A Data Analysis and Coordination Center for the Human Microbiome Project

Owen White
Institute for Genome Sciences
University of Maryland
School of Medicine

HMP Initiatives

- **Initiative 1: Data Resource Generation** - sequencing of 400 strains of prokaryotic microbes from different body regions; recruitment of donors; collection of samples; metagenomic sequence analysis;
- **Initiative 2: Demonstration Projects** - relationship between changes in the human microbiome and health or disease onset;
- **Initiative 3: Technology Development** - development of improved culturing techniques; individual microbe sequencing;
- **Initiative 4: Ethical, Legal, and Social Implications Research** - clinical and health; forensics; uses of new technologies; ownership of microbiome;
- **Initiative 5: Data Analysis and Coordinating Center** - tracking, storing and distributing data; data retrieval tools; coordination of analyses and metadata standards; creation of a portal for international activities; and
- **Initiative 6: Computational Tool Development** - new tool development; next generation sequencing platforms; large, complex sequence data; functional data and metadata.

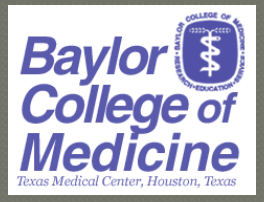
DACC Roles and Responsibilities

- Tracking, storing and distributing data
- Data and metadata standardization
- Distribution of software tools and pipelines
- Support for data analysis
- Providing a repository of protocols and SOPs
- Development of a comprehensive web portal

DACC Collaborators

- The Institute for Genome Sciences
  - Project Coordination
  - Web Portal
  - Core Pipelines
  - Data and Metadata Management
- The Joint Genome Institute
  - HMP Project Catalog (GOLD)
  - Metagenome Analysis Strategies
- Lawrence Berkeley National Lab
  - 16S Data Management (greengenes)
  - HMP Data Analysis System (IMG)
- University of Colorado at Boulder
  - Metadata Standards
  - Statistical and Analytical Tools

In partnership with….



Human Microbiome Project DACC - Home - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.hmpdacc.org/

Human Microbiome Project DACC - H...

REFERENCE GENOMES | MICROBIOME ANALYSES | IMPACTS ON HEALTH | TECHNOLOGY DEVELOPMENT | ETHICAL IMPLICATIONS | OUTREACH

SEARCH

HUMAN MICROBIOME PROJECT

Welcome to the Data Analysis and Coordination Center (DACC) for the Human Microbiome Project (HMP),

launched by the National Institutes of Health Roadmap for Medical Research, and designed to fuel research into the multitude of microbes that live in the various environments of the human body. A major goal of the HMP is to look for correlations between changes in the microbiome and human health. The HMP DACC is the central repository for all HMP data. More information about the project can be found on the NIH Roadmap site at http://nihroadmap.nih.gov/hmp.

Focus Areas of the HMP

Impacts On Health

Demonstration projects to determine the relationship between human health and changes in the human microbiome through the use of 16S and whole metagenome sequencing of clinical samples of interest...

Microbial Reference Genomes | Human Microbial Sampling | Impacts on Health | Technology Development | Ethical Implications | Outreach

CURRENT NEWS

- Expanding Knowledge about the Human Microbiome Will Lead to New Clinical Pathology Laboratory Tests
- Interactions between human and microbial cells determine health, physical well-being: Researchers. The Medical News
- Data acquisition and coordination key to human microbiome project

+ DACC MEMBER ORGANIZATIONS
+ NIH SITES
+ SEQUENCING CENTERS
+ INTERNATIONAL SITES

DATA

- Get Sequences
- BLAST against Reference Genomes
- Get WGS Sequences

TOOLS & PROTOCOLS

- By research area
- By type

PUBLICATIONS

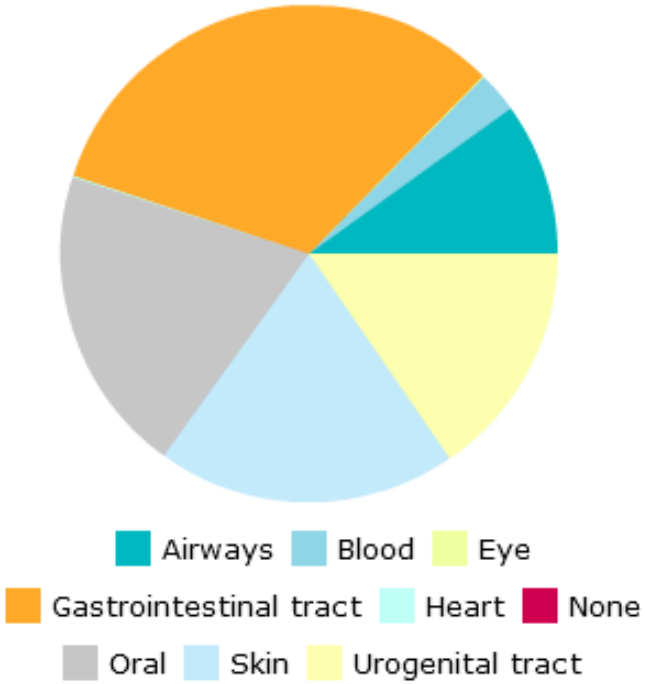
- The right to ignore genetic status of late onset genetic disease in the genomic era...
- Drawing the line between commensal and pathogenic Gardnerella vaginalis through genome analysis...
- Expansion of ribosomally produced natural products: a nitrile hydratase...

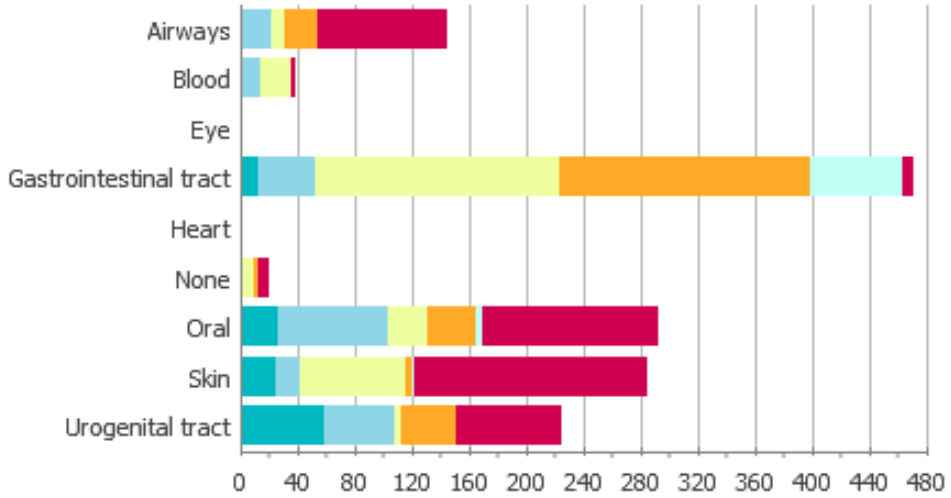
Human Microbiome Research

For Internal HMP Use

By HMP body isolation site

Breakdown by body site



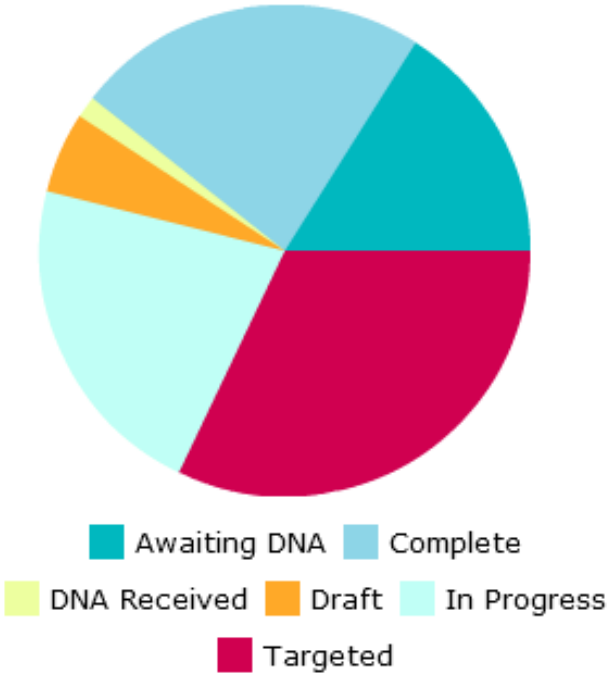


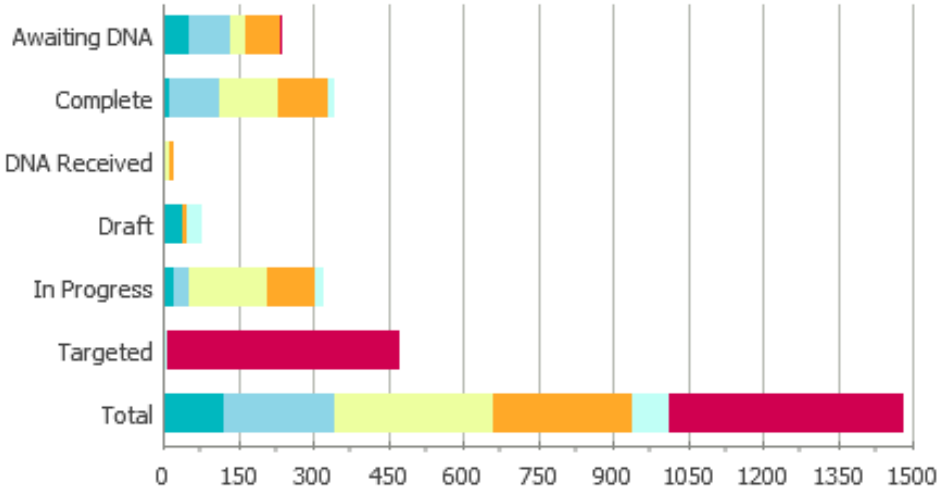
	JCVI	Baylor	WashU	Broad	other	unassigned	total
Airways	1	21	9	23	0	91	145
Blood	0	14	21	0	1	3	39
Eye	0	0	0	0	1	0	1
Gastrointestinal tract	12	40	171	176	64	8	471
Heart	0	2	0	0	0	0	2
None	1	1	7	4	0	7	0
Oral	26	77	28	33	5	124	293
Skin	24	18	73	5	2	163	285
Urogenital tract	58	49	5	38	1	74	225

1482

By project status

Breakdown by project status





	JCVI	Baylor	WashU	Broad	other	unassigned	total
Awaiting DNA	50	85	29	71	0	3	238
Complete	13	101	116	101	14	0	345
DNA Received	1	4	10	6	0	0	21
Draft	37	2	2	6	32	0	79
In Progress	21	30	158	95	19	0	323
Targeted	0	0	0	0	9	467	476
Total	122	222	315	279	74	470	1482

HMP Project Catalog

- Relational data model
- Tracks project status
- Stores comprehensive metadata
- Links to public data resources
- Provides search/filtering options

Human Microbiome Projects

Category	Count
Airways	145
Blood	54
Bone	2
Ear	2
Eye	3
Gastrointestinal tract	474
Heart	2
Lymph nodes	1
Oral	313
Skin	290
Spinal Cord	1
Urogenital tract	284
Wound	4
Unclassified	16
All Strains	1570

The **Human Microbiome Project (HMP) Catalog** records sequencing projects related to the NIH Human Microbiome Project.

Metadata collected for sequencing projects complies with the Genomic Standards Consortium MIGS/MIMS minimum information requirements.

The HMP Catalog is based on Genomes OnLine (GOLD) resource and the IMG-GOLD system for collecting genome and metagenome project information.

