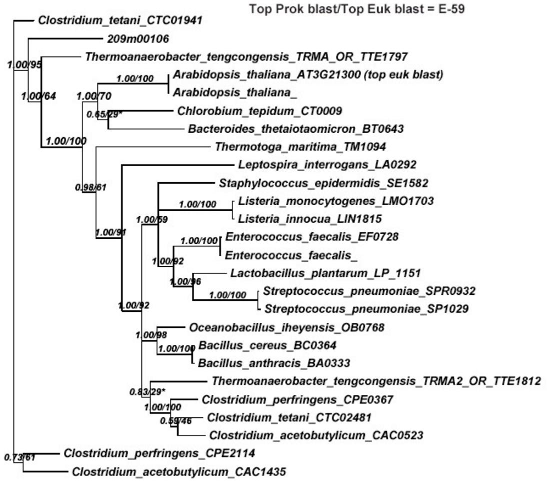
- Search EST databases and hits added to the alignments
- Feed G-blocked alignment into MRBAYES (Huelsenbeck & Ronquist 2001)
  - WAG matrix, gamma correction for site rate variation and a proportion (pinvar) of invariant sites.
- Make Bayesian consensus trees with posterior probabilities as support values

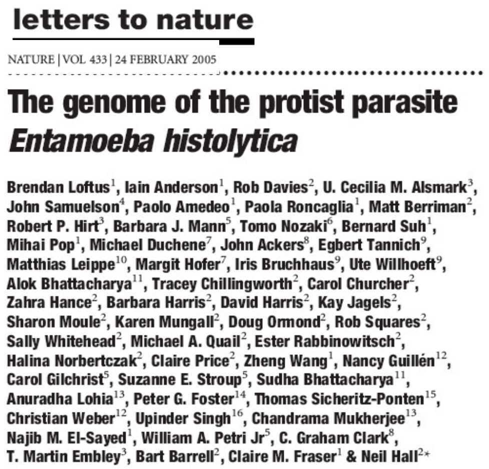
PyPhy - the secondary screen II

- Bootstrapping to provide an additional indication of support for relationships
- Each data set is bootstrapped (100 replicates) and used to make distance matrices under the same evolutionary model as in the Bayesian analysis, using custom (P4) software
- Trees are made from the distance matrices using FastME (Desper & Gascuel 2004) and a bootstrap consensus tree made using P4
- Transfer BS support values onto the Bayesian consensus tree

tRNA (Uracil-5-) - methyltransferase

Top Prok blast/Top Euk blast = E-59


Clostridium_tetani_CTC01941
209m00106
Thermoanaerobacter_tengcongensis_TRMA_OR_TTE1797
1.00/95
1.00/100
Arabidopsis_thaliana_AT3G21300 (top euk blast)
Arabidopsis_thaliana_
1.00/64
1.00/70
0.65/29*
Chlorobium_tepidum_CT0009
Bacteroides_thetaiotaomicron_BT0643
1.00/100
Thermotoga_maritima_TM1094
Leptospira_interrogans_LA0292
Staphylococcus_epidermidis_SE1582
0.98/61
1.00/100
Listeria_monocytogenes_LMO1703
1.00/59
Listeria_innocua_LIN1815
1.00/91
1.00/100
Enterococcus_faecalis_EF0728
Enterococcus_faecalis_
1.00/92
Lactobacillus_plantarum_LP_1151
1.00/96
1.00/100
Streptococcus_pneumoniae_SPR0932
Streptococcus_pneumoniae_SP1029
1.00/92
Oceanobacillus_iheyensis_OB0768
1.00/98
1.00/100
Bacillus_cereus_BC0364
Bacillus_anthracis_BA0333
Thermoanaerobacter_tengcongensis_TRMA2_OR_TTE1812
0.83/29*
Clostridium_perfringens_CPE0367
1.00/100
0.59/46
Clostridium_tetani_CTC02481
Clostridium_acetobutylicum_CAC0523
0.73/61
Clostridium_perfringens_CPE2114
Clostridium_acetobutylicum_CAC1435



letters to nature

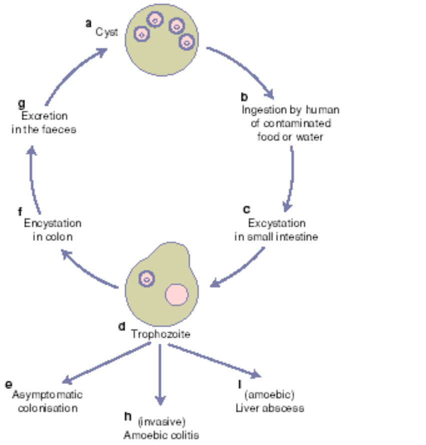
NATURE | VOL 433 | 24 FEBRUARY 2005

The genome of the protist parasite *Entamoeba histolytica*

Brendan Loftus[1], Iain Anderson[1], Rob Davies[2], U. Cecilia M. Alsmark[3], John Samuelson[4], Paolo Amedeo[1], Paola Roncaglia[1], Matt Berriman[2], Robert P. Hirt[3], Barbara J. Mann[5], Tomo Nozaki[6], Bernard Suh[1], Mihai Pop[1], Michael Duchene[7], John Ackers[8], Egbert Tannich[9], Matthias Leippe[10], Margit Hofer[7], Iris Bruchhaus[9], Ute Willhoeft[9], Alok Bhattacharya[11], Tracey Chillingworth[2], Carol Churcher[2], Zahra Hance[2], Barbara Harris[2], David Harris[2], Kay Jagels[2], Sharon Moule[2], Karen Mungall[2], Doug Ormond[2], Rob Squares[2], Sally Whitehead[2], Michael A. Quail[2], Ester Rabbinowitsch[2], Halina Norbertczak[2], Claire Price[2], Zheng Wang[1], Nancy Guillén[12], Carol Gilchrist[5], Suzanne E. Stroup[5], Sudha Bhattacharya[11], Anuradha Lohia[13], Peter G. Foster[14], Thomas Sicheritz-Ponten[15], Christian Weber[12], Upinder Singh[16], Chandrama Mukherjee[13], Najib M. El-Sayed[1], William A. Petri Jr[5], C. Graham Clark[8], T. Martin Embley[3], Bart Barrell[2], Claire M. Fraser[1] & Neil Hall[2]*

