Evaluation of New Methods & Approaches for Comparative Metagenomic Studies

Emmanuel F. Mongodin

J. Craig Venter™ INSTITUTE

mongodin@jcvi.org

Metagenomics

- Rapidly emerging field of research for studying microbial communities without the need for isolation and lab cultivation of individual species.

- Term "metagenomics" was first coined by Jo Handelsman (U. of Wisconsin): "*the application of modern genomics techniques to the study of complex communities of microbial organisms directly in their natural environments*".

- In complex environments where most species are not cultivable:
  - Key questions: "Who's there?" and "What are they doing?"
  - Metagenomics aims to capture the full measure of microbial, and shed light on the biological processes present in the environment studied.

- Problem:
  - Most of the tools & methods used to analyze metagenomic data sets are adapted from the ones developed for the analysis of single genomes.
  - The genetic & taxonomic complexity (genomes of multiple strains/species present in highly variable abundance) of metagenomic data sets imposes new challenges on existing analysis tools.

- New analysis tools and pipelines specifically dedicated to metagenomics sets is needed.

Genomics of the Human Colonic Microbiome

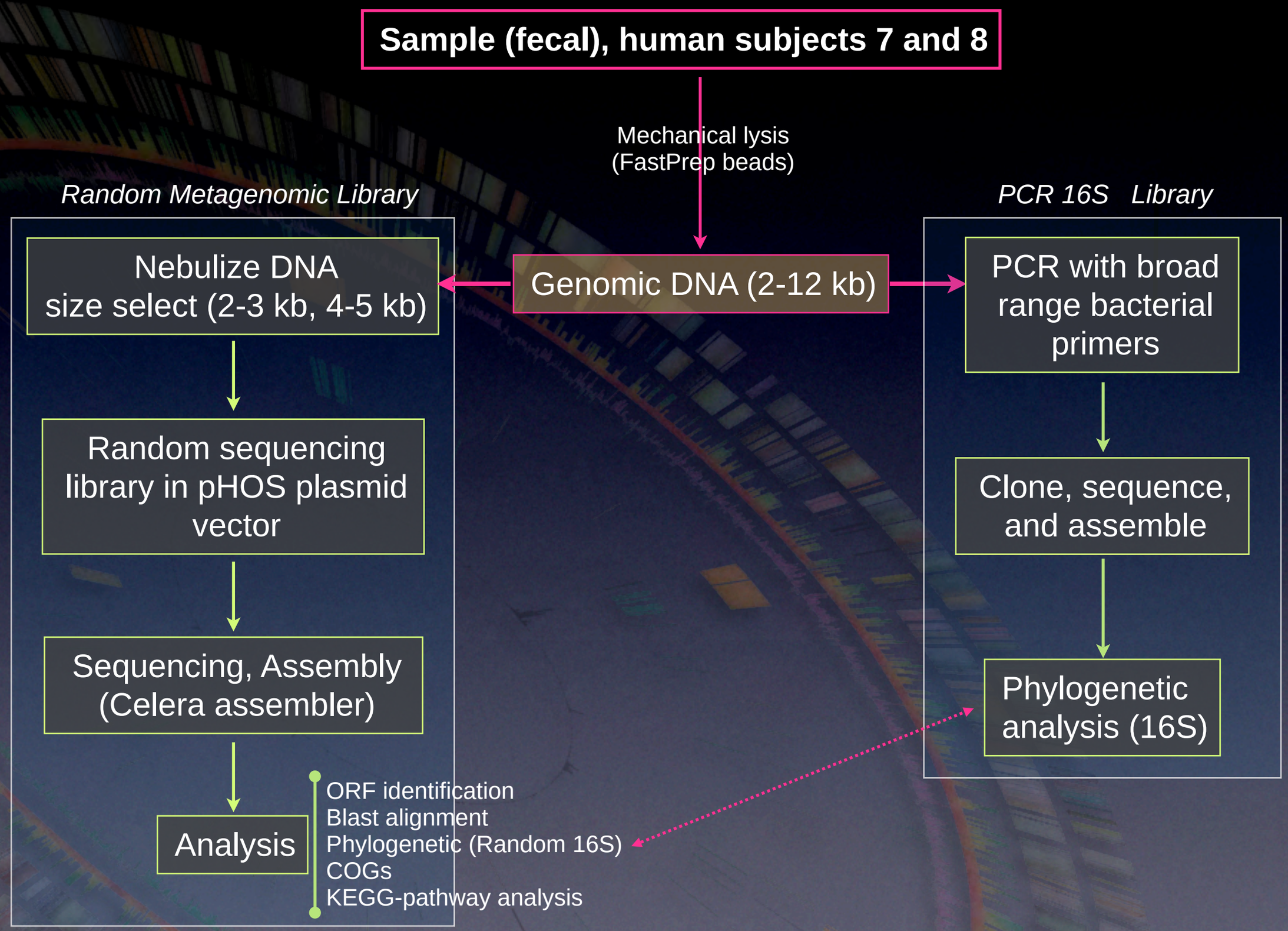
Metagenomic Analysis of the Human Distal Gut Microbiome

Steven R. Gill,[1*‡] Mihai Pop,[1†] Robert T. DeBoy,[1] Paul B. Eckburg,[2,3,4]
Peter J. Turnbaugh,[5] Buck S. Samuel,[5] Jeffrey I. Gordon,[5] David A. Relman,[2,3,4]
Claire M. Fraser-Liggett,[1,6] Karen E. Nelson[1]

The human intestinal microbiota is composed of 10^{13} to 10^{14} microorganisms whose collective genome ("microbiome") contains at least 100 times as many genes as our own genome. We analyzed ~78 million base pairs of unique DNA sequence and 2062 polymerase chain reaction–amplified 16S ribosomal DNA sequences obtained from the fecal DNAs of two healthy adults. Using metabolic function analyses of identified genes, we compared our human genome with the average content of previously sequenced microbial genomes. Our microbiome has significantly enriched metabolism of glycans, amino acids, and xenobiotics; methanogenesis; and 2-methyl-D-erythritol 4-phosphate pathway–mediated biosynthesis of vitamins and isoprenoids. Thus, humans are superorganisms whose metabolism represents an amalgamation of microbial and human attributes.

www.sciencemag.org SCIENCE VOL 312 2 JUNE 2006

Genomics of the Human Colonic Microbiome



J. Craig Venter INSTITUTE

Comparison between different pipelines for ORFs and functions predictions in metagenomic dataset

	This study			Gill et al. 2006		IMG/M pipeline [1]	
	#7	#8	#7&8	#7	#8	#7	#8
# of reads	65 059	74 462	139 521	139 521		65 059	74 462
# of contig	7 647	10 471	17 783	17 668		10 411	12 020
Mean/max/min length of contigs, bp	1767 /19 104 /131	1879 /25 060 /93	1856/45 326/92	1910 /44 747 /92		1521 /20 238 /67	1705 /25 061 /67
% reads in contigs	48%	63%		60%		?	?
# of singletons (% of all reads)	33 925 (52%)	27 658 (37%)	58 977 (42%)	56 292 (40%)		?	?
Average depth of coverage in contigs	1.92	2.04	2.07	2.13		?	?
# of ORFs	69 829	67 766	132 943	-	-	21 210	26 696
mean / max length of proteins, aa	172 /2091	181 /2696	177 /2554	211 /1942	218 /1499	216 /6848	223 /3118
mean length of contig proteins, aa	219	222	222	246	252	216	223
# of proteins with function prediction (% of all proteins)	38 352 (55%)	38 206 (56%)	24 559[2] (57%)	25 077 (-)	25 087 (-)	13 345 (62%)	16 666 (62%)
# of unique database matches	23 302	21 925	nd	13 293	12 273	?	?
% of coding sequence[3]	86.6	86.4	86.4	40.9		86.7	87.3
# of ORFs in PFAMs	23 342	24 017	nd	14335	14805	11 829	15 007
# of ORFs in TIGRFAMs	8809	9058	nd	4810	5072		
# of PFAMs	2183	2174	nd	1743	1759	1868	1927
# of TIGRFAMS	1344	1361	nd	1143	1137		
# of enzymes (% of all ORFs)	8673 (12%)	8618 (13%)	nd	6166 (25%)[4]	6320 (25%)[4]	3641 (17%)	4508 (17%)

1. http://img.jgi.doe.gov/cgi-bin/pub/main.cgi , probably included only contigs for ORF prediction and analysis afterwards
* analysed in this study using ORFs predicted by Gill et al. 2006
2. Only ORFs in contig were annotated, % of all ORFs in contigs.
3. Calculated as ratio between length of all ORFs and sum of lengths of contig and singletons , for IMG/M calculated as length of ORFs/length of contig. For Gill et al. (2006) dataset, length of annotated ORFs/length of contigs and singletons
4. % of annotated ORFs

J. Craig Venter INSTITUTE

Gene Finder: MetaGene

- Exhaustive gene identification is a fundamental goal in all metagenomics projects. However, most metagenomic sequences are unassembled anonymous fragments, and conventional gene-finding methods require experimentally identified gene sequences ("training set") and cannot be applied.

- MetaGene utilizes di-codon frequencies estimated by the GC content of a given sequence with other various measures. MetaGene can predict prokaryotic genes from genomic sequences of a few hundred bases, with a sensitivity of 95% and a specificity of 90%.

- MetaGene uses two sets of codon frequency interpolations, one for bacteria and one for archaea, and automatically selects the proper set for a given sequence using the domain classification.

Published online 5 October 2006

Nucleic Acids Research, 2006, Vol. 34, No. 19 ***5623–5630***
doi:10.1093/nar/gkl723

MetaGene: prokaryotic gene finding from environmental genome shotgun sequences

Hideki Noguchi*, Jungho Park and Toshihisa Takagi

Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba 277-8562, Japan

J. Craig Venter INSTITUTE

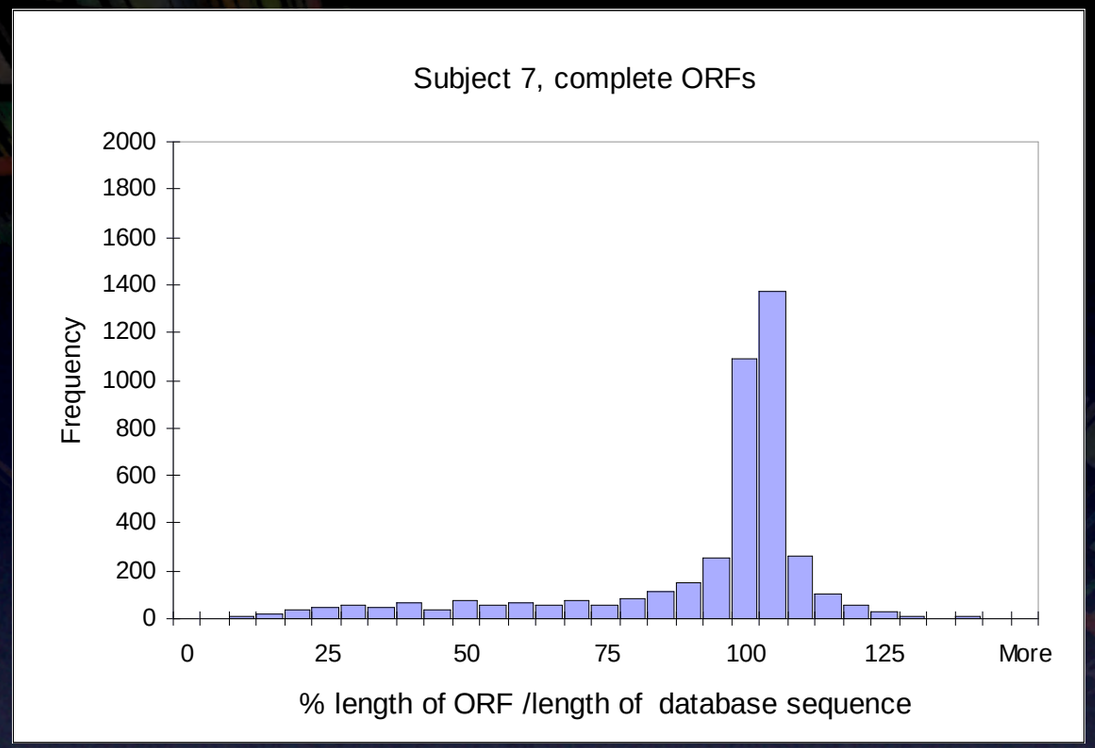
Comparison between different pipelines for ORFs and functions predictions in metagenomic dataset