HMP Project Catalog

Search Field: is not empty is empty Quick Search

Advanced Search Export to Excel

HMP ID	Organism Name	Body Site	HMP Project Status	Finishing Goal	NCBI Project ID	NCBI Submission Status	Genbank ID	Gene Count	IMG/HMP ID	Sequencing Center	Funding Source	Strain Repository
0591	Mycobacterium parascrofulaceum ATCC BAA-614	Urogenital tract	Draft	Level 2: High-Quality Draft	31521	2				BCM-HGSC, USA	NIH-HMP Jumpstart Supplement	ATCC BAA-614
0592	Mycoplasma fermentans Edward ATCC 15474	Oral	Targeted			0				USA	NIH-HMP	ATCC 15474
0593	Mycoplasma hominis ATCC 23114	Gastrointestinal tract	Targeted			0				USA	NIH-HMP	ATCC 23114
0594	Mycoplasma hominis ATCC 14207	Gastrointestinal tract	Targeted			0				USA	NIH-HMP	ATCC 14207
0595	Neisseria cinerea ATCC 14685	Airways	Draft	Level 2: High-Quality Draft	30469	6	ACDY00000000	2191	643886151	Washington Univ, USA	NIH-HMP Jumpstart Supplement	ATCC 14685
0596	Neisseria elongata glycolytica ATCC 29315	Airways	Draft	Level 2: High-Quality Draft	30471	4	ADBF00000000			Washington Univ, USA	NIH-HMP Jumpstart Supplement	ATCC 29315
0597	Neisseria elongata glycolytica	Airways	Targeted			0				USA	NIH-HMP	
0599	Neisseria flavescens NRL30031/H210	Airways	Draft	Level 2: High-Quality	30473	6	ACEN00000000	2595	643886198	Washington Univ, USA	NIH-HMP Jumpstart	

Count: 1294 HMP Master List

Contains a complete list of all Reference Strains along with detailed metadata about each. Provides both “quick” and “advanced” search and download options.

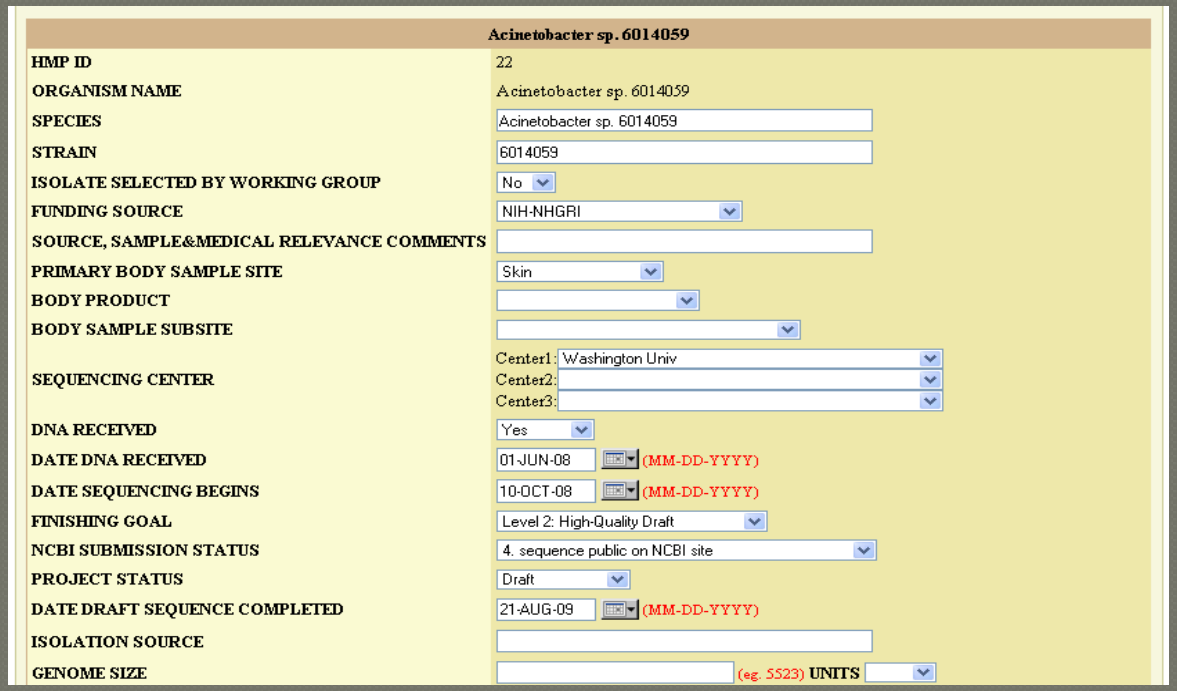
Reference Genomes: MIGS Compliance

HMP 0022 (Acinetobacter sp. 6014059)

MIGS-ID	Organism Info	
MIGS 3 (*)	Organism Name	Acinetobacter sp. 6014059
	NCBI Taxon ID	525242
	NCBI Kingdom	Bacteria
	NCBI Phylum	Proteobacteria
	NCBI Class	Gammaproteobacteria
	NCBI Order	Pseudomonadales
	NCBI Family	Moraxellaceae
	NCBI Genus	Acinetobacter
	NCBI Species	Acinetobacter sp. 6014059
MIGS 2 (*)	Domain	BACTERIAL
	Phylogeny	PROTEOBACTERIA-GAMMA
	Genus	Acinetobacter
	Species	sp.
	Strain	6014059
MIGS 13 (*)	Strain Repository	NCTC
MIGS-ID	HMP Metadata	
	HMP ID	0022
	HMP Project Status	Draft
	BEI Status	Not Yet Available
	NCBI Submission Status	4. sequence public on NCBI site
	Isolate Selected by Working Group	No
	Finishing Goal	Level 2: High-Quality Draft
	DNA Received	Yes
	Date DNA Received (DD-MON-YY)	01-JUN-08
	Date Sequencing Begins (DD-MON-YY)	10-OCT-08
	Date Draft Sequencing Completed (DD-MON-YY)	21-AUG-09
MIGS-ID	Project Info	

MIGS-ID	Project Info	
	ER Submission Project OID	14024
	GOLD Stamp ID	Gi03496
	GCAT ID	004282_GCAT
	Greengenes ID	6014059
MIGS 1.1 (*)	NCBI Project ID	33073
	GOLD Web Page Code	1
	Project Type	Genome-Isolate
	Availability	Public
	Contact Name	WashU
	Contact Email	hmpstrainswashu@watson.wustl.edu
	Funding Program	NIH-NHGRI
	IMG Contact	hhcreasy (hhuot@som.umaryland.edu)
	Add Date	05-OCT-08
	Last Modify Date	14-DEC-09
	Last Modified By	dinos007 (dinos007@yahoo.com)
	Project Relevance	Human Microbiome Project (HMP), Medical
MIGS 32; MIGS 33	Data Links (URLs)	Data, GenBank, ACYS01000000, URL Funding, NIH, , URL Information, Entrez, 33073, URL Information, HMP, , URL Information, Taxonomy, 525242, URL Seq Center, Washington Univ, , URL

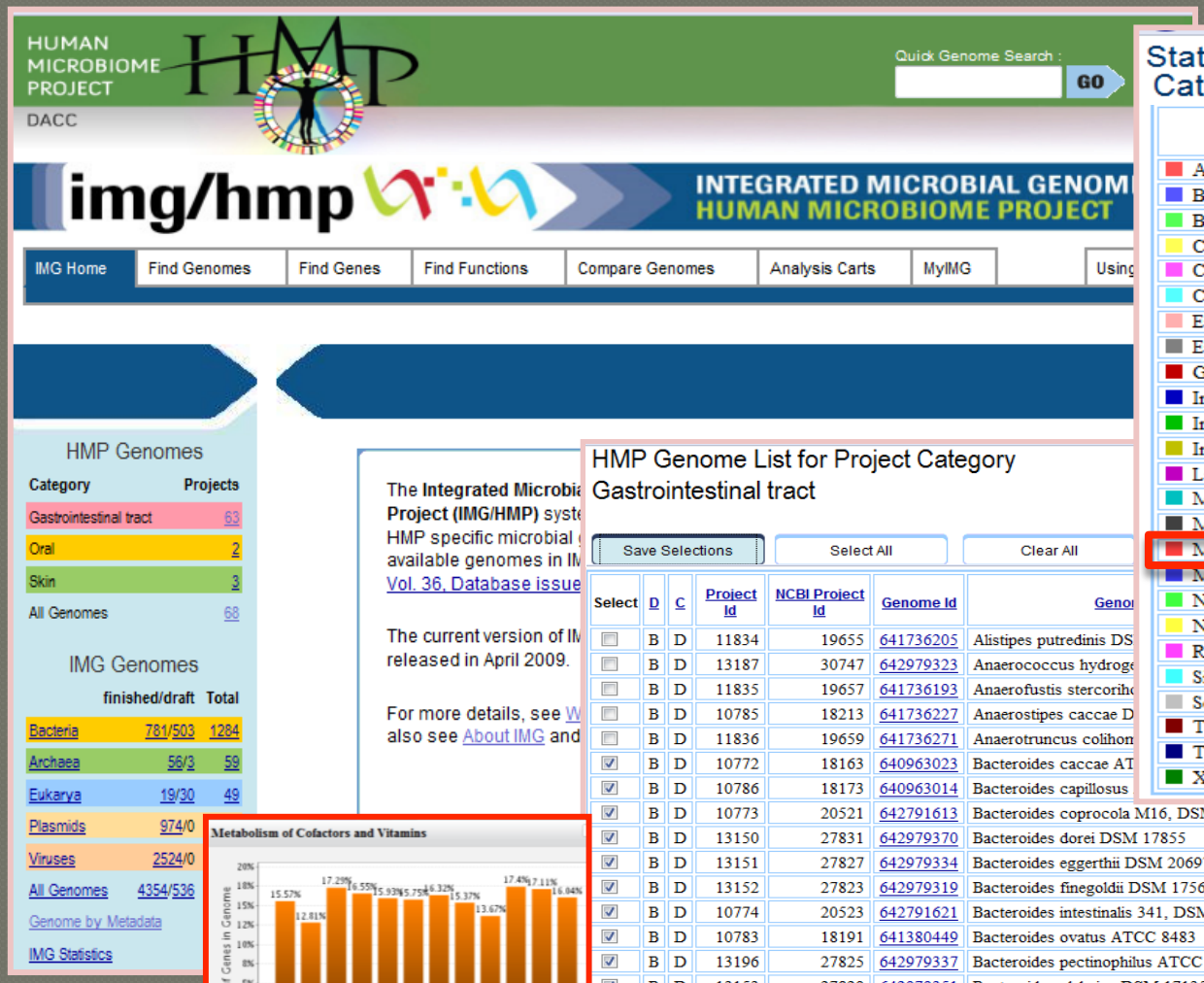
DACC Management Web Interface

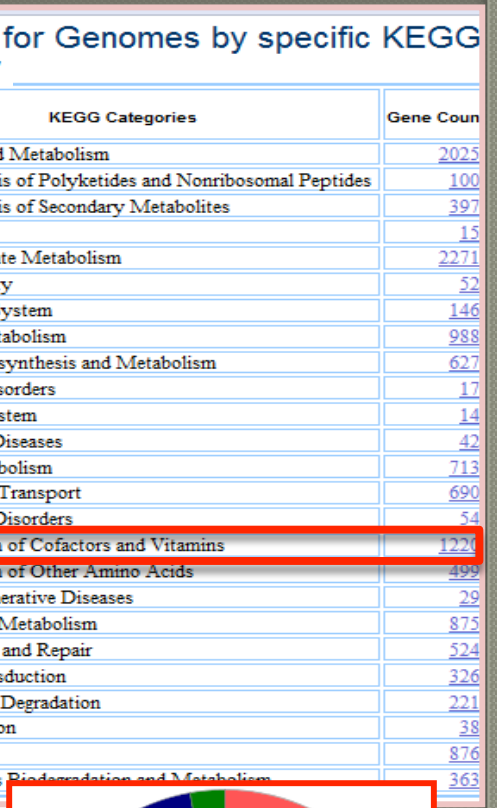


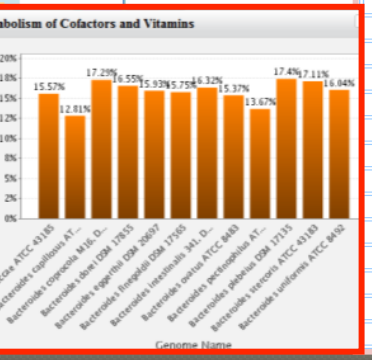
Acinetobacter sp. 6014059	
HMP ID	22
ORGANISM NAME	Acinetobacter sp. 6014059
SPECIES	Acinetobacter sp. 6014059
STRAIN	6014059
ISOLATE SELECTED BY WORKING GROUP	No
FUNDING SOURCE	NIH-NHGRI
SOURCE, SAMPLE&MEDICAL RELEVANCE COMMENTS	
PRIMARY BODY SAMPLE SITE	Skin
BODY PRODUCT	
BODY SAMPLE SUBSITE	
SEQUENCING CENTER	Center1: Washington Univ Center2: Center3:
DNA RECEIVED	Yes
DATE DNA RECEIVED	01-JUN-08 (MM-DD-YYYY)
DATE SEQUENCING BEGINS	10-OCT-08 (MM-DD-YYYY)
FINISHING GOAL	Level 2: High-Quality Draft
NCBI SUBMISSION STATUS	4. sequence public on NCBI site
PROJECT STATUS	Draft
DATE DRAFT SEQUENCE COMPLETED	21-AUG-09 (MM-DD-YYYY)
ISOLATION SOURCE	
GENOME SIZE	(eg. 5523) UNITS