E. histolytica trophozoites in situ

(D. Mirelman 1996)


a Cyst
b Ingestion by human of contaminated food or water
c Excystation in small intestine
d Trophozoite
e Asymptomatic colonisation
f Encystation in colon
g Excretion in the faeces
h (invasive) Amoebic colitis
i (amoebic) Liver abscess

Life cycle of *Entamoeba histolytica* and the clinical manifestations of infection in humans

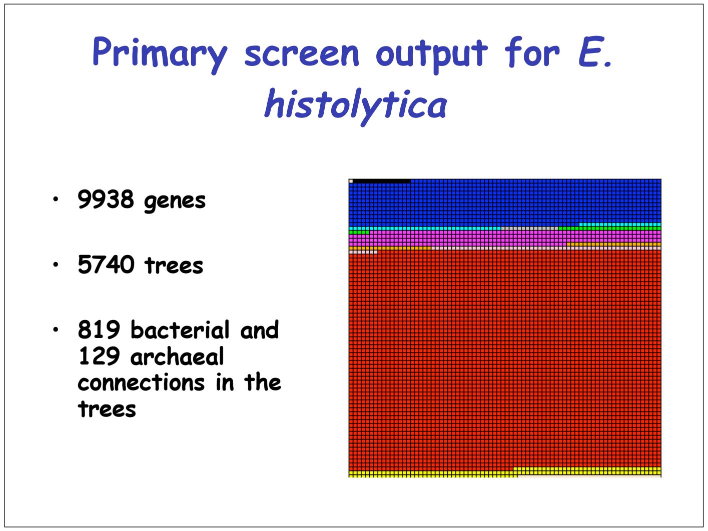
Expert Reviews in Molecular Medicine ©1999 Cambridge University Press

The *E. histolytica* genome

- Sequenced by TIGR and the Sanger institute (PIs Brendan Loftus and Neil Hall)
- ~ 23.7 Mbp
- 9938 predicted genes
- 1/3rd of genes have no detectable homologue in public data bases
- Large gene families, duplications are common

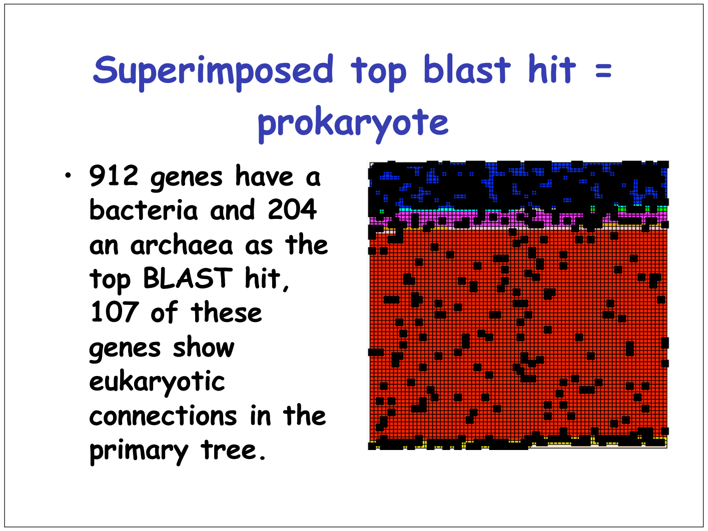
Aim of our HGT screen

- To identify the strongest candidate 'recent' HGT from prokaryotes to *Entamoeba*
  - The tip of the HGT iceberg?
  - Help to validate HGT as a plausible explanation for topological incongruence deeper in tree
- These cases shoudl be the easiest to defend - paralogy or poor tree-building is unlikely to be an equally plausible explanation to HGT
- This screen is not seeking to detect more 'ancient' transfers, for example - at the base of major groups or from the alpha-proteobacterial mitochondrial endosymbiont



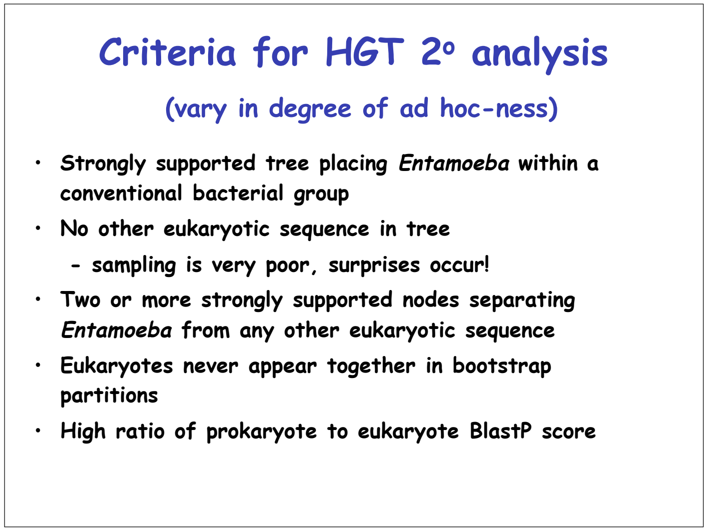
Primary screen output for *E. histolytica*

- 9938 genes
- 5740 trees
- 819 bacterial and 129 archaeal connections in the trees



Superimposed top blast hit = prokaryote

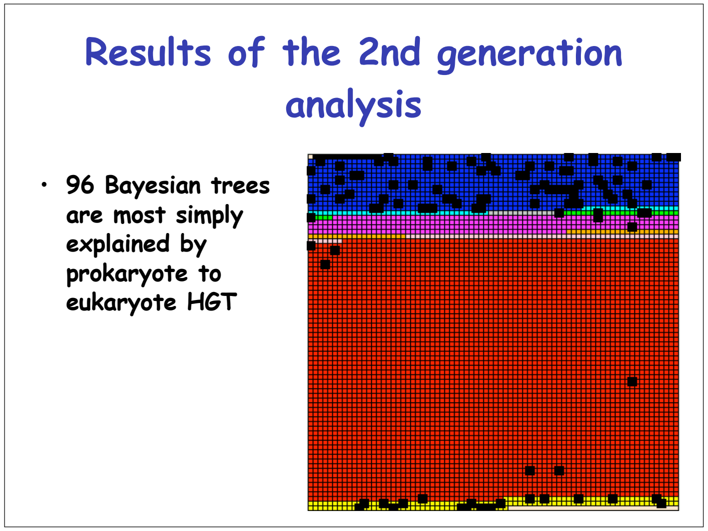
- 912 genes have a bacteria and 204 an archaea as the top BLAST hit, 107 of these genes show eukaryotic connections in the primary tree.