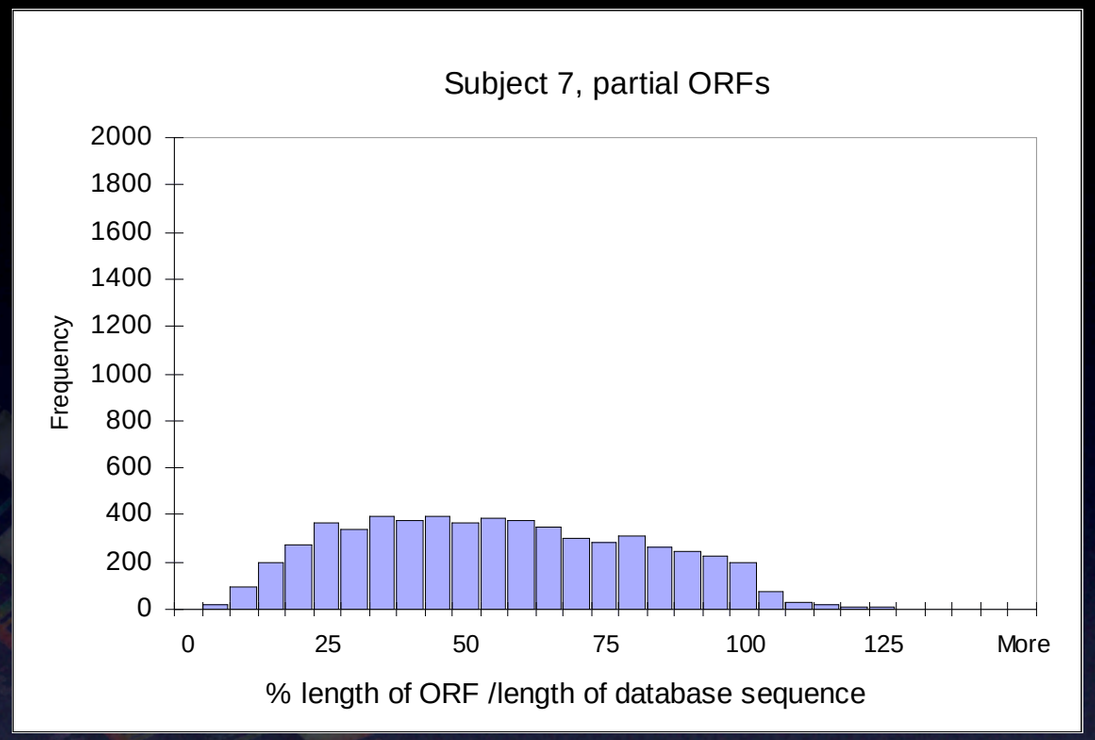
	This study			Gill et al. 2006		IMG/M pipeline [1]	
	#7	#8	#7&8	#7	#8	#7	#8
# of reads	65 059	74 462	139 521	139 521		65 059	74 462
# of contig	7 647	10 471	17 783	17 668		10 411	12 020
Mean/max/min length of contigs, bp	1767 /19 104 /131	1879 /25 060 /93	1856/45 326/92	1910 /44 747 /92		1521 /20 238 /67	1705 /25 061 /67
% reads in contigs	48%	63%		60%		?	?
# of singletons (% of all reads)	33 925 (52%)	27 658 (37%)	58 977 (42%)	56 292 (40%)		?	?
Average depth of coverage in contigs	1.92	2.04	2.07	2.13		?	?
# of ORFs	69 829	67 766	132 943	-	-	21 210	26 696
mean / max length of proteins, aa	172 /2091	181 /2696	177 /2554	211 /1942	218 /1499	216 /6848	223 /3118
mean length of contig proteins, aa	219	222	222	246	252	216	223
# of proteins with function prediction (% of all proteins)	38 352 (55%)	38 206 (56%)	24 559[2] (57%)	25 077 (-)	25 087 (-)	13 345 (62%)	16 666 (62%)
# of unique database matches	23 302	21 925	nd	13 293	12 273	?	?
% of coding sequence[3]	86.6	86.4	86.4	40.9		86.7	87.3
# of ORFs in PFAMs	23 342	24 017	nd	14335	14805	11 829	15 007
# of ORFs in TIGRFAMs	8809	9058	nd	4810	5072		
# of PFAMs	2183	2174	nd	1743	1759	1868	1927
# of TIGRFAMS	1344	1361	nd	1143	1137		
# of enzymes (% of all ORFs)	8673 (12%)	8618 (13%)	nd	6166 (25%)[4]	6320 (25%)[4]	3641 (17%)	4508 (17%)

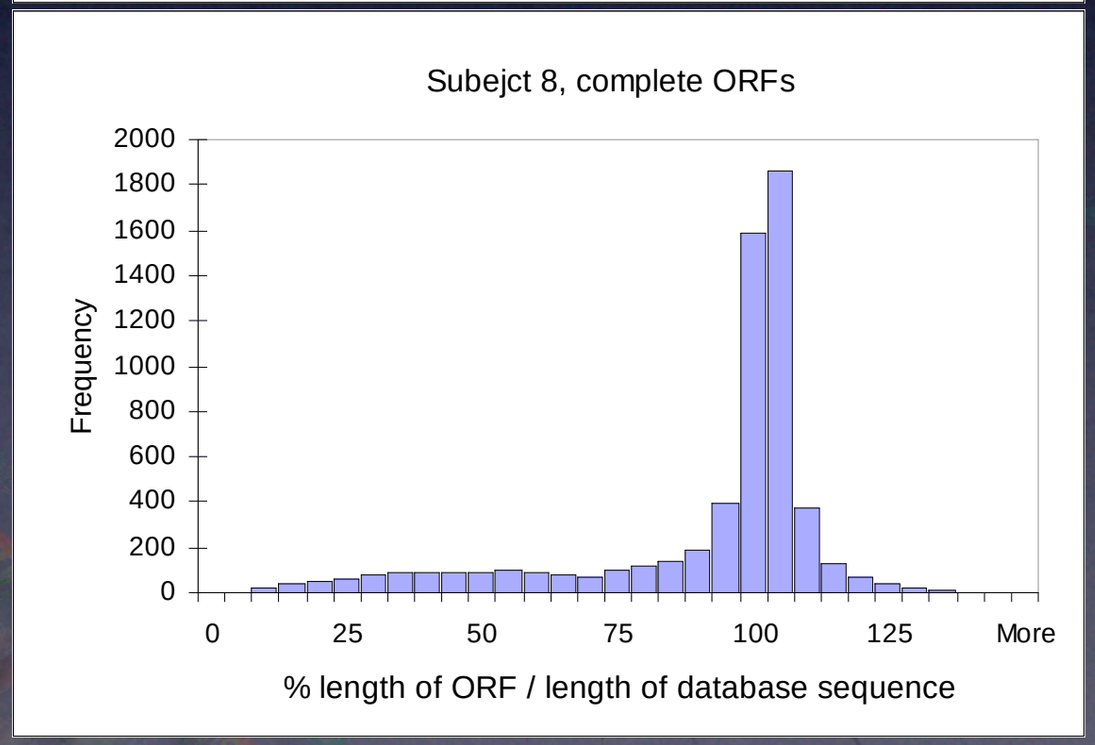
1. http://img.jgi.doe.gov/cgi-bin/pub/main.cgi , probably included only contigs for ORF prediction and analysis afterwards
* analysed in this study using ORFs predicted by Gill et al. 2006
2. Only ORFs in contig were annotated, % of all ORFs in contigs.
3. Calculated as ratio between length of all ORFs and sum of lengths of contig and singletons , for IMG/M calculated as length of ORFs/length of contig. For Gill et al. (2006) dataset, length of annotated ORFs/length of contigs and singletons
4. % of annotated ORFs

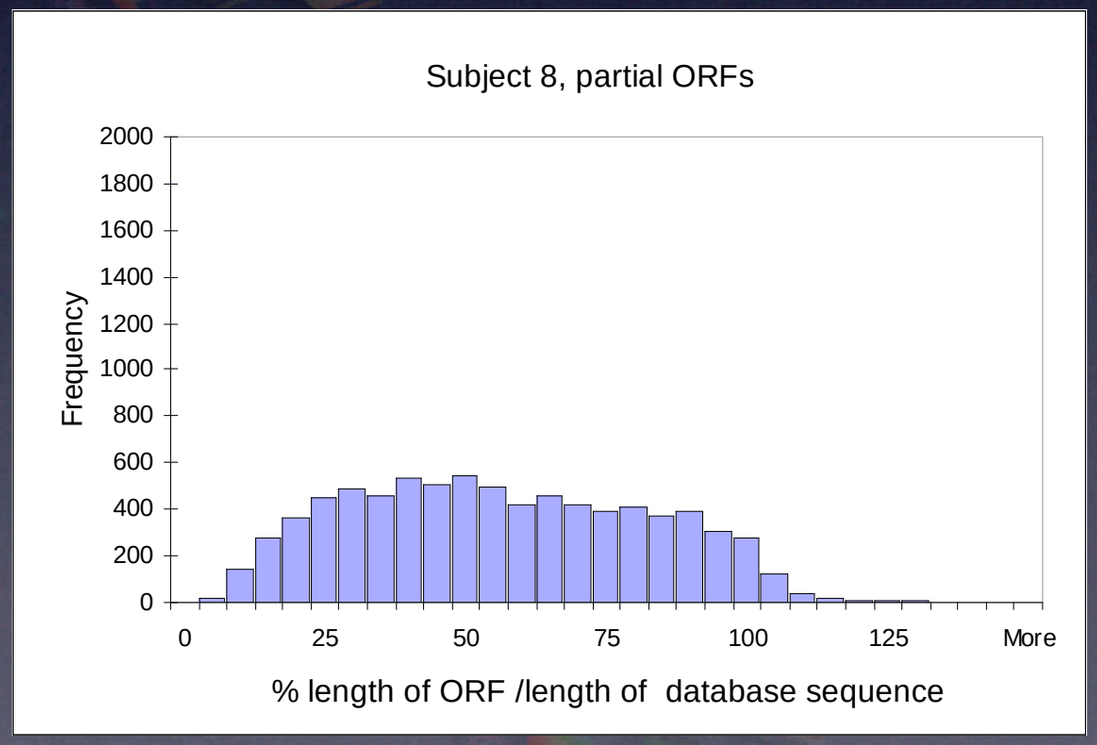
J. Craig Venter INSTITUTE

- Comparison of the length of the Metagene ORFs to their matching database sequences. Only the ORFs in contigs were compared.