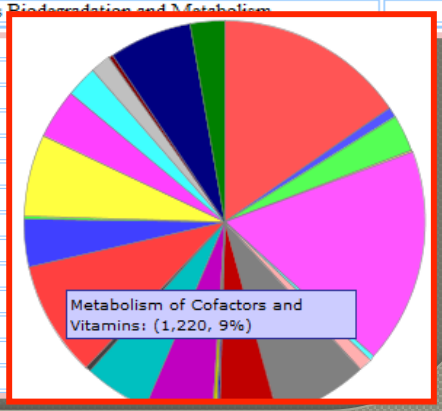
- Enforces the population of required fields
- Restricts contents of fields with controlled vocabularies
- Provides both individual and bulk update options
- Followed by QC steps prior to incorporation into the Catalog

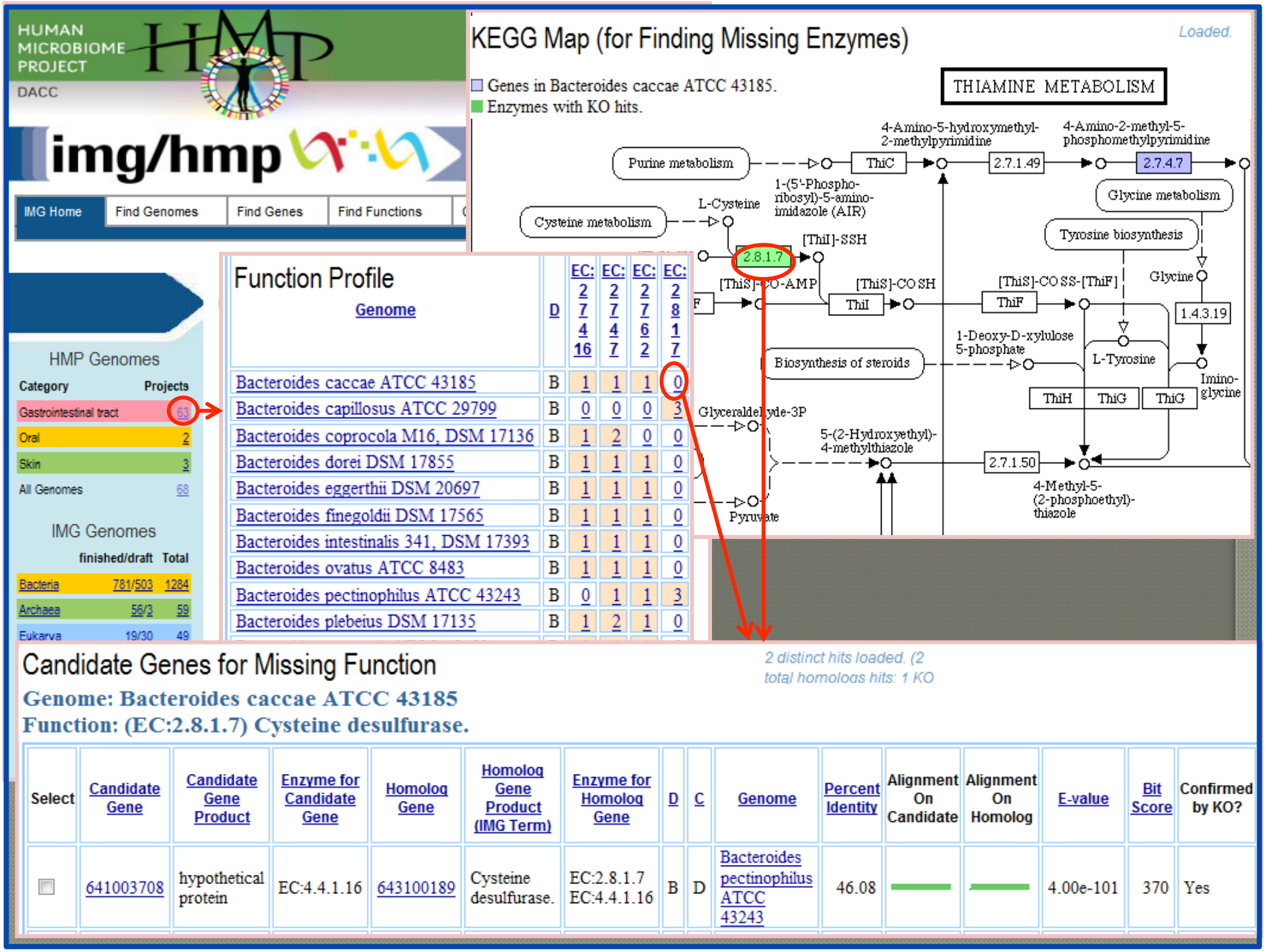
Genome Analysis at IMG







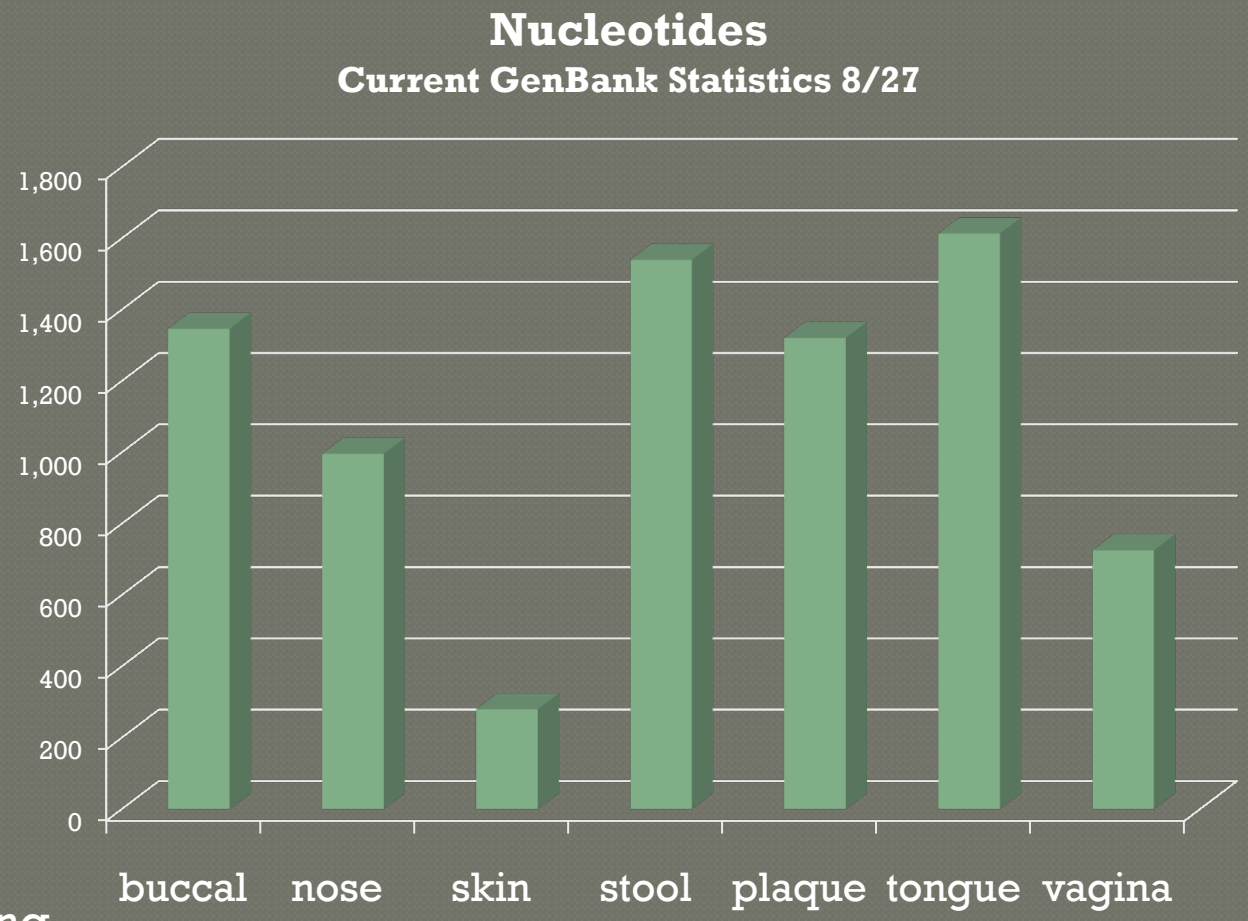



HUMAN MICROBIOME PROJECT
DACC
HMP
img/hmp
IMG Home
Find Genomes
Find Genes
Find Functions
KEGG Map (for Finding Missing Enzymes)
Loaded.
Genes in Bacteroides caccae ATCC 43185.
Enzymes with KO hits.
THIAMINE METABOLISM
Purine metabolism
ThiC
4-Amino-5-hydroxymethyl-2-methylpyrimidine
2.7.1.49
4-Amino-2-methyl-5-phosphomethylpyrimidine
2.7.4.7
Glycine metabolism
Tyrosine biosynthesis
Cysteine metabolism
L-Cysteine
1-(5'-Phospho-ribosyl)-5-amino-imidazole (AIR)
[ThiI]-SSH
2.8.1.7
[ThiS]-CO-AMP
[ThiS]-COSH
ThiI
[ThiS]-COSS-[ThiF]
ThiF
Glycine
1.4.3.19
1-Deoxy-D-xylulose 5-phosphate
L-Tyrosine
Biosynthesis of steroids
Imino-glycine
ThiH
ThiG
ThiG
Glyceraldehyde-3P
5-(2-Hydroxyethyl)-4-methylthiazole
2.7.1.50
4-Methyl-5-(2-phosphoethyl)-thiazole
Pyruvate
HMP Genomes
Category
Projects
Gastrointestinal tract 63
Oral 2
Skin 3
All Genomes 68
IMG Genomes
finished/draft
Total
Bacteria 781/503 1284
Archaea 56/3 59
Eukarya 19/30 49
Function Profile
Genome
D
EC: 2.7.4.16
EC: 2.7.4.7
EC: 2.7.6.2
EC: 2.8.1.7
Bacteroides caccae ATCC 43185 B 1 1 1 0
Bacteroides capillosus ATCC 29799 B 0 0 0 3
Bacteroides coprocola M16, DSM 17136 B 1 2 0 0
Bacteroides dorei DSM 17855 B 1 1 1 0
Bacteroides eggerthii DSM 20697 B 1 1 1 0
Bacteroides finegoldii DSM 17565 B 1 1 1 0
Bacteroides intestinalis 341, DSM 17393 B 1 1 1 0
Bacteroides ovatus ATCC 8483 B 1 1 1 0
Bacteroides pectinophilus ATCC 43243 B 0 1 1 3
Bacteroides plebeius DSM 17135 B 1 2 1 0
2 distinct hits loaded. (2 total homologs hits: 1 KO
Candidate Genes for Missing Function
Genome: Bacteroides caccae ATCC 43185
Function: (EC:2.8.1.7) Cysteine desulfurase.
Select
Candidate Gene
Candidate Gene Product
Enzyme for Candidate Gene
Homolog Gene
Homolog Gene Product (IMG Term)
Enzyme for Homolog Gene
D
C
Genome
Percent Identity
Alignment On Candidate
Alignment On Homolog
E-value
Bit Score
Confirmed by KO?
641003708
hypothetical protein
EC:4.4.1.16
643100189
Cysteine desulfurase.
EC:2.8.1.7 EC:4.4.1.16
B
D
Bacteroides pectinophilus ATCC 43243
46.08
4.00e-101
370
Yes