Criteria for HGT 2^o analysis

(vary in degree of ad hoc-ness)

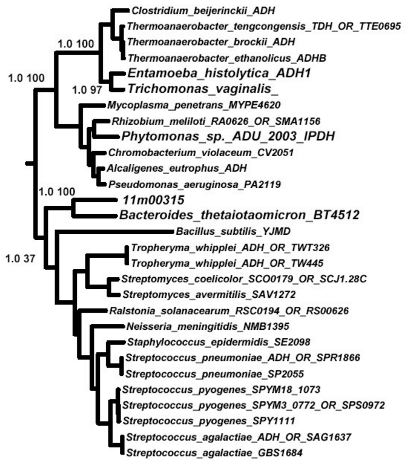
- Strongly supported tree placing *Entamoeba* within a conventional bacterial group
- No other eukaryotic sequence in tree
  - sampling is very poor, surprises occur!
- Two or more strongly supported nodes separating *Entamoeba* from any other eukaryotic sequence
- Eukaryotes never appear together in bootstrap partitions
- High ratio of prokaryote to eukaryote BlastP score



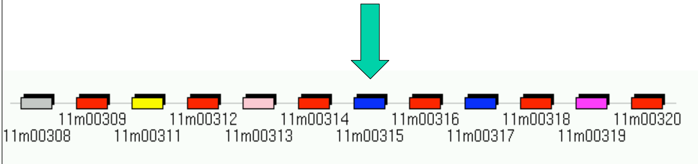
Results of the 2nd generation analysis

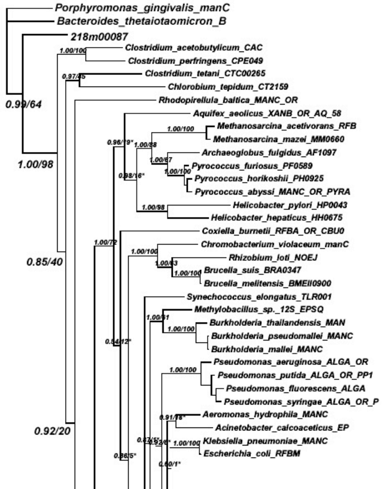
- 96 Bayesian trees are most simply explained by prokaryote to eukaryote HGT

Zinc alcohol dehydrogenase (subtree)


Clostridium_beijerinckii_ADH
Thermoanaerobacter_tengcongensis_TDH_OR_TTE0695
1.0 100
Thermoanaerobacter_brockii_ADH
1.0 100
Thermoanaerobacter_ethanolicus_ADHB
Entamoeba_histolytica_ADH1
1.0 97
Trichomonas_vaginalis_
Mycoplasma_penetrans_MYPE4620
Rhizobium_meliloti_RA0626_OR_SMA1156
Phytomonas_sp._ADU_2003_IPDH
Chromobacterium_violaceum_CV2051
Alcaligenes_eutrophus_ADH
Pseudomonas_aeruginosa_PA2119
1.0 100
11m00315
Bacteroides_thetaiotaomicron_BT4512
Bacillus_subtilis_YJMD
Tropheryma_whipplei_ADH_OR_TWT326
1.0 37
Tropheryma_whipplei_ADH_OR_TW445
Streptomyces_coelicolor_SCO0179_OR_SCJ1.28C
Streptomyces_avermitilis_SAV1272
Ralstonia_solanacearum_RSC0194_OR_RS00626
Neisseria_meningitidis_NMB1395
Staphylococcus_epidermidis_SE2098
Streptococcus_pneumoniae_ADH_OR_SPR1866
Streptococcus_pneumoniae_SP2055
Streptococcus_pyogenes_SPYM18_1073
Streptococcus_pyogenes_SPYM3_0772_OR_SPS0972
Streptococcus_pyogenes_SPY1111
Streptococcus_agalactiae_ADH_OR_SAG1637
Streptococcus_agalactiae_GBS1684

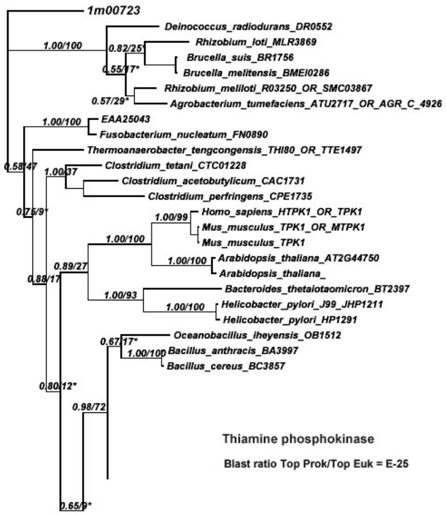
Mapping HGT genes onto scaffolds




Porphyromonas_gingivalis_manC
Bacteroides_thetaiotaomicron_B
218m00087
1.00/100
Clostridium_acetobutylicum_CAC
Clostridium_perfringens_CPE049
0.97/85
Clostridium_tetani_CTC00265
Chlorobium_tepidum_CT2159
0.99/64
Rhodopirellula_baltica_MANC_OR
Aquifex_aeolicus_XANB_OR_AQ_58
1.00/100
Methanosarcina_acetivorans_RFB
0.96/19*
1.00/88
Methanosarcina_mazei_MM0660
Archaeoglobus_fulgidus_AF1097
1.00/67
Pyrococcus_furiosus_PF0589
1.00/100
0.98/16*
Pyrococcus_horikoshii_PH0925
Pyrococcus_abyssi_MANC_OR_PYRA
1.00/98
1.00/98
Helicobacter_pylori_HP0043
Helicobacter_hepaticus_HH0675
Coxiella_burnetii_RFBA_OR_CBU0
1.00/72
Chromobacterium_violaceum_manC
0.85/40
1.00/100
Rhizobium_loti_NOEJ
1.00/63
1.00/100
Brucella_suis_BRA0347
Brucella_melitensis_BMEII0900
Synechococcus_elongatus_TLR001
Methylobacillus_sp._12S_EPSQ
1.00/61
Burkholderia_thailandensis_MAN
1.00/100
0.54/12*
Burkholderia_pseudomallei_MANC
Burkholderia_mallei_MANC
Pseudomonas_aeruginosa_ALGA_OR
1.00/100
Pseudomonas_putida_ALGA_OR_PP1
Pseudomonas_fluorescens_ALGA
Pseudomonas_syringae_ALGA_OR_P
0.92/20
0.91/18*
Aeromonas_hydrophila_MANC
Acinetobacter_calcoaceticus_EP
0.87/15*
0.92/6*
1.00/100
Klebsiella_pneumoniae_MANC
0.86/5*
Escherichia_coli_RFBM
0.60/1*