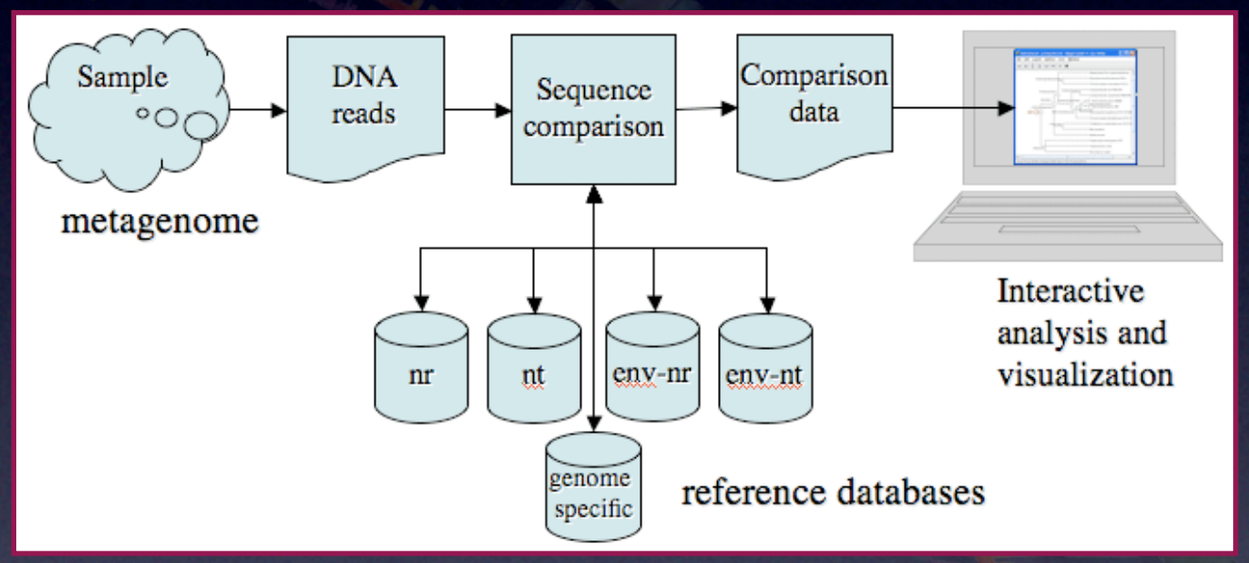




J. Craig Venter INSTITUTE

MEGAN: MEtaGenome ANalizer

- Allows the analysis of large metagenomic datasets:
  - The set of DNA reads (or contigs) is compared against databases of known sequences using BLAST or other comparison tools.
  - MEGAN is then used to compute and explore the taxonomical content of the dataset, employing the NCBI taxonomy.

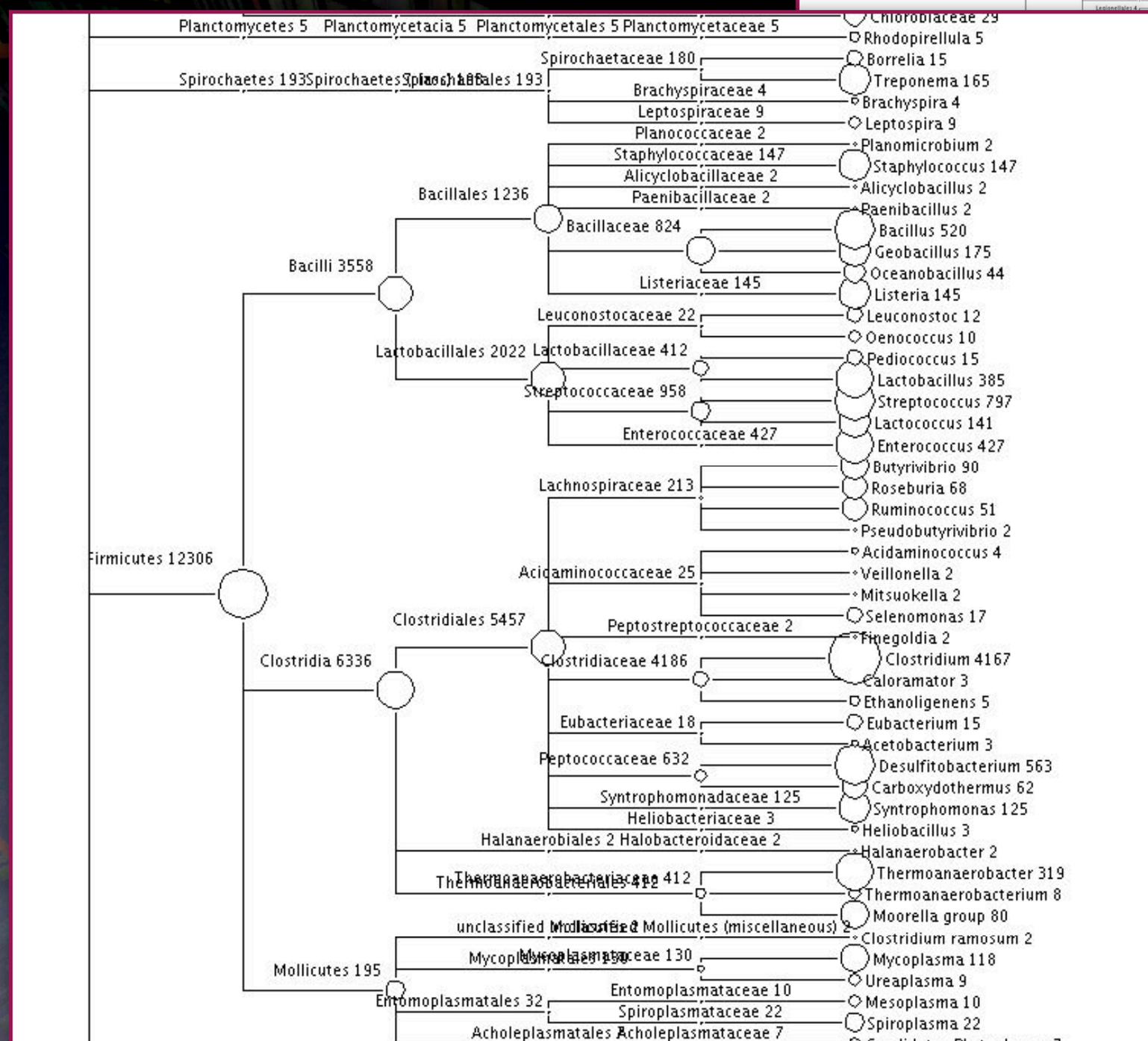


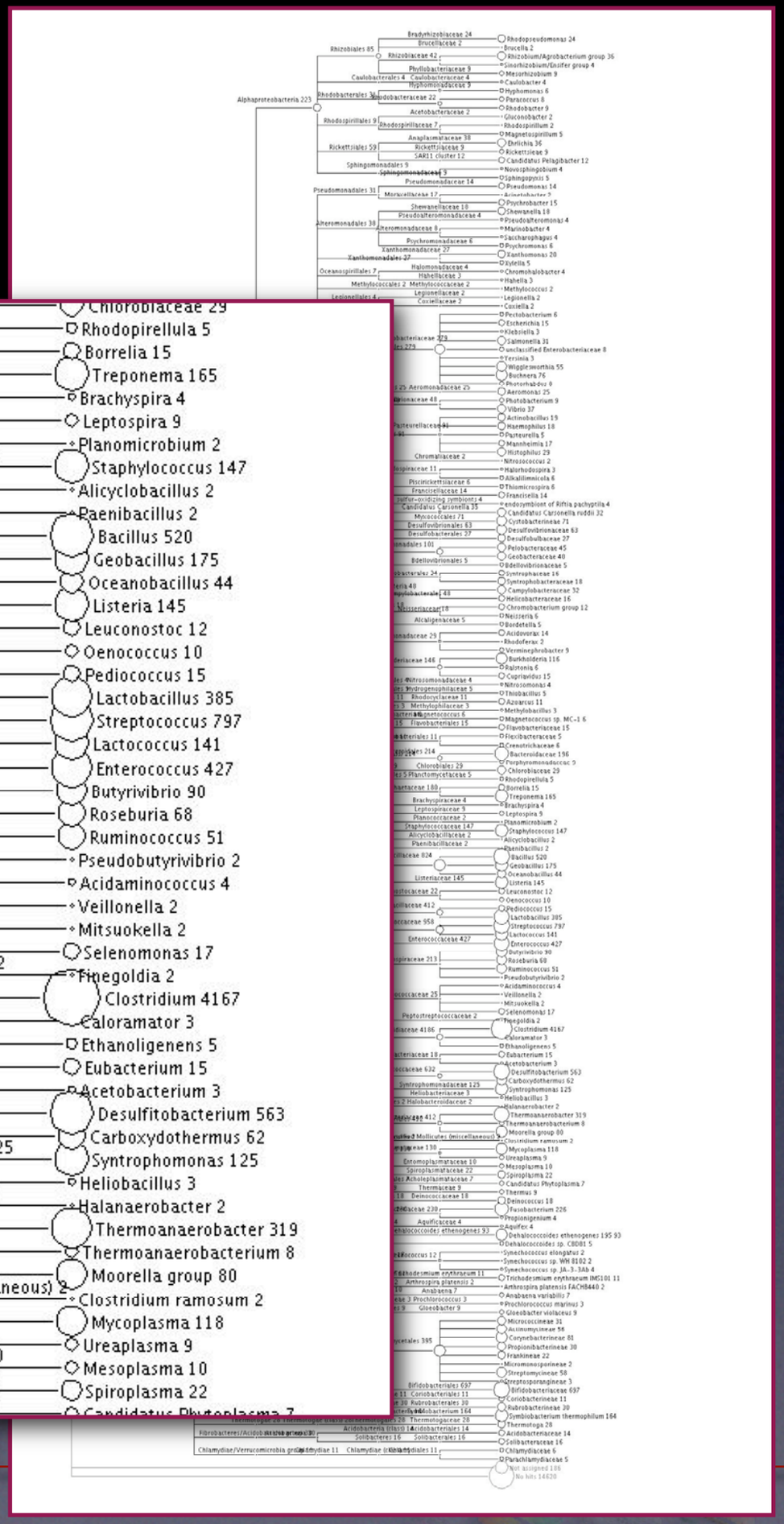
- Assignment of reads to Taxa: the taxonomical level of the assigned taxon reflects the level of conservation of the sampled sequence.
- MEGAN allows dissection of large datasets without the need for assembly or the targeting of specific phylogenetic markers. It provides graphical and statistical output for the comparison of different data sets.