Microbiome Sequencing

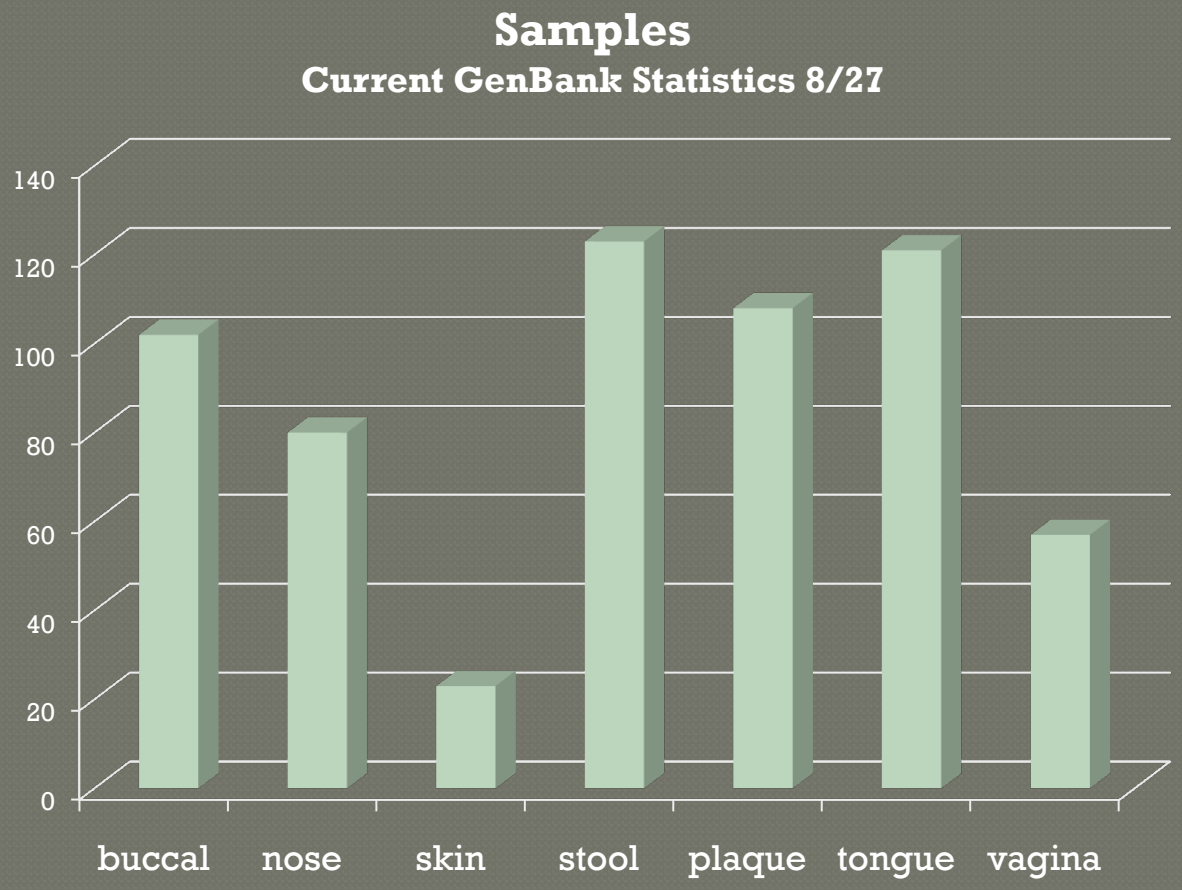
	16S rRNA Sequencing				Metagenomic wgs Sequencing			
Study	Sequencing Center	16S NCBI PID	16S SRA Study ID	16S Trace Archives	Sequencing Center	wgs NCBI PID	wgs SRA Study ID	dbGap Study accession
Production Phase I	BCM, BI, JCVI, WashU	48333	SRP002395		BCM, BI, JCVI, WashU	48479	SRP002163	phs000228
	Description: Pyrosequencing of clinical samples collected from multiple body sites from hundreds of subjects. Metagenomic sequencing represents a subset of the subjects for which 16S sequencing was performed.							
Clinical Production Pilot	BCM, BI, JCVI, WashU	48335	SRP002012					phs000228
	Description: Pyrosequencing of clinical samples representing multiple body sites from a small set of subjects, using the XLR SOP v4.2 16S Sequencing protocol Each sample was sent to two sequencing centers for quantification of variability among replicates, and to gain a preliminary understanding of the variability of community structures across different body sites. DACC Clinical Pilot Production Study (PPS) 16S data download page							
Sanger Clinical Production Pilot	BCM, BI, JCVI, WashU	34129		Link to TA				phs000228
	Description: Dideoxy seqeuncing of clinical samples representing multiple body sites from a small set of subjects. Each sample was sent to two sequencing centers for quantification of variability among replicates, and to gain a preliminary understanding of the variability of community structures across different body sites.							
CEFoS Clinical Pilot	BCM, BI, JCVI, WashU	48339	SRP002396					phs000228
	Description: Pyrosequencing of a single clinical stool sample, using the XLR SOP v4.2 16S Sequencing protocol							
Pre-CEFoS Clinical Pilot	BI	48467	SRP002440					phs000228
	Description: Pyrosequencing of a single clinical stool sample, using an early common 454 16S sequencing protocol no longer in use							
CEFoS Mock Pilot	BCM, BI, JCVI, WashU	48341	SRP002397					
	Description: Pyrosequencing of HMP even & staggered mock community samples, using the XLR SOP v4.2 16S Sequencing protocol							
Pre-CEFoS Mock Pilot	BCM, BI, JCVI, WashU	48465	SRP002443					
	Description: Pyrosequencing of HMP even & staggered mock community samples, using an early common 454 16S sequencing protocol no longer in use							

Metagenomic WGS Data

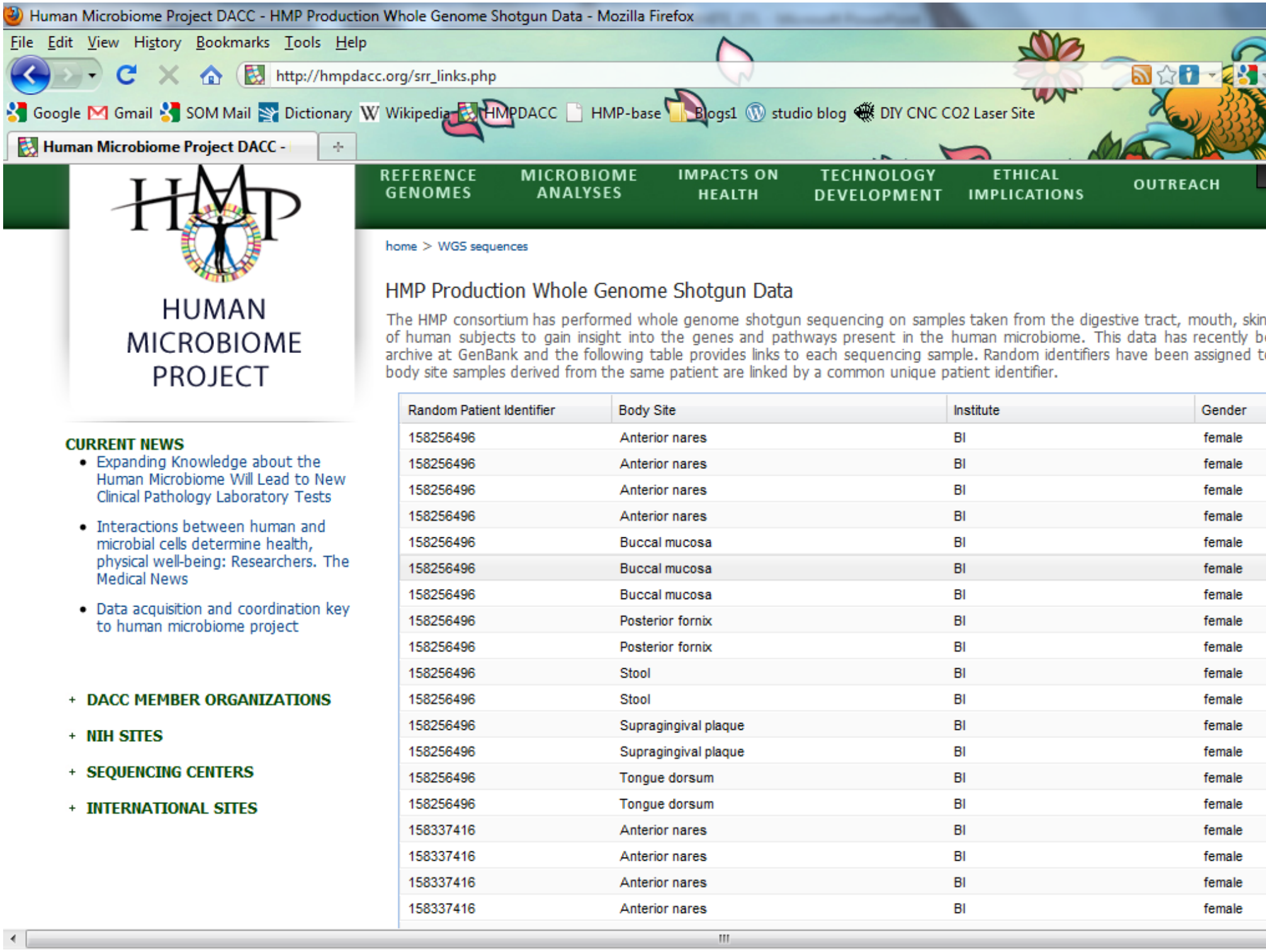


Sarah Young
John Martin
HMP Data Processing Working Group

Metagenomic WGS Data



Sarah Young
John Martin
HMP Data Processing Working Group



Random Patient Identifier	Body Site	Institute	Gender
158256496	Anterior nares	BI	female
158256496	Anterior nares	BI	female
158256496	Anterior nares	BI	female
158256496	Anterior nares	BI	female
158256496	Buccal mucosa	BI	female
158256496	Buccal mucosa	BI	female
158256496	Buccal mucosa	BI	female
158256496	Posterior fornix	BI	female
158256496	Posterior fornix	BI	female
158256496	Stool	BI	female
158256496	Stool	BI	female
158256496	Supragingival plaque	BI	female
158256496	Supragingival plaque	BI	female
158256496	Tongue dorsum	BI	female
158256496	Tongue dorsum	BI	female
158337416	Anterior nares	BI	female
158337416	Anterior nares	BI	female
158337416	Anterior nares	BI	female
158337416	Anterior nares	BI	female