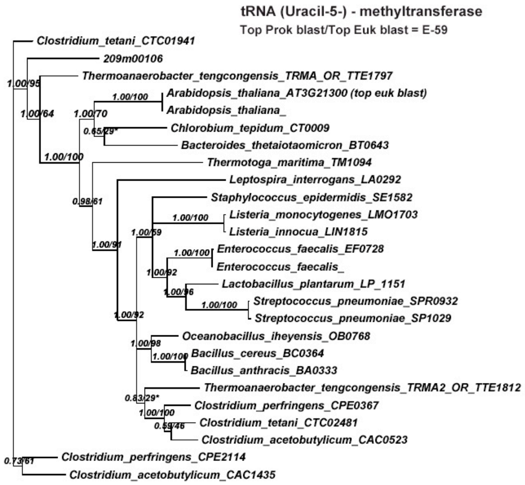
Mannose phosphate guanyltransferase

Top Prok Blast/Top Euk = E-88


1m00723
Deinococcus_radiodurans_DR0552
1.00/100
0.82/25*
Rhizobium_loti_MLR3869
0.55/17*
Brucella_suis_BR1756
Brucella_melitensis_BMEI0286
Rhizobium_meliloti_R03250_OR_SMC03867
0.57/29*
Agrobacterium_tumefaciens_ATU2717_OR_AGR_C_4926
1.00/100
EAA25043
Fusobacterium_nucleatum_FN0890
Thermoanaerobacter_tengcongensis_THI80_OR_TTE1497
0.58/47
Clostridium_tetani_CTC01228
1.00/37
Clostridium_acetobutylicum_CAC1731
Clostridium_perfringens_CPE1735
0.75/9*
1.00/99
Homo_sapiens_HTPK1_OR_TPK1
Mus_musculus_TPK1_OR_MTPK1
Mus_musculus_TPK1
1.00/100
1.00/100
Arabidopsis_thaliana_AT2G44750
0.89/27
Arabidopsis_thaliana_
0.88/17
Bacteroides_thetaiotaomicron_BT2397
1.00/93
1.00/100
Helicobacter_pylori_J99_JHP1211
Helicobacter_pylori_HP1291
0.67/17*
Oceanobacillus_iheyensis_OB1512
1.00/100
Bacillus_anthracis_BA3997
Bacillus_cereus_BC3857
0.80/12*
0.98/72
0.65/9*

Thiamine phosphokinase

Blast ratio Top Prok/Top Euk = E-25



tRNA (Uracil-5-) - methyltransferase

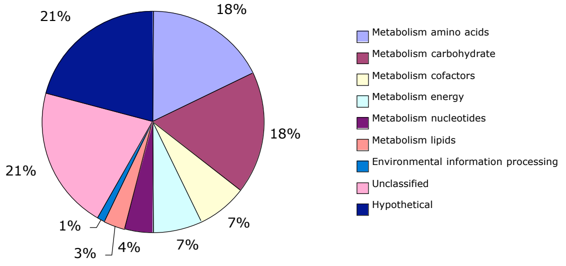
Top Prok blast/Top Euk blast = E-59

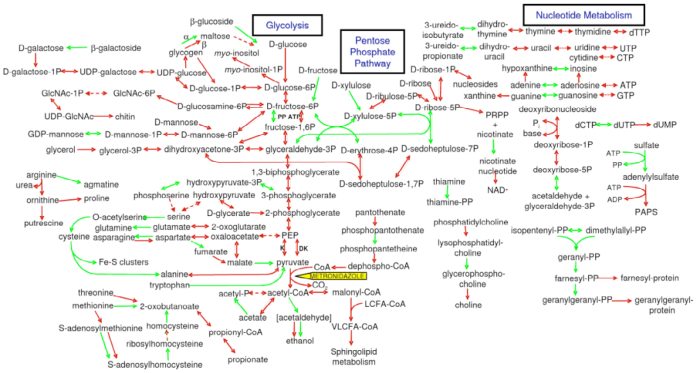
Clostridium_tetani_CTC01941
209m00106
Thermoanaerobacter_tengcongensis_TRMA_OR_TTE1797
1.00/95
1.00/100
Arabidopsis_thaliana_AT3G21300 (top euk blast)
Arabidopsis_thaliana_
1.00/64
1.00/70
0.65/29*
Chlorobium_tepidum_CT0009
Bacteroides_thetaiotaomicron_BT0643
1.00/100
Thermotoga_maritima_TM1094
Leptospira_interrogans_LA0292
0.98/61
Staphylococcus_epidermidis_SE1582
1.00/100
Listeria_monocytogenes_LMO1703
1.00/59
Listeria_innocua_LIN1815
1.00/91
1.00/100
Enterococcus_faecalis_EF0728
1.00/92
Enterococcus_faecalis_
Lactobacillus_plantarum_LP_1151
1.00/96
1.00/100
Streptococcus_pneumoniae_SPR0932
1.00/92
Streptococcus_pneumoniae_SP1029
1.00/98
Oceanobacillus_iheyensis_OB0768
1.00/100
Bacillus_cereus_BC0364
Bacillus_anthracis_BA0333
Thermoanaerobacter_tengcongensis_TRMA2_OR_TTE1812
0.83/29*
Clostridium_perfringens_CPE0367
1.00/100
Clostridium_tetani_CTC02481
0.59/46
Clostridium_acetobutylicum_CAC0523
0.73/61
Clostridium_perfringens_CPE2114
Clostridium_acetobutylicum_CAC1435