J. Craig Venter INSTITUTE

www-ab.informatik.uni-tuebingen.de/software/megan/welcome.html

MEGAN analysis: genus level

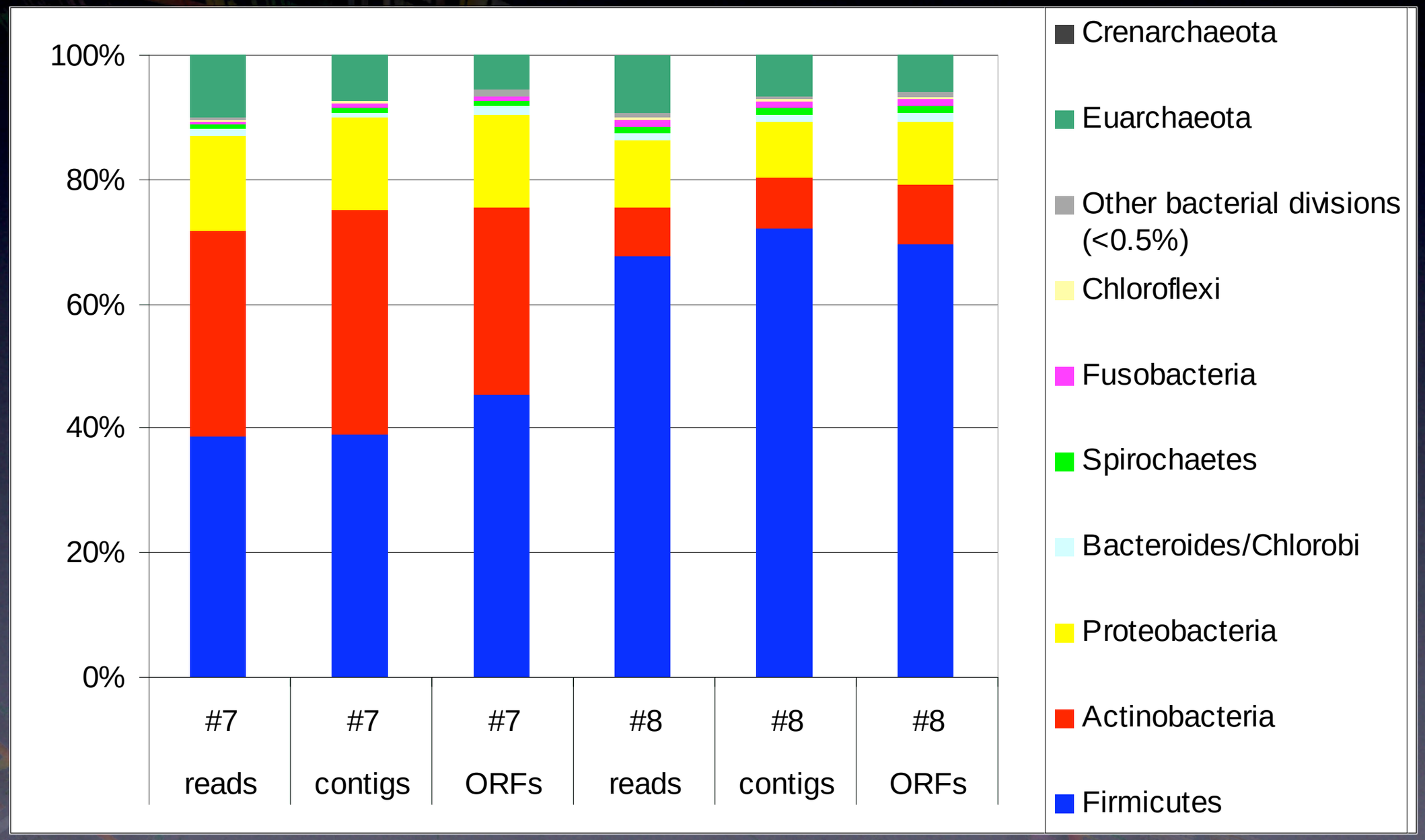




J. Craig Venter INSTITUTE

Human Colonic Microbiome: Taxonomic Assignments

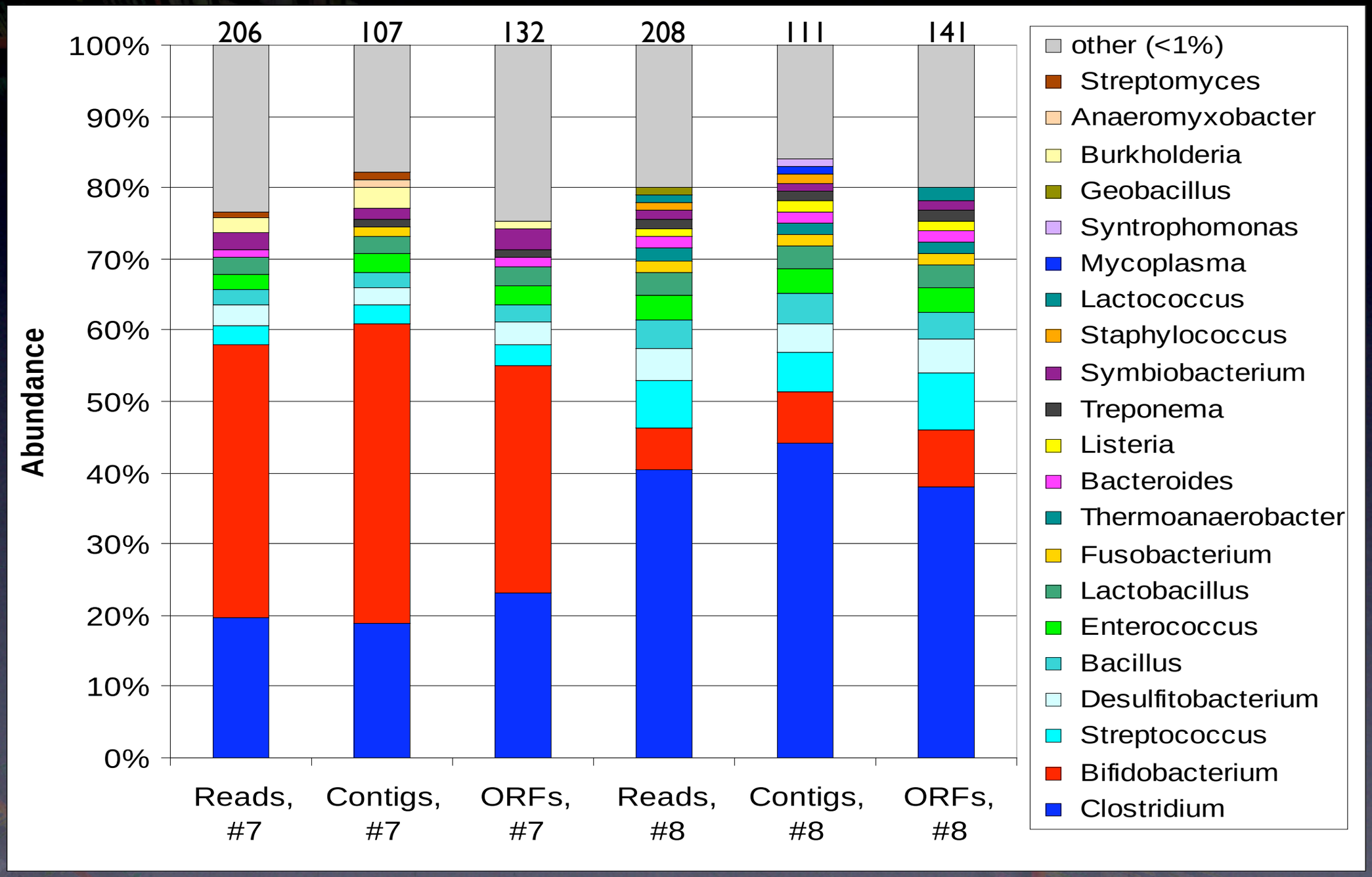
- Taxonomic assignment of reads, ORFs and contigs from subject 7 and 8 into Bacterial and Archaeal divisions by BlastN and Megan analyses.



J. Craig Venter INSTITUTE

Human Colonic Microbiome: Taxonomic Assignments

- Assignment of reads, ORFs and contigs from subject 7 and 8 into bacterial genera by BlastN and Megan analyses.



The "other" category includes genera which abundance was <1%.
The numbers above each bar shown total number of genera detected.

J. Craig Venter INSTITUTE

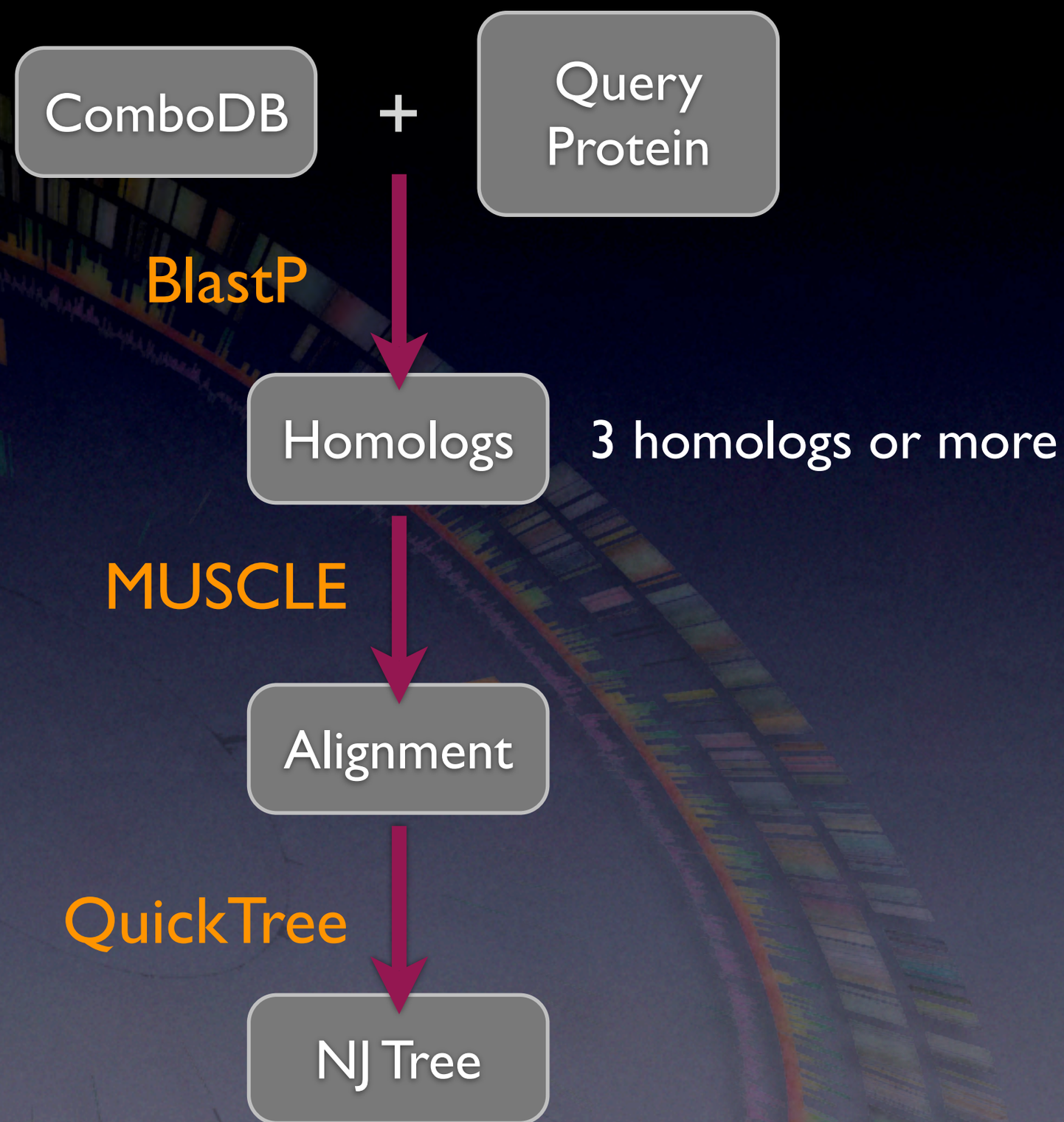
APIS

Rationale for APIS:

- Traditional taxonomy based on one gene - 16S rRNA.
- Sometimes other markers such as HSP70 used, and sometimes this conflicts with 16S.
- But systematic phylogeny of all proteins in genome individually not done -- would take too much manual effort.
- APIS makes this possible; it makes trees for each protein and summarizes results.