This coming year

img/mhmp

IMG/M Home | Find Genomes

Reference context for metagenome analysis

IMG/M Genomes

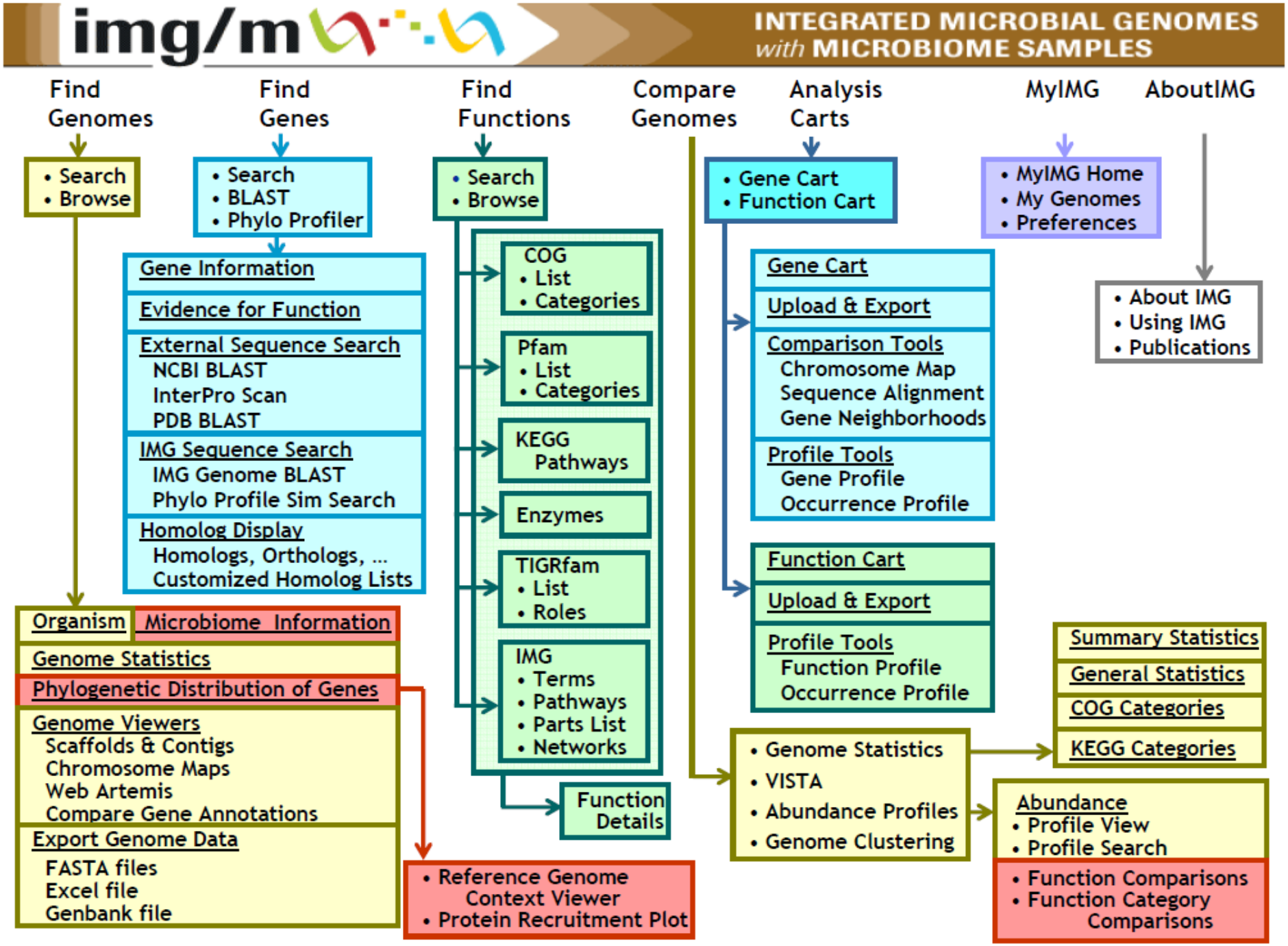
	finished/draft	Total
Bacteria	806/591	1397
Archaea	59/14	73
Eukarya	19/30	49
Plasmids	974/0	974
Viruses	2524/0	2524
Microbiomes	0/283	283
All Genomes	4382/918	5300
User Private	victor	226

Microbiome Projects Map

Hands on training available at the Microbial Genomics & Metagenomics Workshop

01 *Microbiome All None
02 Endobiotic All None
03 Animals All None
04 Gastrointestinal tract All None
05 Intestinal microbiome of Mouse lean and obese All None
06 Sample All None
08 ☑ Mouse Gut Community lean1 [D]
08 ☑ Mouse Gut Community lean2 [D]
08 ☑ Mouse Gut Community lean3 [D]
08 ☑ Mouse Gut Community ob1 [D]
08 ☑ Mouse Gut Community ob2 [D]
03 Human All None
04 Gastrointestinal All None
05 Fecal microbiome of Human from Obese and Lean Twins All None
06 Sample All None
08 ☑ Human distal gut (mom, family 1, overweight, TS3) [D]
08 ☑ Human distal gut (mom, family 2, obese, TS6) [D]
08 ☑ Human distal gut (mom, family 3, overweight, TS9) [D]
08 ☑ Human distal gut (mom, family 4, obese, TS21) [D]
08 ☑ Human distal gut (mom, family 5, overweight, TS30) [D]
08 ☑ Human distal gut (twin, family 1, lean, TS1) [D]
08 ☑ Human distal gut (twin, family 1, lean, TS2) [D]
08 ☑ Human distal gut (twin, family 2, lean, TS4) [D]
08 ☑ Human distal gut (twin, family 2, lean, TS5) [D]
08 ☑ Human distal gut (twin, family 3, lean, TS7) [D]
08 ☑ Human distal gut (twin, family 3, lean, TS8) [D]
08 ☑ Human distal gut (twin, family 4, obese, TS19) [D]
08 ☑ Human distal gut (twin, family 4, obese, TS20) [D]
08 ☑ Human distal gut (twin, family 4, obese, TS49) [D]
08 ☑ Human distal gut (twin, family 4, obese, TS50) [D]
08 ☑ Human distal gut (twin, family 5, obese, TS28) [D]
08 ☑ Human distal gut (twin, family 5, obese, TS29) [D]
08 ☑ Human distal gut (twin, family 5, obese, TS51) [D]

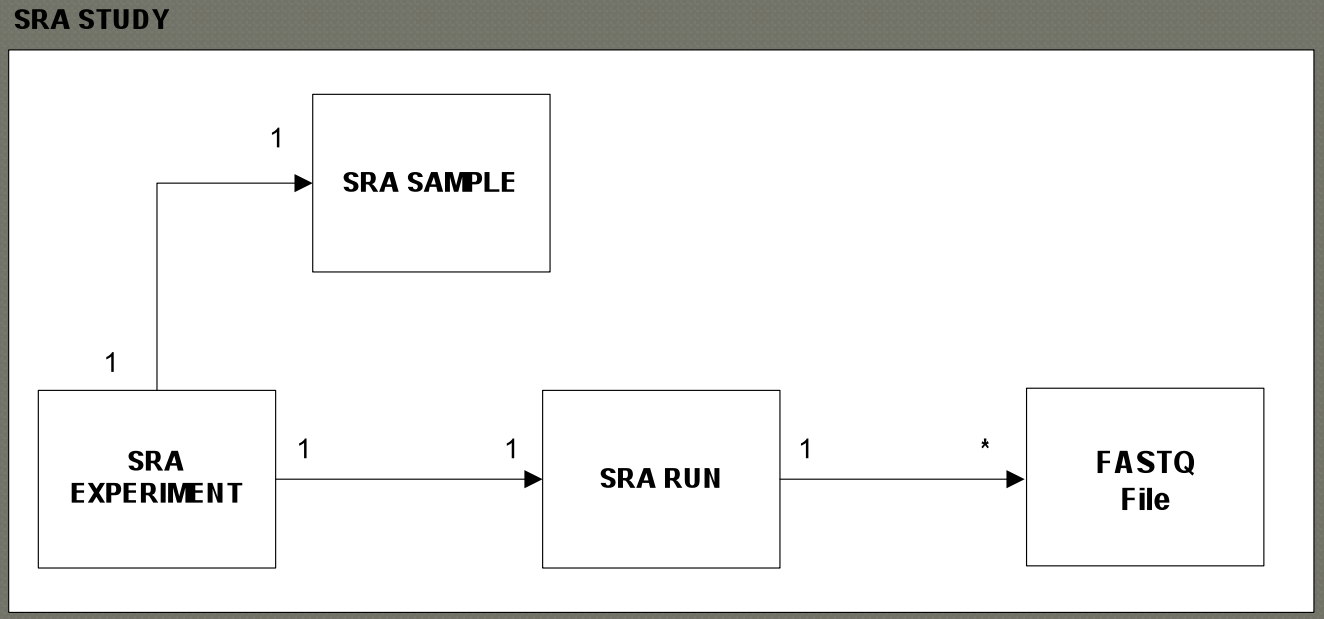
JGI: Victor Markowitz
Nikos Kyrpides


img/m
INTEGRATED MICROBIAL GENOMES
with MICROBIOME SAMPLES
Find Genomes
• Search
• Browse
Organism
Microbiome Information
Genome Statistics
Phylogenetic Distribution of Genes
Genome Viewers
Scaffolds & Contigs
Chromosome Maps
Web Artemis
Compare Gene Annotations
Export Genome Data
FASTA files
Excel file
Genbank file
Find Genes
• Search
• BLAST
• Phylo Profiler
Gene Information
Evidence for Function
External Sequence Search
NCBI BLAST
InterPro Scan
PDB BLAST
IMG Sequence Search
IMG Genome BLAST
Phylo Profile Sim Search
Homolog Display
Homologs, Orthologs, ...
Customized Homolog Lists
• Reference Genome Context Viewer
• Protein Recruitment Plot
Find Functions
• Search
• Browse
COG
• List
• Categories
Pfam
• List
• Categories
KEGG Pathways
Enzymes
TIGRfam
• List
• Roles
IMG
• Terms
• Pathways
• Parts List
• Networks
Function Details
Compare Genomes
• Genome Statistics
• VISTA
• Abundance Profiles
• Genome Clustering
Abundance
• Profile View
• Profile Search
• Function Comparisons
• Function Category Comparisons
Analysis Carts
• Gene Cart
• Function Cart
Gene Cart
Upload & Export
Comparison Tools
Chromosome Map
Sequence Alignment
Gene Neighborhoods
Profile Tools
Gene Profile
Occurrence Profile
Function Cart
Upload & Export
Profile Tools
Function Profile
Occurrence Profile
Summary Statistics
General Statistics
COG Categories
KEGG Categories
MyIMG
• MyIMG Home
• My Genomes
• Preferences
AboutIMG
• About IMG
• Using IMG
• Publications

WGS submission to NCBI

• Centers and DACC are working with NCBI to use common schema and relevant metadata.