Transferred genes are involved in metabolism

121m00124	Tartrate dehydrogenase	METABOLISM
133m00129	Fructokinase	METABOLISM
400m00033	Arginine decarboxylase	METABOLISM
78m00151	Lysophospholipase L2	METABOLISM
9m00390	HesA/MoeB/ThiF family protein	METABOLISM
24m00307	NADH dehydrogenase, similar to nitrite reductase	ENERGY METABOLISM
61m00186	5-nitroimidazole antibiotic resistance protein	CELL RESCUE, DEFENCE AND VIRULENCE
30m00272	ABC transporter	TRANSPORT
289m00068	Conserved hypothetical protein	UNKNOWN FUNCTION

Functional distribution of HGT



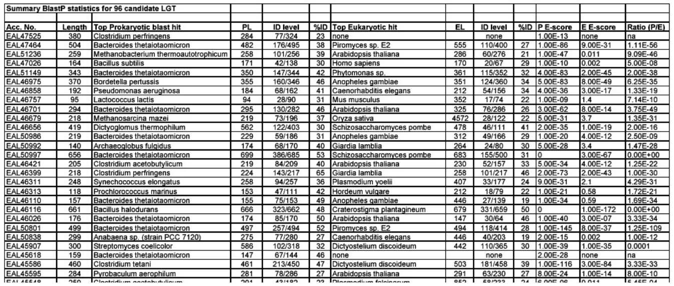


We see a broad collection of donors

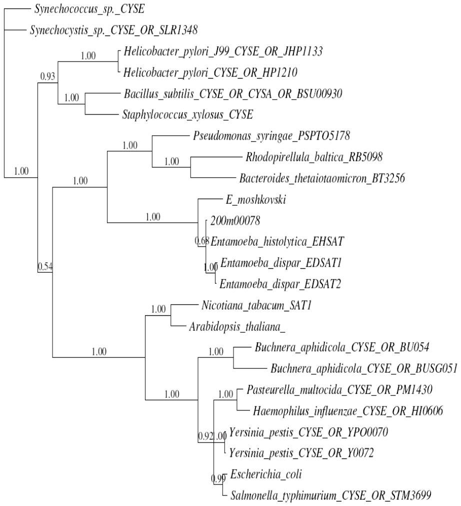
Name	NN MrBayes	Best Prokaryote blast hit
100m00122	DT Deinococcus	Clostridium perfringens
189m00102	Archeal like	Pyrobaculum aerophilum
10m00331	Proteo /Xanthobact	Methanobacterium thermoautotrophicum
119m00142	Proteo /Vibrio	Bacillus subtilis
11m00315	Bact/Chlorob	Bacteroides thetaiotaomicron
209m00106	Firmicutes	Clostridium perfringen
126m00107	Bacteria like	Pseudomonas aeruginosa
130m00124	Firmicutes	Lactococcus lactis (subsp. lactis) (Streptococcus lactis)
133m00129	Bacteria like	Bacteroides thetaiotaomicron
133m00136	Bact/Chlorob	Methanosarcina mazei (Methanosarcina frisia)

Summary BlastP Statistics

(provided as supplemental info to Loftus et al 2005)