APIS Outline

For each query protein in genome / metagenome:



APIS: Phylogenomic breakdown of proteins from *Gut Metagenomics*

Search for ORF # or description

Recent Lineage Spe

Kingdom

Contained within B
Outgroup of Bacteri
Closest relative is ur
Contained within A
Outgroup of Archae
Outgroup of Viruse
Outgroup of Eukary
Contained within E
Contained within V
Outgroup of Eukary

Phylum

Closest relati
Outgroup of
Contained w
Contained w
Outgroup of
Outgroup of
Contained w
Outgroup of
Contained w
Outgroup of
Outgroup of
Outgroup of
Outgroup of
Outgroup of
Outgroup of
Outgroup of
Outgroup of
Outgroup of
Outgroup of
Contained w
Outgroup of
Outgroup of
Outgroup of
Outgroup of

Class

Closest r
Outgroup
Containe
Containe
Outgroup
Outgroup
Containe
Outgroup
Outgroup
Containe
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup

Order

Closest re
Outgroup
Containe
Containe
Outgroup
Outgroup
Outgroup
Containe
Outgroup
Outgroup
Outgroup
Containe
Outgroup
Outgroup
Outgroup
Containe
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup

Family

Closest r
Outgroup
Outgroup
Containe
Containe
Outgroup
Outgroup
Outgroup
Outgroup
Containe
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup
Outgroup

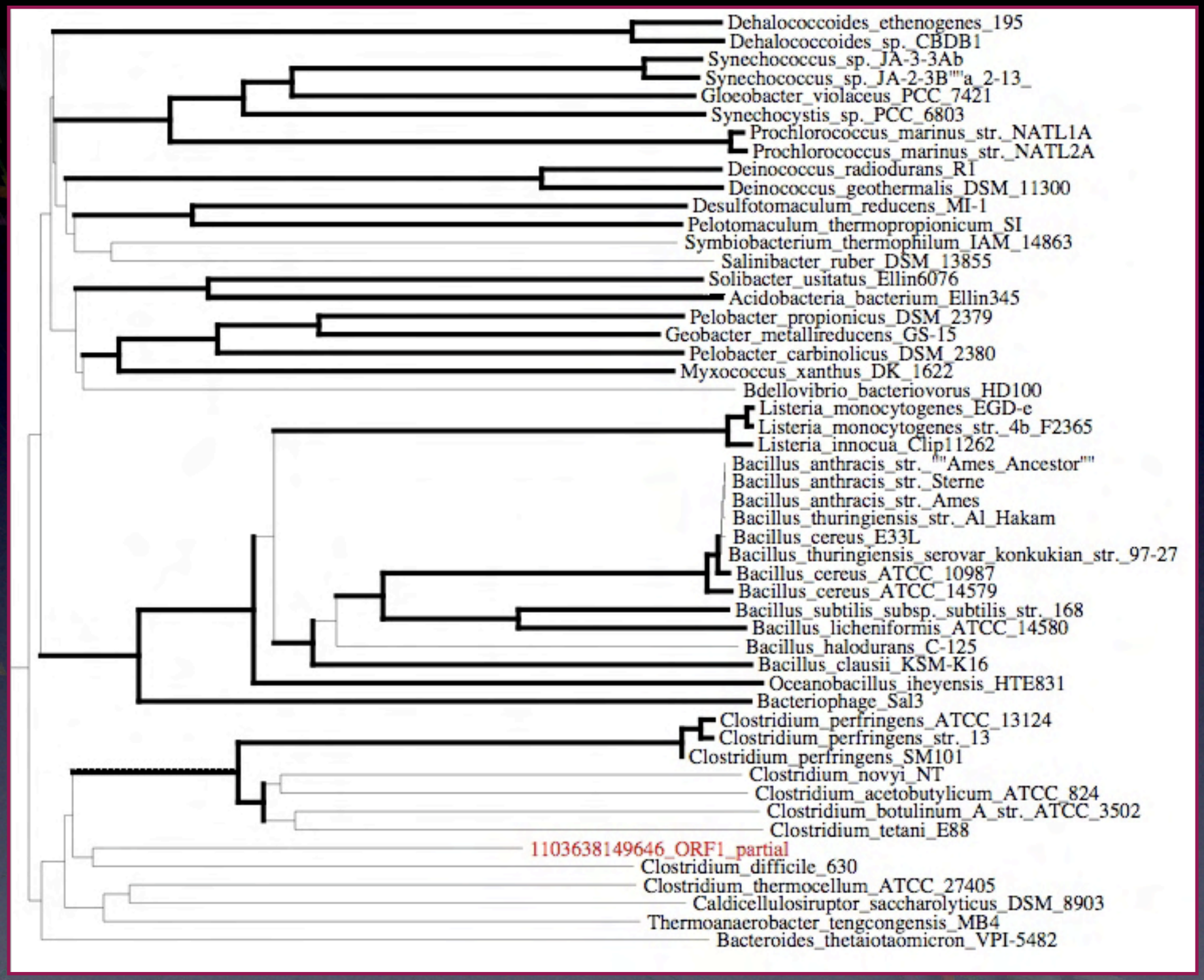
Genus

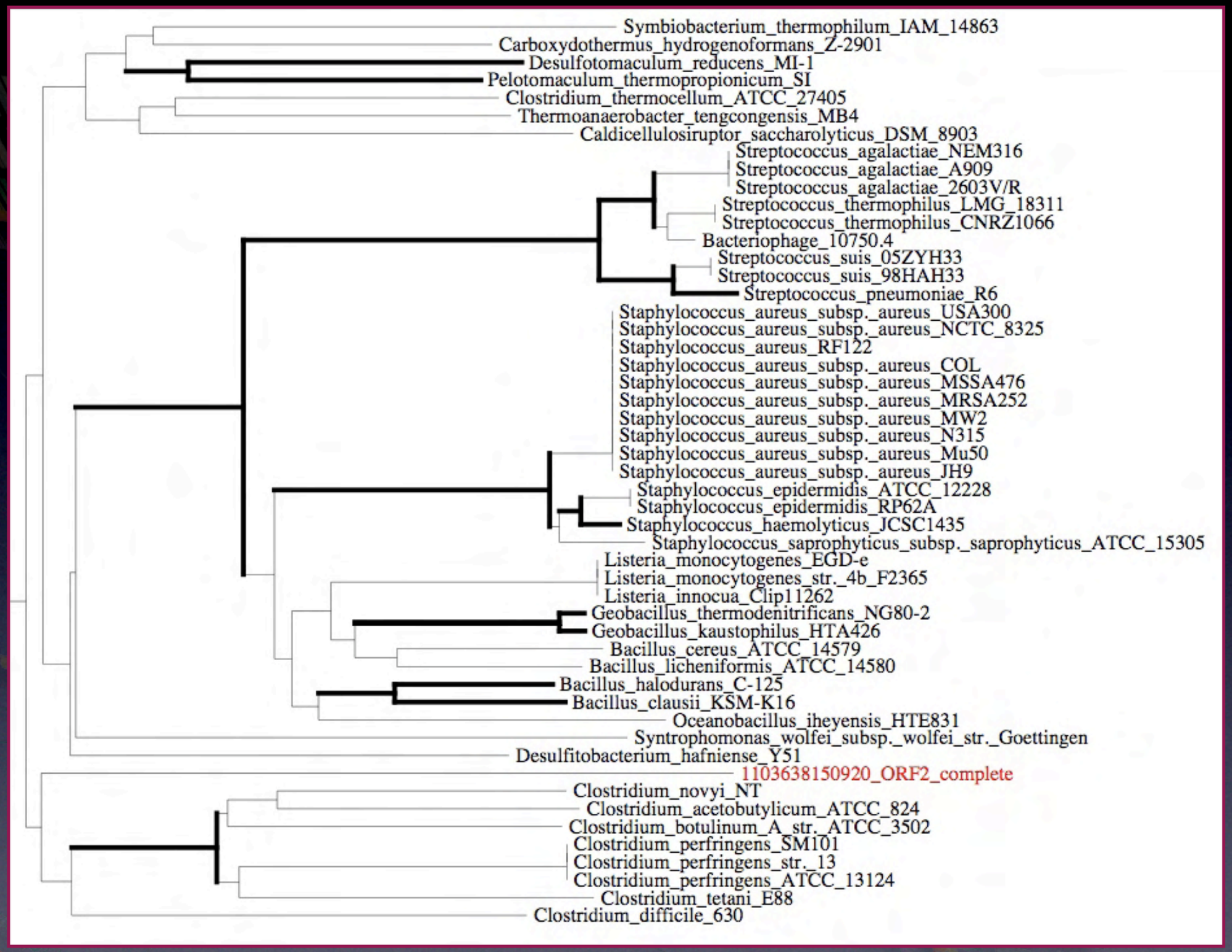
Closest relative is unresolved at genus level 11933 (39.1%)
Outgroup of Clostridium 3088 (10.1%)
Outgroup of Bifidobacteriaceae 1791 (5.9%)
Contained within Bifidobacteriaceae 1576 (5.2%)
Contained within Clostridium 1062 (3.5%)
Outgroup of Desulfitobacterium 666 (2.2%)
Outgroup of Streptococcus 585 (1.9%)
Outgroup of Lactobacillus 504 (1.7%)
Outgroup of Methanosphaera 459 (1.5%)
Outgroup of Methanothermobacter 457 (1.5%)
Outgroup of Bacillus 393 (1.3%)
Outgroup of Treponema 369 (1.2%)
Outgroup of Fusobacterium 324 (1.1%)
Outgroup of Caldicellulosiruptor 320 (1.0%)
Outgroup of Bacteroidaceae 276 (0.9%)
Outgroup of Enterococcus 265 (0.9%)
Outgroup of Symbiobacterium thermophilum 193 (0.6%)