•Submission guide, usage Aspera client, usage of QIIME available



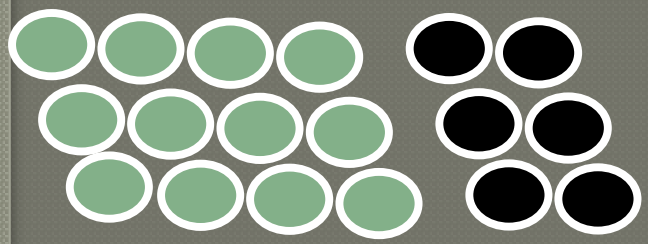
NCBI Projects

HMP Top Level
(43021)

Reference Genome Top Level
(28331)

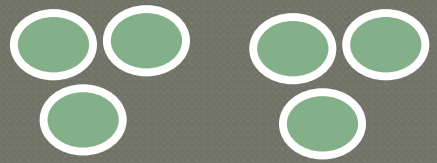
Characterizing microbiome
of healthy individuals
Source = HMP Centers

Associating microbiome
with disease
Source = Demo Projects
(46305)



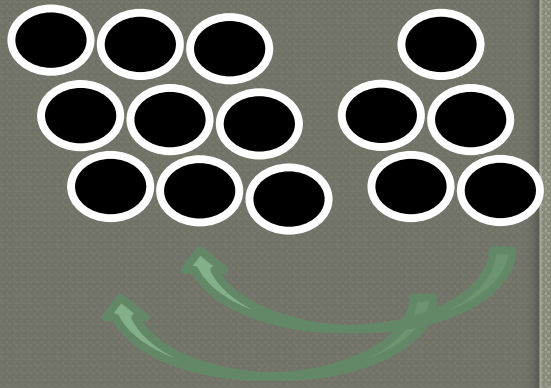
16S
(48489)

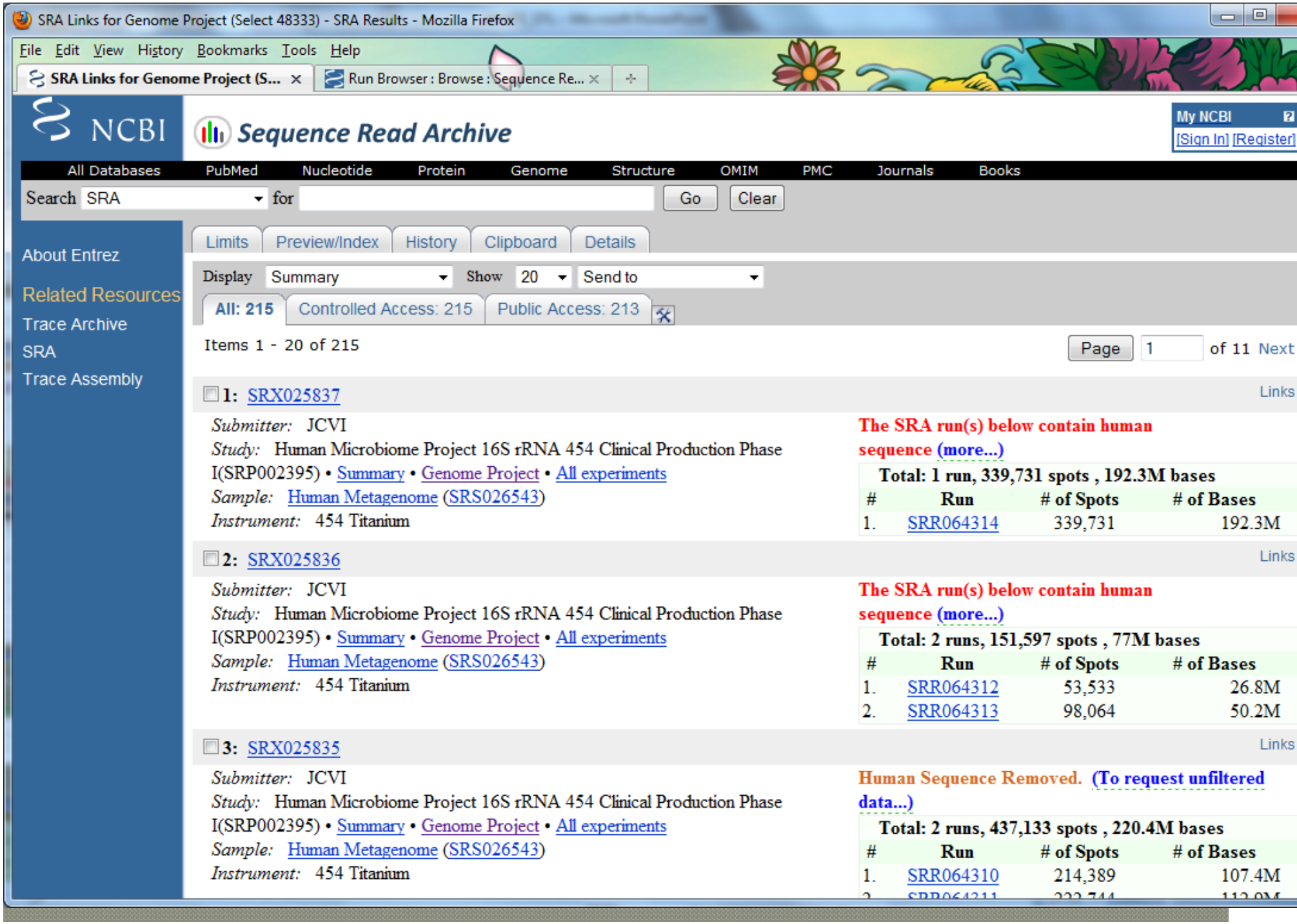
WGS
(43017)



16S

WGS




SRA Links for Genome Project (Select 48333) - SRA Results - Mozilla Firefox
File Edit View History Bookmarks Tools Help
SRA Links for Genome Project (S...
Run Browser : Browse : Sequence Re...
NCBI
Sequence Read Archive
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search SRA for Go Clear
About Entrez
Related Resources
Trace Archive
SRA
Trace Assembly
Limits Preview/Index History Clipboard Details
Display Summary Show 20 Send to
All: 215 Controlled Access: 215 Public Access: 213
Items 1 - 20 of 215
Page 1 of 11 Next
1: SRX025837
Links
Submitter: JCVI
Study: Human Microbiome Project 16S rRNA 454 Clinical Production Phase I(SRP002395) • Summary • Genome Project • All experiments
Sample: Human Metagenome (SRS026543)
Instrument: 454 Titanium
The SRA run(s) below contain human sequence (more...)
Total: 1 run, 339,731 spots , 192.3M bases
Run # of Spots # of Bases
1. SRR064314 339,731 192.3M
2: SRX025836
Links
Submitter: JCVI
Study: Human Microbiome Project 16S rRNA 454 Clinical Production Phase I(SRP002395) • Summary • Genome Project • All experiments
Sample: Human Metagenome (SRS026543)
Instrument: 454 Titanium
The SRA run(s) below contain human sequence (more...)
Total: 2 runs, 151,597 spots , 77M bases
Run # of Spots # of Bases
1. SRR064312 53,533 26.8M
2. SRR064313 98,064 50.2M
3: SRX025835
Links
Submitter: JCVI
Study: Human Microbiome Project 16S rRNA 454 Clinical Production Phase I(SRP002395) • Summary • Genome Project • All experiments
Sample: Human Metagenome (SRS026543)
Instrument: 454 Titanium
Human Sequence Removed. (To request unfiltered data...)
Total: 2 runs, 437,133 spots , 220.4M bases
Run # of Spots # of Bases
1. SRR064310 214,389 107.4M

NCBI

dbGaP GENOTYPES and PHENOTYPES

NIH Human Microbiome Project - Core Microbiome Sampling Protocol A (HMP-A)

Study Accession: phs000228.v1.p1

Study | Variables | Documents | Analyses | Datasets

Study Description

This first clinical study of the Human Microbiome Project (HMP) addresses whether individuals share a core human microbiome. It involves broad determination of the microbiota found in five anatomical sites: the oral cavity, skin, nasal cavity, gastrointestinal tract and vagina. This study will enroll approximately 250 healthy male and female adults, 18-40 years old, from two geographic regions of the US: Houston, TX and St. Louis, MO. The participation of healthy individuals will create a baseline for discovery of the core microbiota typically found in various areas of the human body. The information from this initial study can then be used to help assess the changes in the complement of microbiota found on or within diseased individuals.

- Study Type: Population-Based Control Set
- Number of participants in study:
  - 0 phenotyped subjects

Authorized Access Data

Individual level data will be coming soon

Publicly Available Data (Public ftp)

Estimated availability to be determined

Study Inclusion/Exclusion Criteria

Inclusion Criteria:

In order to be eligible for participation in this study, subjects must meet the following criteria:

- Male or female subjects 18 years of age, but not more than 40 years of age at the time of enrollment.
- Must be able to provide signed and dated informed consent.
- Healthy subjects willing and able to provide blood, as well as oral cavity, skin, nasal cavity and stool specimens; female subjects must be willing to provide a vaginal specimen and must either have regular menstrual cycles (between 21 and 35 days) or, for subjects on hormonal contraception influencing cycle length, have a history of regular 21 to 35 day menstrual cycles prior to initiating hormonal contraception. At study enrollment, female subjects may be using any contraception method except a combination hormone vaginal ring (see Exclusion Criteria).