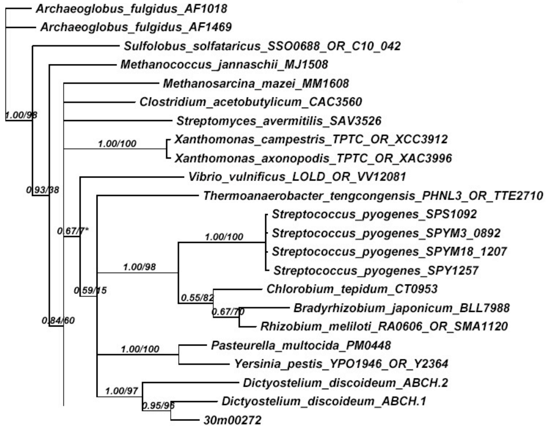


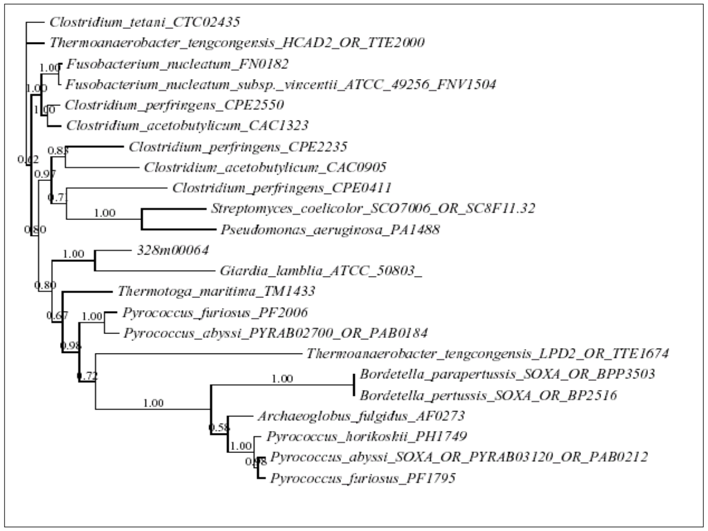
Summary BlastP statistics for 96 candidate LGT												
Acc. No.	Length	Top Prokaryotic blast hit	PL	ID level	%ID	Top Eukaryotic hit	EL	ID level	%ID	P E-score	E E-score	Ratio (P/E)
EAL47525	380	Clostridium perfringens	284	77/324	23	none		none		1.00E-13	none	na
EAL47464	504	Bacteroides thetaiotaomicron	482	176/495	38	Piromyces sp. E2	555	110/400	27	1.00E-86	9.00E-31	1.11E-56
EAL51236	259	Methanobacterium thermoautotrophicum	258	101/256	39	Arabidopsis thaliana	286	60/276	21	1.00E-47	0.011	9.09E-46
EAL47026	164	Bacillus subtilis	171	42/138	30	Homo sapiens	170	20/67	29	1.00E-10	0.002	5.00E-08
EAL51149	343	Bacteroides thetaiotaomicron	350	147/344	42	Phytomonas sp.	361	115/352	32	4.00E-83	2.00E-45	2.00E-38
EAL46975	370	Bordetella pertussis	355	160/346	46	Anopheles gambiae	351	124/360	34	5.00E-83	8.00E-49	6.25E-35
EAL46858	192	Pseudomonas aeruginosa	184	68/162	41	Caenorhabditis elegans	212	54/156	34	4.00E-36	3.00E-17	1.33E-19
EAL46757	95	Lactococcus lactis	94	28/90	31	Mus musculus	352	17/74	22	1.00E-09	1.4	7.14E-10
EAL46701	294	Bacteroides thetaiotaomicron	295	130/282	46	Arabidopsis thaliana	325	76/286	26	3.00E-62	8.00E-14	3.75E-49
EAL46679	218	Methanosarcina mazei	219	73/196	37	Oryza sativa	4572	28/122	22	5.00E-31	3.7	1.35E-31
EAL46656	419	Dictyoglomus thermophilum	562	122/403	30	Schizosaccharomyces pombe	478	46/111	41	2.00E-35	1.00E-19	2.00E-16
EAL50986	219	Bacteroides thetaiotaomicron	229	59/186	31	Anopheles gambiae	312	49/166	29	1.00E-20	4.00E-12	2.50E-09
EAL50992	140	Archaeoglobus fulgidus	174	68/170	40	Giardia lamblia	264	24/80	30	5.00E-28	3.4	1.47E-28
EAL50997	656	Bacteroides thetaiotaomicron	699	386/685	53	Schizosaccharomyces pombe	683	155/500	31	0	3.00E-67	0.00E+00
EAL46421	205	Clostridium acetobutylicum	219	84/209	40	Arabidopsis thaliana	230	52/157	33	5.00E-34	4.00E-12	1.25E-22
EAL46399	218	Clostridium perfringens	224	143/217	65	Giardia lamblia	258	101/217	46	2.00E-73	2.00E-43	1.00E-30
EAL46311	248	Synechococcus elongatus	258	94/257	36	Plasmodium yoelii	407	33/177	24	9.00E-31	2.1	4.29E-31
EAL46313	118	Prochlorococcus marinus	153	47/111	42	Hordeum vulgare	212	18/79	22	1.00E-21	0.58	1.72E-21
EAL46110	157	Bacteroides thetaiotaomicron	155	75/153	49	Anopheles gambiae	446	27/139	19	1.00E-34	0.59	1.69E-34
EAL46116	661	Bacillus halodurans	666	323/662	48	Craterostigma plantagineum	679	331/659	50	0	1.00E-172	0.00E+00
EAL46026	176	Bacteroides thetaiotaomicron	174	85/170	50	Arabidopsis thaliana	147	30/64	46	1.00E-40	3.00E-07	3.33E-34
EAL50801	499	Bacteroides thetaiotaomicron	497	257/494	52	Piromyces sp. E2	494	118/414	28	1.00E-145	8.00E-37	1.25E-109
EAL50838	299	Anabaena sp. (strain PCC 7120)	275	77/280	27	Caenorhabditis elegans	446	40/203	19	2.00E-15	0.002	1.00E-12
EAL45907	300	Streptomyces coelicolor	586	102/318	32	Dictyostelium discoideum	442	110/365	30	1.00E-39	1.00E-35	0.0001
EAL45618	159	Bacteroides thetaiotaomicron	147	67/144	46	none		none		2.00E-28	none	na
EAL45586	460	Clostridium tetani	461	213/450	47	Dictyostelium discoideum	503	181/458	39	1.00E-116	3.00E-84	3.33E-33
EAL45595	284	Pyrobaculum aerophilum	281	78/286	27	Arabidopsis thaliana	291	63/230	27	8.00E-24	1.00E-14	8.00E-10
EAL45548	250	Clostridium acetobutylicum	201	43/182	23	Plasmodium falciparum	852	58/233	24	6.00E-06	0.011	5.45E-04

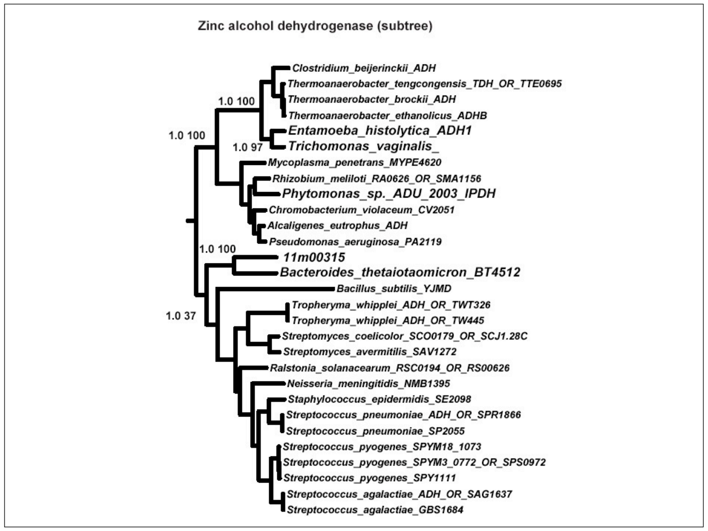

Synechococcus_sp._CYSE
Synechocystis_sp._CYSE_OR_SLR1348
1.00
Helicobacter_pylori_J99_CYSE_OR_JHP1133
0.93
Helicobacter_pylori_CYSE_OR_HP1210
1.00
Bacillus_subtilis_CYSE_OR_CYSA_OR_BSU00930
Staphylococcus_xylosus_CYSE
1.00
Pseudomonas_syringae_PSPTO5178
1.00
Rhodopirellula_baltica_RB5098
1.00
1.00
Bacteroides_thetaiotaomicron_BT3256
E_moshkovski
1.00
200m00078
0.68
Entamoeba_histolytica_EHSAT
0.54
1.00
Entamoeba_dispar_EDSAT1
Entamoeba_dispar_EDSAT2
1.00
Nicotiana_tabacum_SAT1
Arabidopsis_thaliana_
1.00
1.00
Buchnera_aphidicola_CYSE_OR_BU054
Buchnera_aphidicola_CYSE_OR_BUSG051
1.00
1.00
Pasteurella_multocida_CYSE_OR_PM1430
Haemophilus_influenzae_CYSE_OR_HI0606
0.92
1.00
Yersinia_pestis_CYSE_OR_YPO0070
Yersinia_pestis_CYSE_OR_Y0072
0.99
Escherichia_coli
Salmonella_typhimurium_CYSE_OR_STM3699

HGT at the base of the amoebozoa?