J. Craig Venter™ INSTITUTE

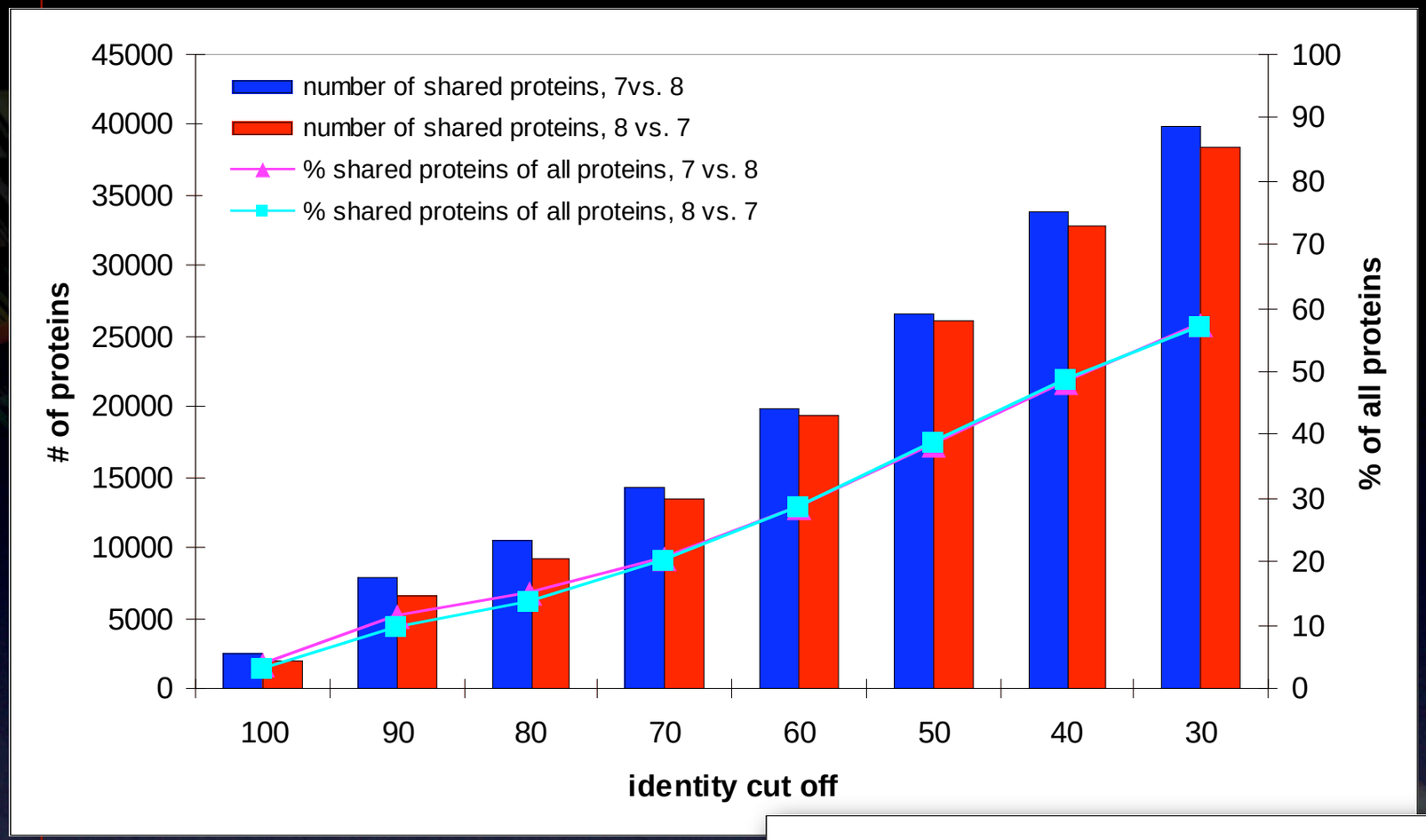
APIS: *Gut Metagenomics*: Contained within Firmicutes

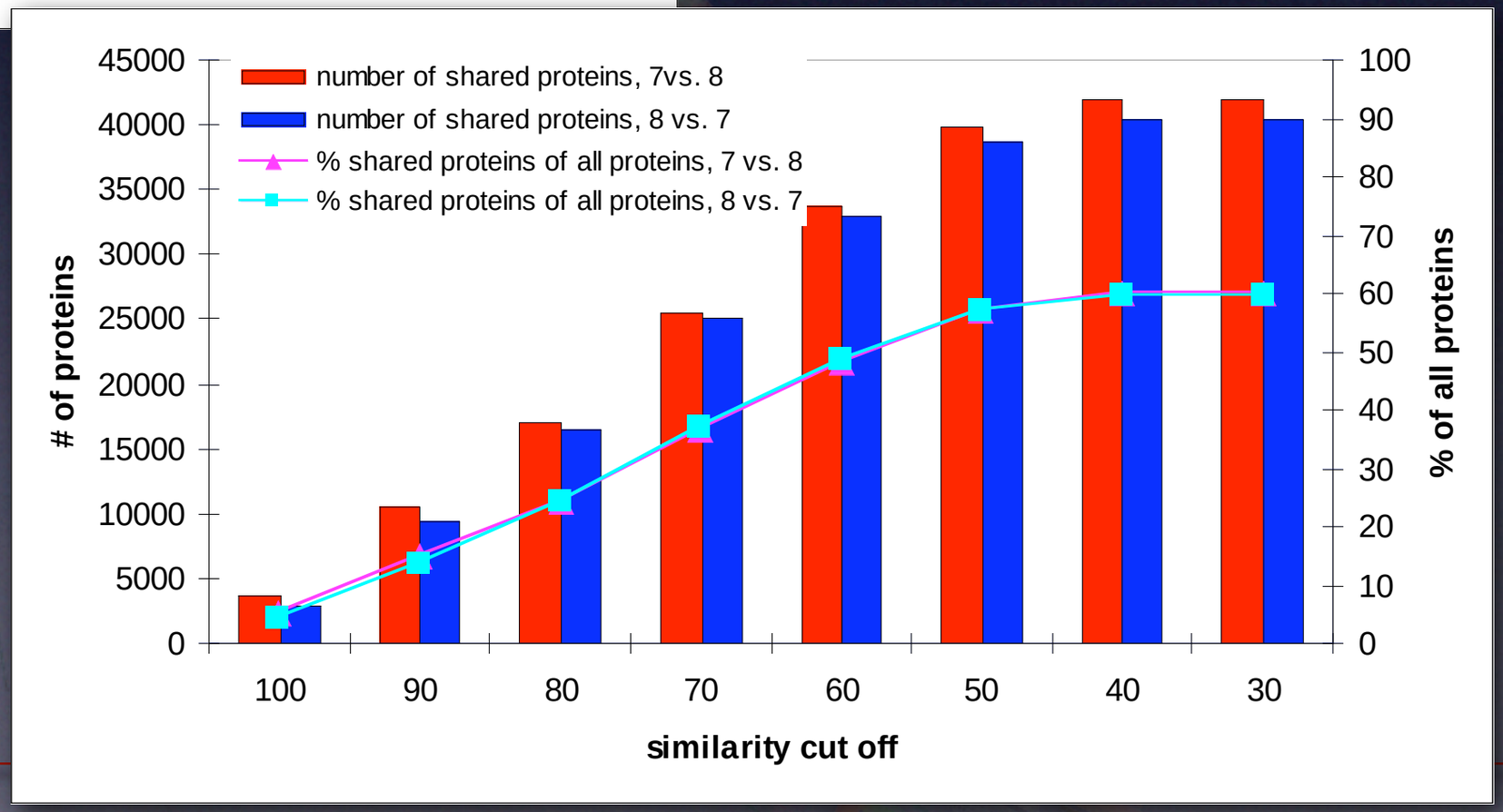
blast tree pdf alignment neighbors 1103638149372_ORF2_complete (frame 1)
blast tree pdf alignment neighbors 1103638149372_ORF3_partial (frame 1)
blast tree pdf alignment neighbors 1103638149388_ORF1_partial (frame 1)
blast tree pdf alignment neighbors 1103638149396_ORF1_partial (frame 1)
blast tree pdf alignment neighbors 1103638149408_ORF2_partial (frame 1)
blast tree pdf alignment neighbors 1103638149412_ORF1_complete (frame 1)
blast tree pdf alignment neighbors 1103638149419_ORF2_complete (frame 1)
blast tree pdf alignment neighbors 1103638149420_ORF2_complete (frame 1)
blast tree pdf alignment neighbors 1103638149443_ORF1_partial (frame 1)
blast tree pdf alignment neighbors 1103638149453_ORF1_partial (frame 1)
blast tree pdf alignment neighbors 1103638149453_ORF2_complete (frame 1)
blast tree pdf alignment neighbors 1103638149456_ORF2_complete (frame 1)
blast tree pdf alignment neighbors 1103638149458_ORF2_complete (frame 1)
blast tree pdf alignment neighbors 1103638149458_ORF4_partial (frame 1)
blast tree pdf alignment neighbors 1103638149462_ORF3_partial (frame 1)
blast tree pdf alignment neighbors 1103638149512_ORF2_complete (frame 1)
blast tree pdf alignment neighbors 1103638149548_ORF1_partial (frame 1)
blast tree pdf alignment neighbors 1103638149552_ORF1_complete (frame 1)
blast tree pdf alignment neighbors 1103638149556_ORF1_partial (frame 1)
blast tree pdf alignment neighbors 1103638149561_ORF1_partial (frame 1)


Dehalococcoides_ethenogenes_195
Dehalococcoides_sp._CBDB1
Synechococcus_sp._JA-3-3Ab
Synechococcus_sp._JA-2-3B""a_2-13_
Gloeobacter_violaceus_PCC_7421
Synechocystis_sp._PCC_6803
Prochlorococcus_marinus_str._NATL1A
Prochlorococcus_marinus_str._NATL2A
Deinococcus_radiodurans_R1
Deinococcus_geothermalis_DSM_11300
Desulfotomaculum_reducens_MI-1
Pelotomaculum_thermopropionicum_SI
Symbiobacterium_thermophilum_IAM_14863
Salinibacter_ruber_DSM_13855
Solibacter_usitatus_Ellin6076
Acidobacteria_bacterium_Ellin345
Pelobacter_propionicus_DSM_2379
Geobacter_metallireducens_GS-15
Pelobacter_carbinolicus_DSM_2380
Myxococcus_xanthus_DK_1622
Bdellovibrio_bacteriovorus_HD100
Listeria_monocytogenes_EGD-e
Listeria_monocytogenes_str._4b_F2365
Listeria_innocua_Clip11262
Bacillus_anthracis_str._""Ames_Ancestor""
Bacillus_anthracis_str._Sterne
Bacillus_anthracis_str._Ames
Bacillus_thuringiensis_str._Al_Hakam
Bacillus_cereus_E33L
Bacillus_thuringiensis_serovar_konkukian_str._97-27
Bacillus_cereus_ATCC_10987
Bacillus_cereus_ATCC_14579
Bacillus_subtilis_subsp._subtilis_str._168
Bacillus_licheniformis_ATCC_14580
Bacillus_halodurans_C-125
Bacillus_clausii_KSM-K16
Oceanobacillus_iheyensis_HTE831
Bacteriophage_Sal3
Clostridium_perfringens_ATCC_13124
Clostridium_perfringens_str._13
Clostridium_perfringens_SM101
Clostridium_novyi_NT
Clostridium_acetobutylicum_ATCC_824
Clostridium_botulinum_A_str._ATCC_3502
Clostridium_tetani_E88
1103638149646_ORF1_partial
Clostridium_difficile_630
Clostridium_thermocellum_ATCC_27405
Caldicellulosiruptor_saccharolyticus_DSM_8903
Thermoanaerobacter_tengcongensis_MB4
Bacteroides_thetaiotaomicron_VPI-5482