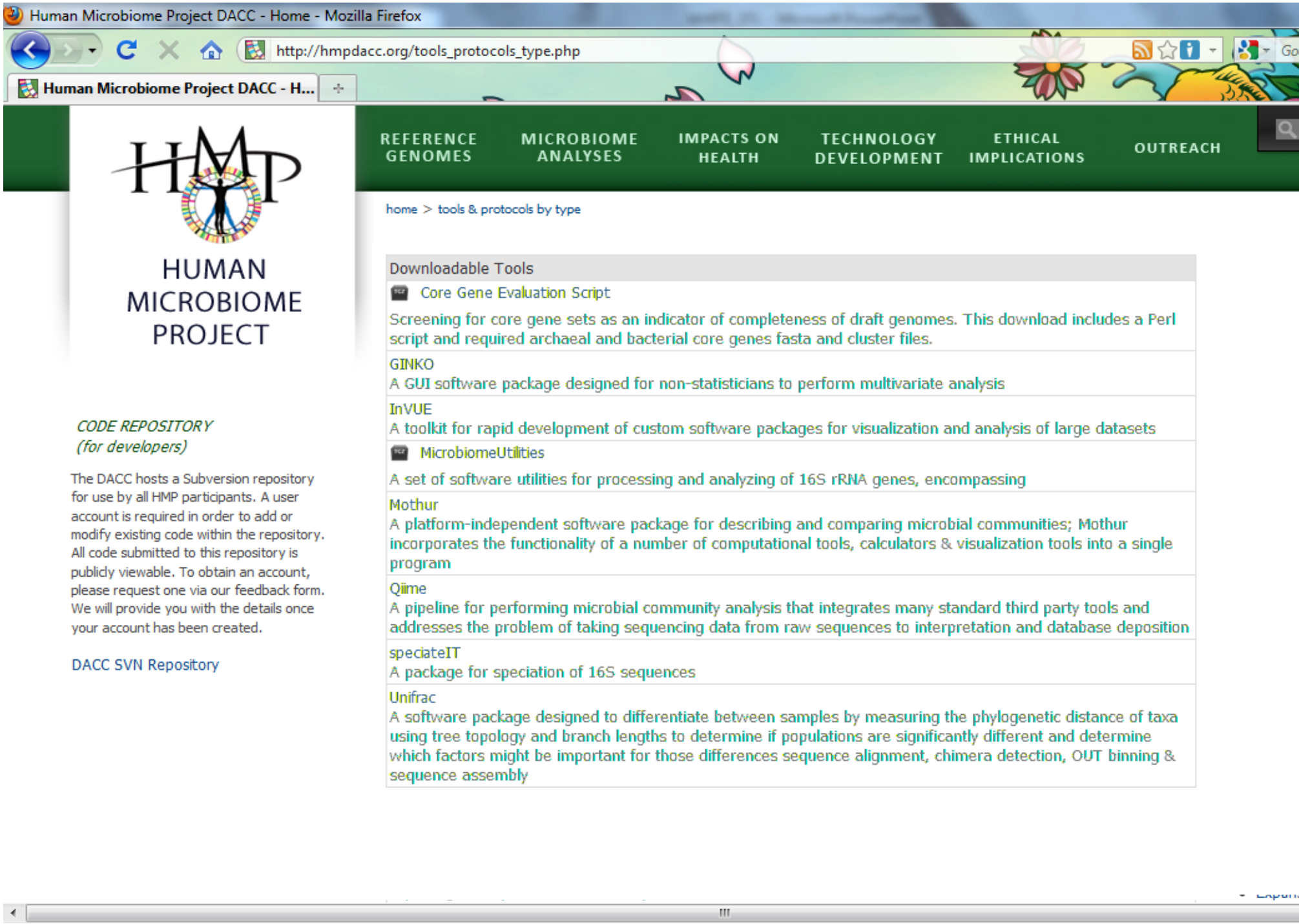
HMP-Wide Patient Phenotype

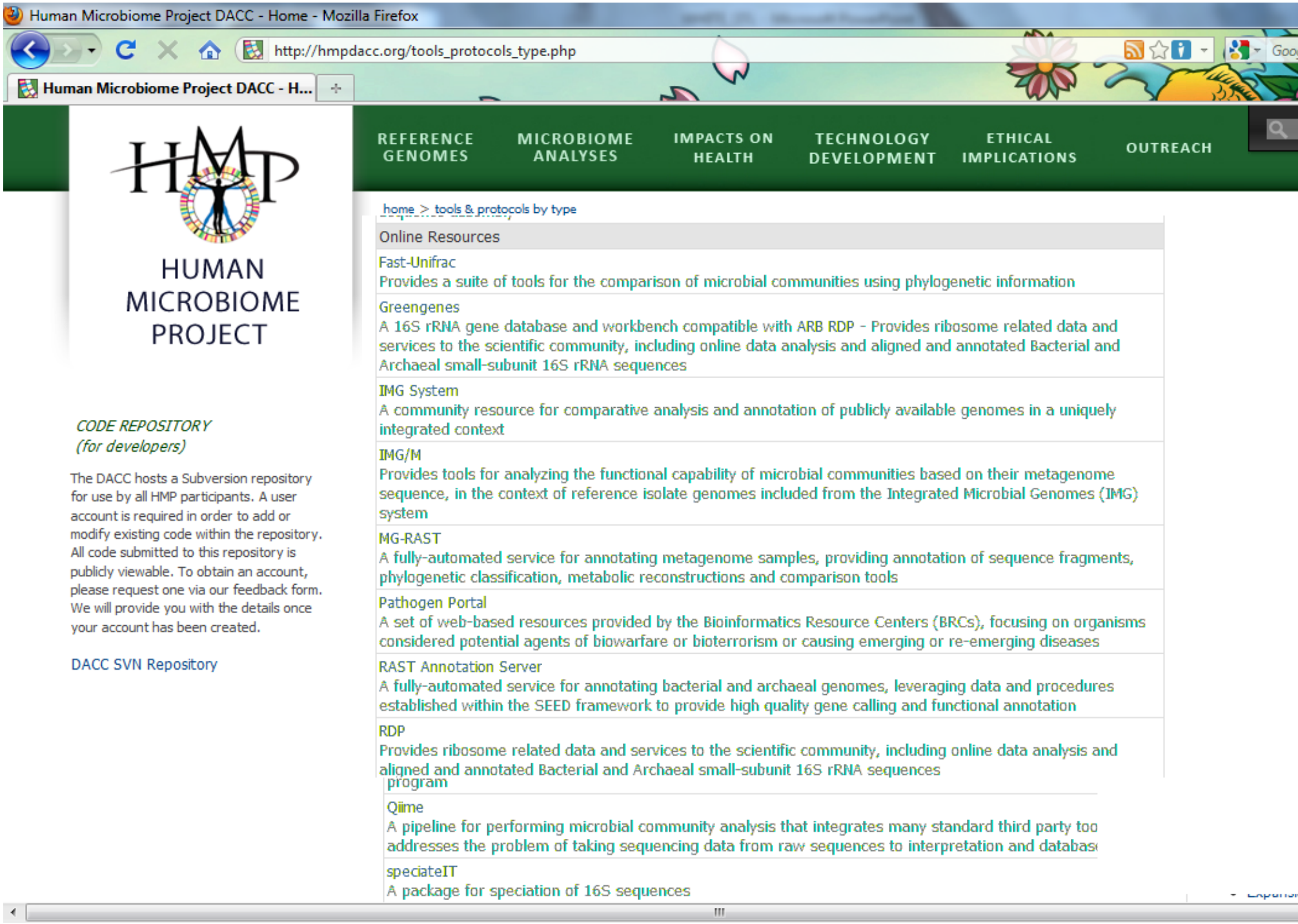
IHMC Variable	Total	Fraction Identical	Mappable	Not mappable	Not present	P1	P2	… PN
SUBJID	1.00	0.88	0.13			SUBJID	SUBJID	
Gender	0.94	0.94			0.06	Gender		
Age	0.88	0.81	0.06	0.06	0.06	Age_at_first_visit	AgeAtEnrollment	
Race	0.81	0.44	0.38		0.19	Race	Race_Other_Text	
Other Rrace	0.56	0.31	0.25		0.44	Other Race	Race_Other	
Smoking	0.38	0.31	0.06		0.63	Smoking_status		
Lab	0.31	0.19	0.13	0.06	0.63	Diagnosis	TID	
Smoking_duration	0.31	0.19	0.13		0.69	Smoking_status		
Drugs	0.31	0.19	0.13		0.69	Antacids, Steroids, Antibiotics		
Weight_kg	0.25	0.25		0.06	0.69			
BP	0.19	0.19			0.81			
Height	0.19	0.19			0.81			
Disease	0.19	0.06	0.13		0.81			
Institution	0.13	0.00	0.13		0.88			
Dose	0.13	0.06	0.06		0.88			
Duration	0.13	0.06	0.06		0.88			
Start_date	0.13	0.13			0.88		TID	
Finish_date	0.13	0.13			0.88		TID	
Location	0.13	0.13		0.06	0.81	Other Country		
Drug_name	0.06		0.06	0.06	0.88			
HIV/AIDS	0.00				1.00			

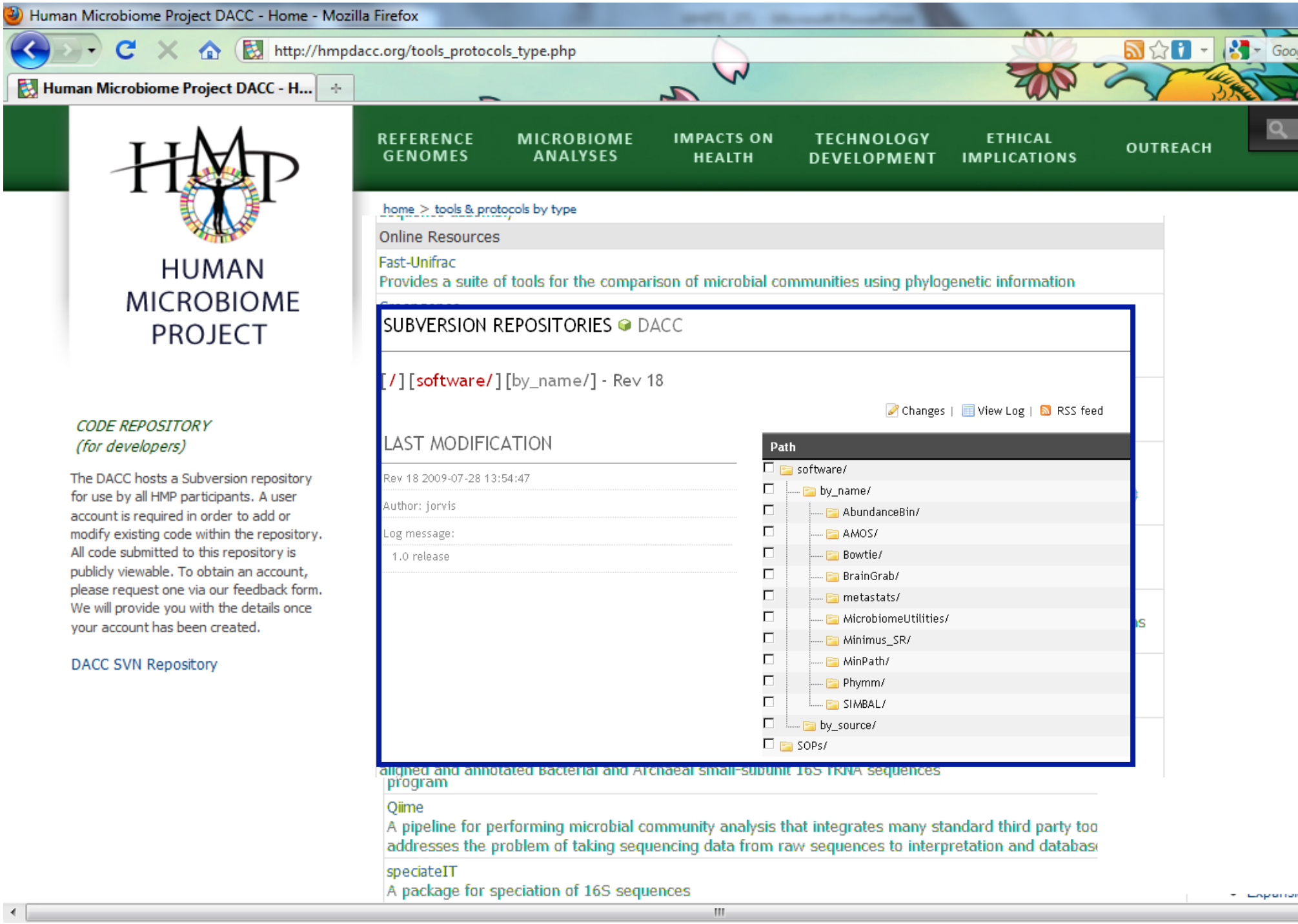
Current data availability status for WGS