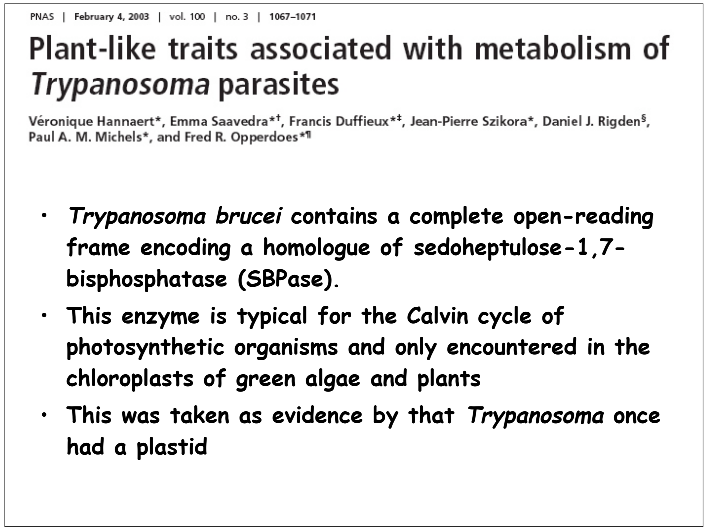
ABC transporter


Archaeoglobus_fulgidus_AF1018
Archaeoglobus_fulgidus_AF1469
Sulfolobus_solfataricus_SSO0688_OR_C10_042
Methanococcus_jannaschii_MJ1508
Methanosarcina_mazei_MM1608
Clostridium_acetobutylicum_CAC3560
1.00/98
Streptomyces_avermitilis_SAV3526
1.00/100
Xanthomonas_campestris_TPTC_OR_XCC3912
Xanthomonas_axonopodis_TPTC_OR_XAC3996
Vibrio_vulnificus_LOLD_OR_VV12081
0.93/38
Thermoanaerobacter_tengcongensis_PHNL3_OR_TTE2710
Streptococcus_pyogenes_SPS1092
0.67/7*
1.00/100
Streptococcus_pyogenes_SPYM3_0892
Streptococcus_pyogenes_SPYM18_1207
1.00/98
Streptococcus_pyogenes_SPY1257
0.59/15
Chlorobium_tepidum_CT0953
0.55/82
Bradyrhizobium_japonicum_BLL7988
0.67/70
0.84/60
Rhizobium_meliloti_RA0606_OR_SMA1120
1.00/100
Pasteurella_multocida_PM0448
Yersinia_pestis_YPO1946_OR_Y2364
Dictyostelium_discoideum_ABCH.2
1.00/97
Dictyostelium_discoideum_ABCH.1
0.95/95
30m00272


Clostridium_tetani_CTC02435
Thermoanaerobacter_tengcongensis_HCAD2_OR_TTE2000
1.00
Fusobacterium_nucleatum_FN0182
1.00
Fusobacterium_nucleatum_subsp._vincentii_ATCC_49256_FNV1504
1.00
Clostridium_perfringens_CPE2550
Clostridium_acetobutylicum_CAC1323
0.85
Clostridium_perfringens_CPE2235
0.62
Clostridium_acetobutylicum_CAC0905
0.97
Clostridium_perfringens_CPE0411
0.71
1.00
Streptomyces_coelicolor_SCO7006_OR_SC8F11.32
0.80
Pseudomonas_aeruginosa_PA1488
1.00
328m00064
Giardia_lamblia_ATCC_50803_
0.80
Thermotoga_maritima_TM1433
0.67
1.00
Pyrococcus_furiosus_PF2006
Pyrococcus_abyssi_PYRAB02700_OR_PAB0184
0.98
Thermoanaerobacter_tengcongensis_LPD2_OR_TTE1674
0.72
1.00
Bordetella_parapertussis_SOXA_OR_BPP3503
1.00
Bordetella_pertussis_SOXA_OR_BP2516
Archaeoglobus_fulgidus_AF0273
0.58
Pyrococcus_horikoshii_PH1749
1.00
Pyrococcus_abyssi_SOXA_OR_PYRAB03120_OR_PAB0212
0.98
Pyrococcus_furiosus_PF1795


Zinc alcohol dehydrogenase (subtree)

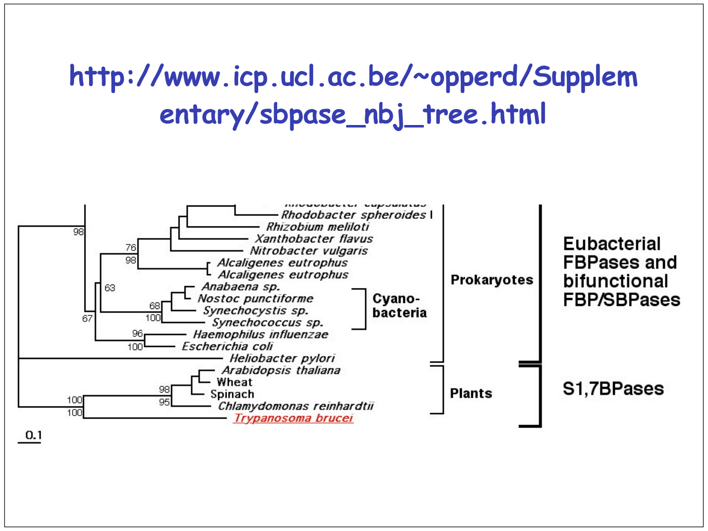
Clostridium_beijerinckii_ADH
Thermoanaerobacter_tengcongensis_TDH_OR_TTE0695
1.0 100
Thermoanaerobacter_brockii_ADH
Thermoanaerobacter_ethanolicus_ADHB
1.0 100
Entamoeba_histolytica_ADH1
1.0 97
Trichomonas_vaginalis_
Mycoplasma_penetrans_MYPE4620
Rhizobium_meliloti_RA0626_OR_SMA1156
Phytomonas_sp._ADU_2003_IPDH
Chromobacterium_violaceum_CV2051
Alcaligenes_eutrophus_ADH
Pseudomonas_aeruginosa_PA2119
1.0 100
11m00315
Bacteroides_thetaiotaomicron_BT4512
Bacillus_subtilis_YJMD
Tropheryma_whipplei_ADH_OR_TWT326
1.0 37
Tropheryma_whipplei_ADH_OR_TW445
Streptomyces_coelicolor_SCO0179_OR_SCJ1.28C
Streptomyces_avermitilis_SAV1272
Ralstonia_solanacearum_RSC0194_OR_RS00626
Neisseria_meningitidis_NMB1395
Staphylococcus_epidermidis_SE2098
Streptococcus_pneumoniae_ADH_OR_SPR1866
Streptococcus_pneumoniae_SP2055
Streptococcus_pyogenes_SPYM18_1073
Streptococcus_pyogenes_SPYM3_0772_OR_SPS0972
Streptococcus_pyogenes_SPY1111
Streptococcus_agalactiae_ADH_OR_SAG1637
Streptococcus_agalactiae_GBS1684