Symbiobacterium_thermophilum_IAM_14863
Carboxydothermus_hydrogenoformans_Z-2901
Desulfotomaculum_reducens_MI-1
Pelotomaculum_thermopropionicum_SI
Clostridium_thermocellum_ATCC_27405
Thermoanaerobacter_tengcongensis_MB4
Caldicellulosiruptor_saccharolyticus_DSM_8903
Streptococcus_agalactiae_NEM316
Streptococcus_agalactiae_A909
Streptococcus_agalactiae_2603V/R
Streptococcus_thermophilus_LMG_18311
Streptococcus_thermophilus_CNRZ1066
Bacteriophage_10750.4
Streptococcus_suis_05ZYH33
Streptococcus_suis_98HAH33
Streptococcus_pneumoniae_R6
Staphylococcus_aureus_subsp._aureus_USA300
Staphylococcus_aureus_subsp._aureus_NCTC_8325
Staphylococcus_aureus_RF122
Staphylococcus_aureus_subsp._aureus_COL
Staphylococcus_aureus_subsp._aureus_MSSA476
Staphylococcus_aureus_subsp._aureus_MRSA252
Staphylococcus_aureus_subsp._aureus_MW2
Staphylococcus_aureus_subsp._aureus_N315
Staphylococcus_aureus_subsp._aureus_Mu50
Staphylococcus_aureus_subsp._aureus_JH9
Staphylococcus_epidermidis_ATCC_12228
Staphylococcus_epidermidis_RP62A
Staphylococcus_haemolyticus_JCSC1435
Staphylococcus_saprophyticus_subsp._saprophyticus_ATCC_15305
Listeria_monocytogenes_EGD-e
Listeria_monocytogenes_str._4b_F2365
Listeria_innocua_Clip11262
Geobacillus_thermodenitrificans_NG80-2
Geobacillus_kaustophilus_HTA426
Bacillus_cereus_ATCC_14579
Bacillus_licheniformis_ATCC_14580
Bacillus_halodurans_C-125
Bacillus_clausii_KSM-K16
Oceanobacillus_iheyensis_HTE831
Syntrophomonas_wolfei_subsp._wolfei_str._Goettingen
Desulfitobacterium_hafniense_Y51
1103638150920_ORF2_complete
Clostridium_novyi_NT
Clostridium_acetobutylicum_ATCC_824
Clostridium_botulinum_A_str._ATCC_3502
Clostridium_perfringens_SM101
Clostridium_perfringens_str._13
Clostridium_perfringens_ATCC_13124
Clostridium_tetani_E88
Clostridium_difficile_630

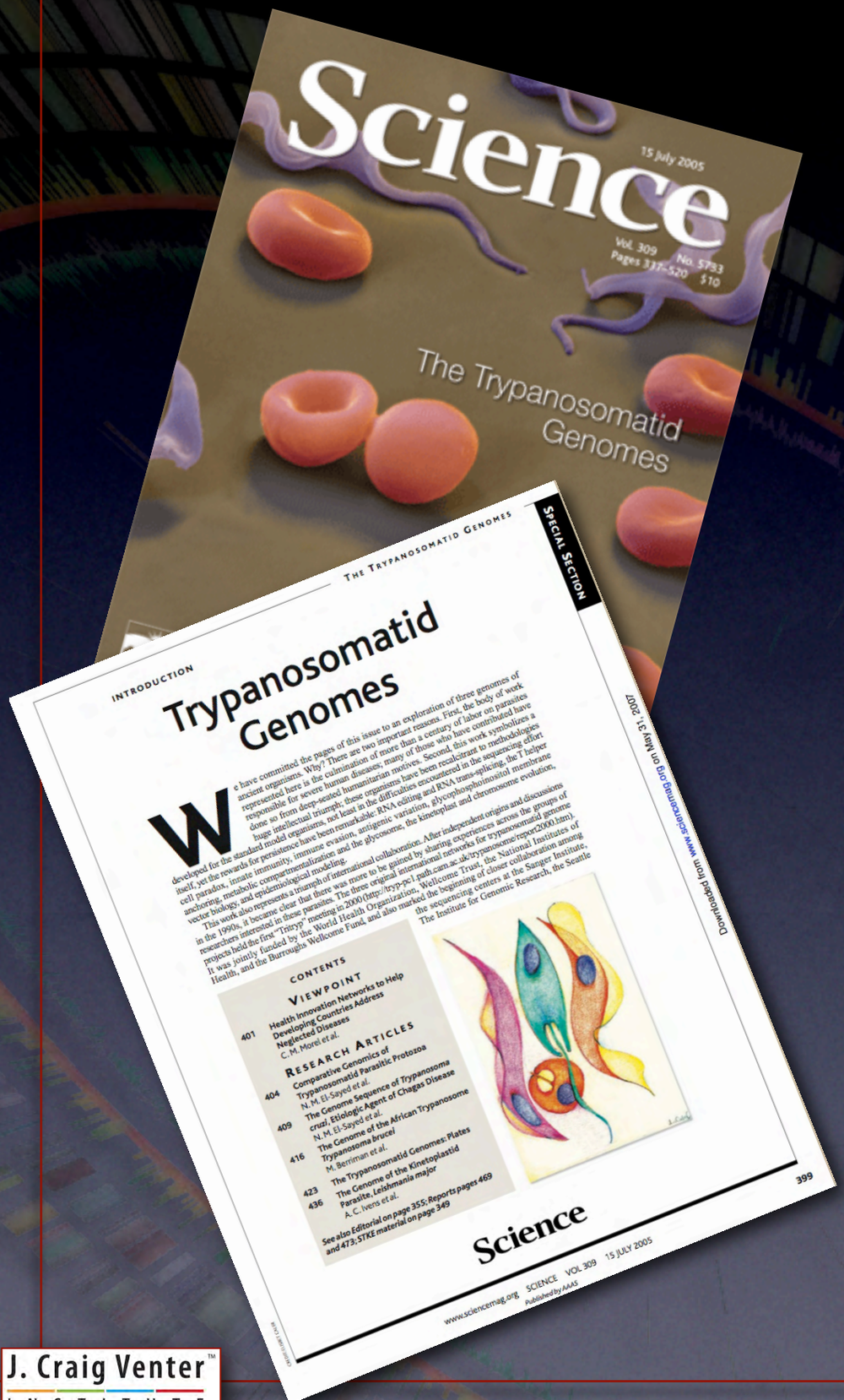
Proteins shared by subjects 7 and 8 based on Blastp





J. Craig Venter INSTITUTE

Comparative Protein Domain Analysis: DOMAINER



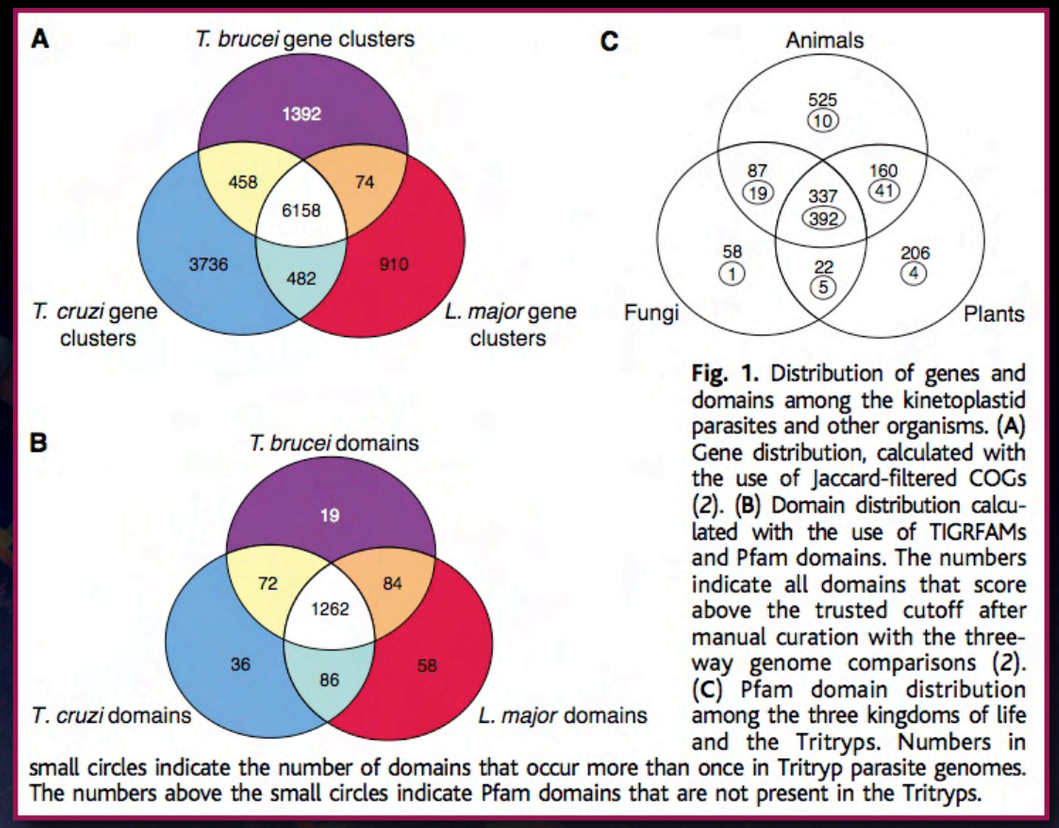


Fig. 1. Distribution of genes and domains among the kinetoplastid parasites and other organisms. (**A**) Gene distribution, calculated with the use of Jaccard-filtered COGs (*2*). (**B**) Domain distribution calculated with the use of TIGRFAMs and Pfam domains. The numbers indicate all domains that score above the trusted cutoff after manual curation with the three-way genome comparisons (*2*). (**C**) Pfam domain distribution among the three kingdoms of life and the Tritryps. Numbers in small circles indicate the number of domains that occur more than once in Tritryp parasite genomes. The numbers above the small circles indicate Pfam domains that are not present in the Tritryps.

<u>Output html file (default):</u>

- Venn diagram
- Domain descriptions
- List of domain containing genes
- links to euk or prok Manatee, pfam webpage, tigrfam webpage and sybil.

<u>Output txt file:</u>

-Tabulated file that can be imported into excel.