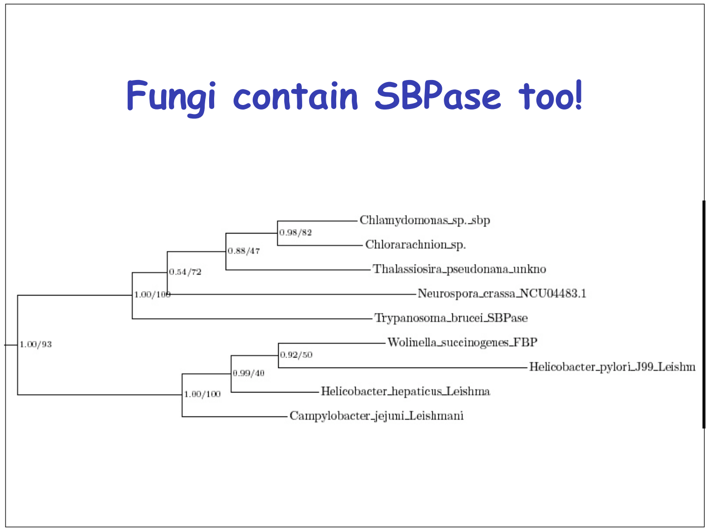

PNAS | February 4, 2003 | vol. 100 | no. 3 | 1067–1071

Plant-like traits associated with metabolism of *Trypanosoma* parasites

Véronique Hannaert*, Emma Saavedra*†, Francis Duffieux*‡, Jean-Pierre Szikora*, Daniel J. Rigden§, Paul A. M. Michels*, and Fred R. Opperdoes*¶

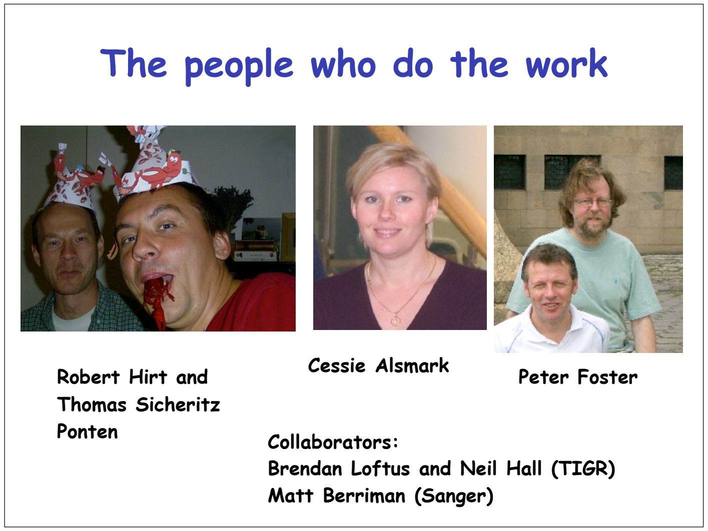
- *Trypanosoma brucei* contains a complete open-reading frame encoding a homologue of sedoheptulose-1,7-bisphosphatase (SBPase).
- This enzyme is typical for the Calvin cycle of photosynthetic organisms and only encountered in the chloroplasts of green algae and plants
- This was taken as evidence by that *Trypanosoma* once had a plastid


http://www.icp.ucl.ac.be/~opperd/Supplementary/sbpase_nbj_tree.html



Fungi contain SBPase too!

Chlamydomonas_sp._sbp
Chlorarachnion_sp.
Thalassiosira_pseudonana_unkno
Neurospora_crassa_NCU04483.1
Trypanosoma_brucei_SBPase
Wolinella_succinogenes_FBP
Helicobacter_pylori_J99_Leishm
Helicobacter_hepaticus_Leishma
Campylobacter_jejuni_Leishmani
0.98/82
0.88/47
0.54/72
1.00/100
1.00/93
0.92/50
0.99/40
1.00/100



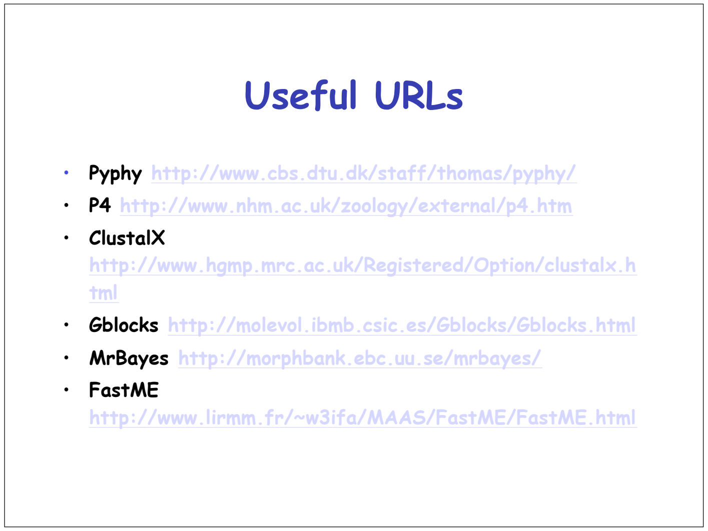
The people who do the work

Robert Hirt and
Thomas Sicheritz
Ponten

Cessie Alsmark

Peter Foster

Collaborators:
Brendan Loftus and Neil Hall (TIGR)
Matt Berriman (Sanger)



Useful URLs

- Pyphy http://www.cbs.dtu.dk/staff/thomas/pyphy/
- P4 http://www.nhm.ac.uk/zoology/external/p4.htm
- ClustalX http://www.hgmp.mrc.ac.uk/Registered/Option/clustalx.html
- Gblocks http://molevol.ibmb.csic.es/Gblocks/Gblocks.html
- MrBayes http://morphbank.ebc.uu.se/mrbayes/
- FastME http://www.lirmm.fr/~w3ifa/MAAS/FastME/FastME.html