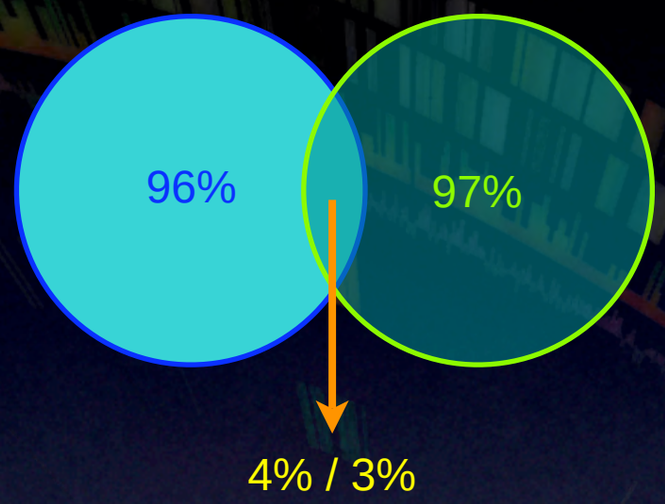
<u>Output parameters:</u>

-o <output file> [STDOUT]
-text T (text output file)

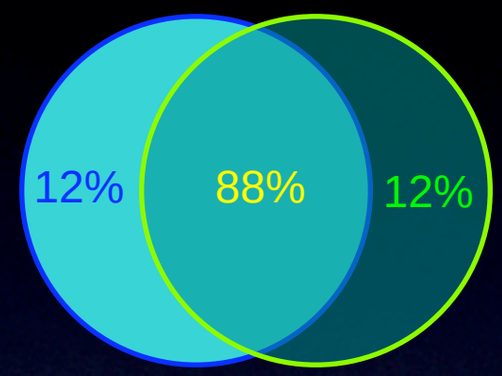
J. Craig Venter INSTITUTE

Elisabet Caler - Hernan Lorenzi / JCVI Euk. Annotation Group

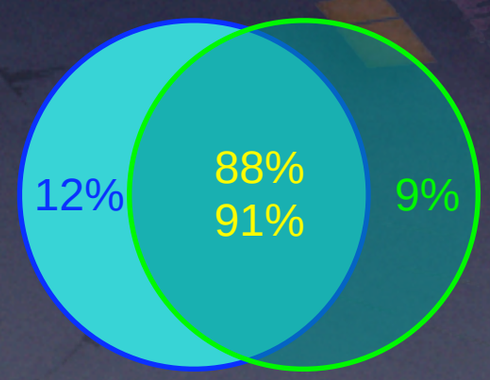
Species diversity *vs.* Functional diversity between patients 7 & 8



Shared proteins
identified by BlastP

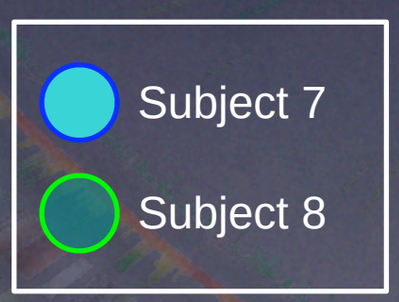


Shared protein domains identified
by Comparative Domain Analysis



Shared genera

BlastN + MEGAN



J. Craig Venter INSTITUTE

Comparative Protein Domain Analysis

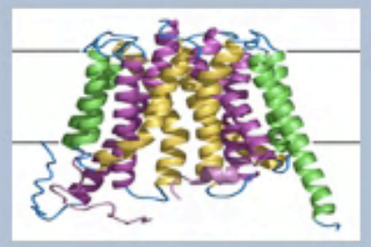
Hypothesis:
Protein Domain <=> Function

Domain		#7	#8
PF02518	ATPase, histidine kinase-, DNA gyrase B-, and HSP90-like domain protein	273	381
PF00528	ABC transporter, permease protein	293	299
PF01381	Helix-turn-helix	204	352
TIGR01494	ATPase, P-type (transporting), HAD superfamily, subfamily IC	265	216
PF01554	MatE	212	251
PF00512	His Kinase A (phosphoacceptor) domain	166	270
PF05738	Cna protein B-type domain	186	230
PF00486	Transcriptional regulatory protein, C terminal	134	220
TIGR00650	GTP-binding conserved hypothetical protein	159	170
PF00589	site-specific recombinase, phage integrase family	123	204
PF04055	radical SAM domain protein	130	180
PF00583	acetyltransferase, GNAT family	139	165
PF02687	efflux ABC transporter, permease protein	120	172
TIGR00231	small GTP-binding protein domain	142	147
PF08281	Sigma-70, region 4	116	154
PF01926	GTPase of unknown function	140	130
PF00271	Helicase conserved C-terminal domain	148	119
PF00702	haloacid dehalogenase-like hydrolase	132	128

- 88% of protein domains were present in both subjects,
- Subject –specific protein domains were present in very low numbers (≤ 7 hits)
- The abundances of shared protein domains were more or less similar between subjects
- Indicates a high functional similarity of gut microbiota in subjects

J. Craig Venter INSTITUTE

Human Colonic Microbiome: Transporter Analysis



< home >

All Organisms (292 total)
-- Select Organism --

Bacteria (235 total)
-- Select Bacteria --

Archaea (24 total)
-- Select Archaea --

Eukaryota (33 total)
-- Select Eukaryota --

TransAAP NEW

Organisms List

Structure

Outer Mem. Channel

Search

Compare Organisms

Blast

Phylogenetics

Links

56354

J. Craig Venter INSTITUTE

Genomic Comparisons of Membrane Transport Systems

TransportDB

TransportDB is a relational database describing the predicted cytoplasmic membrane transport protein complement for organisms whose complete genome sequence are available. For each organism, its complete membrane transport complement was identified, classified into protein families according to the TC classification system, and functional predictions are provided. Complete lists of the transporters from each organism are accessible from the pull down menus on this page.

For each organism, a summary page is available, overviewing the whole transporter system, including transporter types and individual transporter families. For individual transporter types, a detailed list of transporters with their possible substrates is shown with links to individual protein page which contains protein sequence and annotation information. You can also compare the transporter system from two or more different organisms.

A search engine is set up for easy search in our transporter database for transporter type, family, individual proteins and their substrates. You can also blast search your protein sequence against our transporter database.

With the rapid development of genomic sequencing both in TIGR and in other institutes, more and more genomes are available for the analysis of their transporter system. We will keep updating this site with the newly published genomes. If you have any suggestions, corrections, or comments on our site, please contact us. We are currently working on providing additional functionality for this database.

Click to see the complete list of organisms in TransportDB

Watch YouTube Video on F-type ATPase

[redacted]

Transporter Proteins in
Human Gut Metagenomics H7

ATP-Dependent | Ion Channel | Phosphotransferase System | Secondary Transporter | Unclassified Transporters

print all transporters | Tab-delimited format

Genome Size(Mb):	N/A
Total Transporter Proteins:	3379
No. of Transporters per Mb genome:	N/A

TRANSPORTER TYPE	TOTAL PROTEINS (PERCENT)
ATP-Dependent	2201 (65.1%)

	Number	Percent
ABC Family*	1795	81.6%
ArsAB Family	2	0.1%
F-ATPase Family*	158	7.2%
H+-PPase Family	2	0.1%
IISP Family	75	3.4%
P-ATPase Family*	169	7.7%

Ion Channels — 40 (1.2%)

	Number	Percent
ClC Family	17	42.5%
MIP Family	5	12.5%
MIT Family	7	17.5%
MscL Family	3	7.5%
UAC Family	1	2.5%
VIC Family	7	17.5%

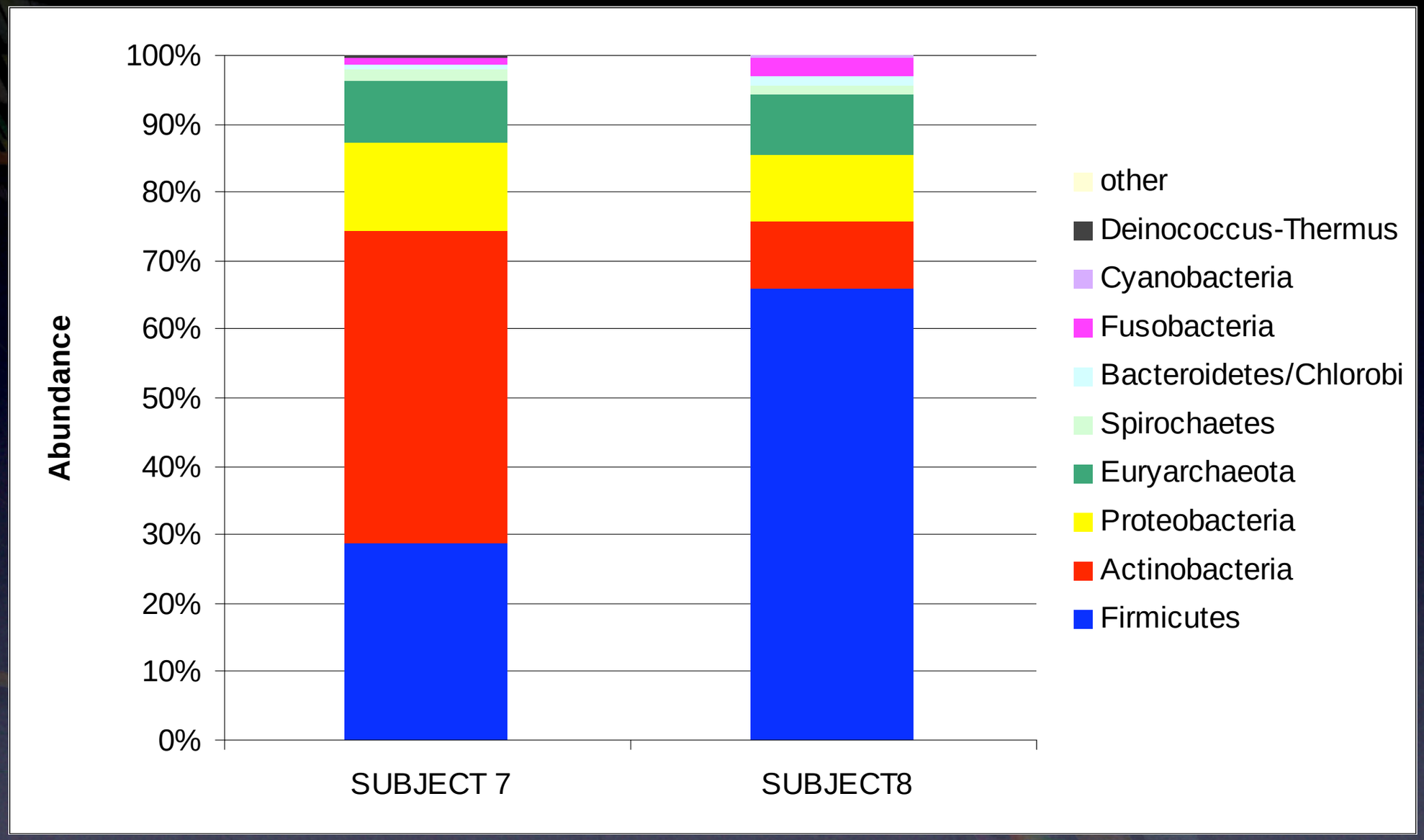
Phosphotransferase System (PTS) — 219 (6.5%)

	Number	Percent
GPTS Family	45	-
SSPTS Family	174	-

J. Craig Venter INSTITUTE

Qinghu Ren & Ian Paulsen - TransportDB - http://172.17.110.151:81/transporter/wb/

Taxonomic assignments of transporter proteins



- 62% of the transporter proteins for subject 7 and 55% of the transporter proteins for subject 8 could be assigned at least at the Division level.
- Transporter proteins originated most commonly from Bifidobacterium (35%) for Subject 7, and Clostridium (26%) for subject 8

J. Craig Venter INSTITUTE

Acknowledgements

J. Craig Venter INSTITUTE

Karen E. Nelson

Pirjo Wacklin

Joanne B. Emerson

Robert T. DeBoy

TransportDB: Ian Paulsen - Qinghu Ren

APIS: Jonathan Badger

SUNY Buffalo:
Steven R. Gill

UMB - SOM:
Claire M. Fraser-Liggett

NIDCR National Institute of Dental and Craniofacial Research

J. Craig Venter INSTITUTE