Dirk Gevers & Ashlee Earl
Broad Institute

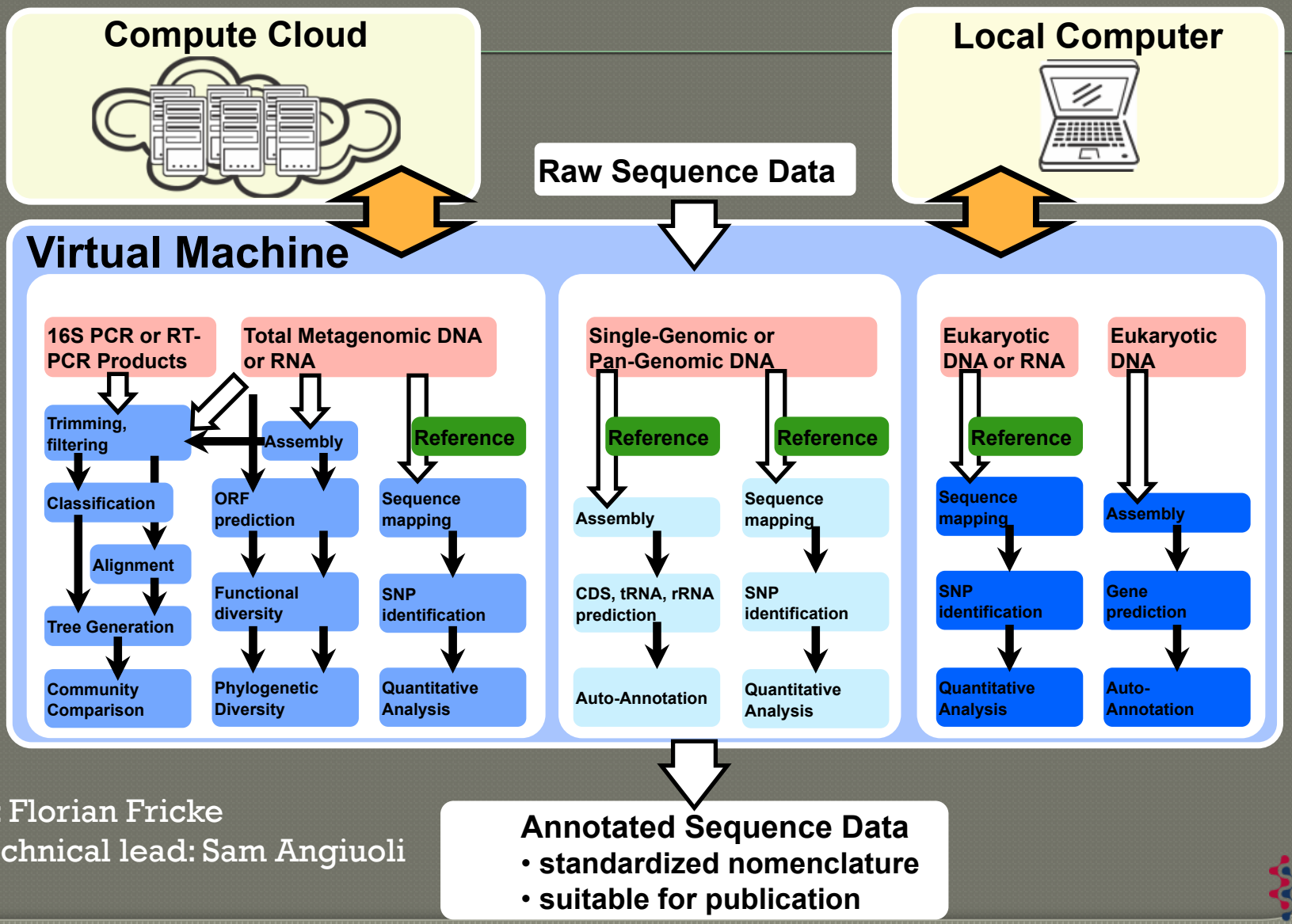
HUMAN MICROBIOME PROJECT HMP


Human Microbiome Project DACC - Home - Mozilla Firefox
http://hmpdacc.org/tools_protocols_type.php
Human Microbiome Project DACC - H...
REFERENCE GENOMES
MICROBIOME ANALYSES
IMPACTS ON HEALTH
TECHNOLOGY DEVELOPMENT
ETHICAL IMPLICATIONS
OUTREACH
HUMAN MICROBIOME PROJECT
home > tools & protocols by type
CODE REPOSITORY
(for developers)
The DACC hosts a Subversion repository for use by all HMP participants. A user account is required in order to add or modify existing code within the repository. All code submitted to this repository is publicly viewable. To obtain an account, please request one via our feedback form. We will provide you with the details once your account has been created.
DACC SVN Repository
Downloadable Tools
Core Gene Evaluation Script
Screening for core gene sets as an indicator of completeness of draft genomes. This download includes a Perl script and required archaeal and bacterial core genes fasta and cluster files.
GINKO
A GUI software package designed for non-statisticians to perform multivariate analysis
InVUE
A toolkit for rapid development of custom software packages for visualization and analysis of large datasets
MicrobiomeUtilities
A set of software utilities for processing and analyzing of 16S rRNA genes, encompassing
Mothur
A platform-independent software package for describing and comparing microbial communities; Mothur incorporates the functionality of a number of computational tools, calculators & visualization tools into a single program
Qiime
A pipeline for performing microbial community analysis that integrates many standard third party tools and addresses the problem of taking sequencing data from raw sequences to interpretation and database deposition
speciateIT
A package for speciation of 16S sequences
Unifrac
A software package designed to differentiate between samples by measuring the phylogenetic distance of taxa using tree topology and branch lengths to determine if populations are significantly different and determine which factors might be important for those differences sequence alignment, chimera detection, OUT binning & sequence assembly


Human Microbiome Project DACC - Home - Mozilla Firefox
http://hmpdacc.org/tools_protocols_type.php
Human Microbiome Project DACC - H...
REFERENCE GENOMES
MICROBIOME ANALYSES
IMPACTS ON HEALTH
TECHNOLOGY DEVELOPMENT
ETHICAL IMPLICATIONS
OUTREACH
HUMAN MICROBIOME PROJECT
home > tools & protocols by type
Online Resources
Fast-Unifrac
Provides a suite of tools for the comparison of microbial communities using phylogenetic information
Greengenes
A 16S rRNA gene database and workbench compatible with ARB RDP - Provides ribosome related data and services to the scientific community, including online data analysis and aligned and annotated Bacterial and Archaeal small-subunit 16S rRNA sequences
IMG System
A community resource for comparative analysis and annotation of publicly available genomes in a uniquely integrated context
IMG/M
Provides tools for analyzing the functional capability of microbial communities based on their metagenome sequence, in the context of reference isolate genomes included from the Integrated Microbial Genomes (IMG) system
MG-RAST
A fully-automated service for annotating metagenome samples, providing annotation of sequence fragments, phylogenetic classification, metabolic reconstructions and comparison tools
Pathogen Portal
A set of web-based resources provided by the Bioinformatics Resource Centers (BRCs), focusing on organisms considered potential agents of biowarfare or bioterrorism or causing emerging or re-emerging diseases
RAST Annotation Server
A fully-automated service for annotating bacterial and archaeal genomes, leveraging data and procedures established within the SEED framework to provide high quality gene calling and functional annotation
RDP
Provides ribosome related data and services to the scientific community, including online data analysis and aligned and annotated Bacterial and Archaeal small-subunit 16S rRNA sequences
program
Qiime
A pipeline for performing microbial community analysis that integrates many standard third party too
addresses the problem of taking sequencing data from raw sequences to interpretation and databas
speciateIT
A package for speciation of 16S sequences
CODE REPOSITORY
(for developers)
The DACC hosts a Subversion repository for use by all HMP participants. A user account is required in order to add or modify existing code within the repository. All code submitted to this repository is publicly viewable. To obtain an account, please request one via our feedback form. We will provide you with the details once your account has been created.
DACC SVN Repository


Human Microbiome Project DACC - Home - Mozilla Firefox
http://hmpdacc.org/tools_protocols_type.php
Human Microbiome Project DACC - H...
REFERENCE GENOMES
MICROBIOME ANALYSES
IMPACTS ON HEALTH
TECHNOLOGY DEVELOPMENT
ETHICAL IMPLICATIONS
OUTREACH
HUMAN MICROBIOME PROJECT
home > tools & protocols by type
Online Resources
Fast-Unifrac
Provides a suite of tools for the comparison of microbial communities using phylogenetic information
SUBVERSION REPOSITORIES DACC
[/] [software/] [by_name/] - Rev 18
Changes | View Log | RSS feed
LAST MODIFICATION
Rev 18 2009-07-28 13:54:47
Author: jorvis
Log message:
1.0 release
Path
software/
by_name/
AbundanceBin/
AMOS/
Bowtie/
BrainGrab/
metastats/
MicrobiomeUtilities/
Minimus_SR/
MinPath/
Phymm/
SIMBAL/
by_source/
SOPs/
CODE REPOSITORY
(for developers)
The DACC hosts a Subversion repository for use by all HMP participants. A user account is required in order to add or modify existing code within the repository. All code submitted to this repository is publicly viewable. To obtain an account, please request one via our feedback form. We will provide you with the details once your account has been created.
DACC SVN Repository
aligned and annotated Bacterial and Archaeal small-subunit 16S rRNA sequences
program
Qiime
A pipeline for performing microbial community analysis that integrates many standard third party too
addresses the problem of taking sequencing data from raw sequences to interpretation and databas
speciateIT
A package for speciation of 16S sequences

CloVR - Cloud Virtual Resource



PI: Florian Fricke
Technical lead: Sam Angiuoli

Large-scale Amazon Deployment

Ganglia .sourceforge.net

CloVR Cluster Report for Sat, 17 Jul 2010 16:55:28 +0000

Metric load_one **Last** day **Sorted** descending

Grid > CloVR > --Choose a Node

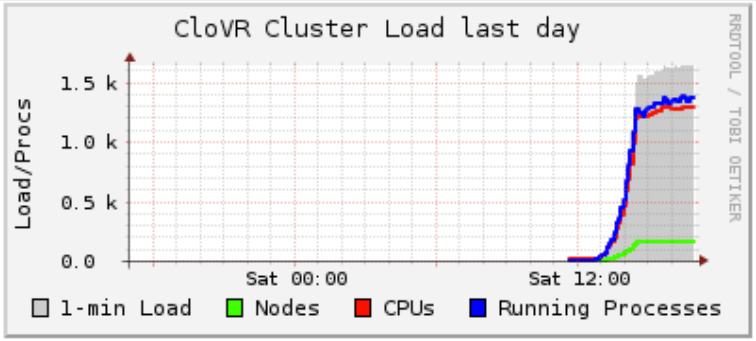
Overview of CloVR

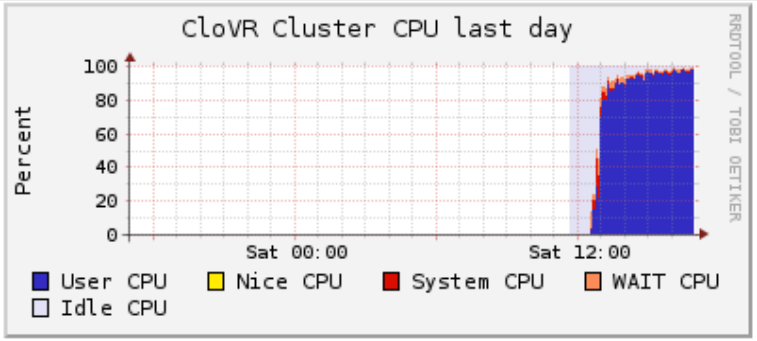
CPUs Total:	**1288**
Hosts up:	**160**
Hosts down:	**1**

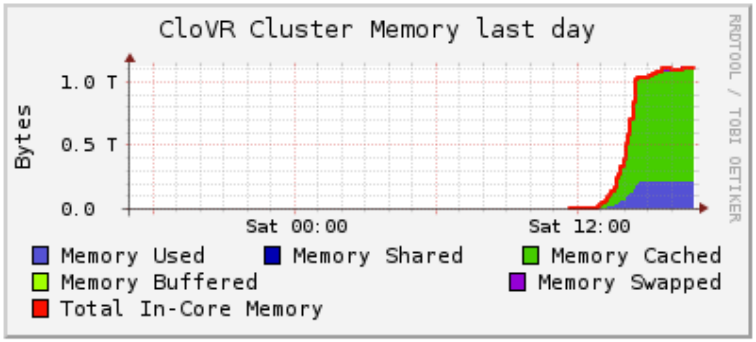
Avg Load (15, 5, 1m): **126%, 127%, 127%**

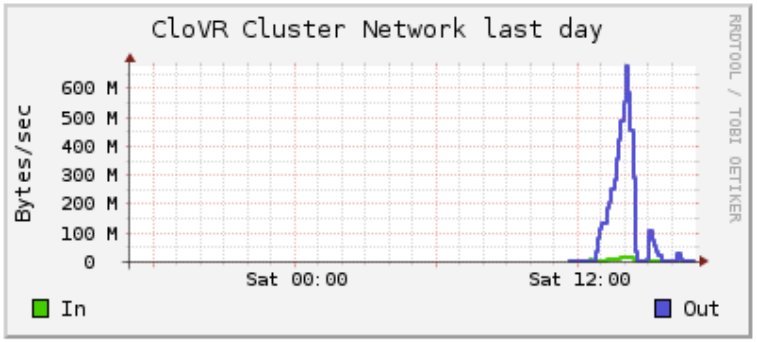
Localtime: **2010-07-17 16:55**

Pie Chart









Phase 2: More Access

- Open access data
  - Annotated data sets, aggregated, searchable
  - Some pre-computes
  - "Reference" data sets
- Research network
  - Processed files
  - Aggregated datasets
  - Metadata

We are surveying the community now!
See:

Heather Huot Creasy
Cathering Jordan

Phase 2

- External users will:
  - Select data sets /results for download
  - Search for specific data
  - Access data archives (may be some with controlled access)
  - See data reports, stats about data, validation process, etc
  - See information about metadata

Phase 3: Analysis Tools

- Annotation Pipelines
  - RAMMCAP Rapid analysis of Multiple Metagenomes with Clustering and Annotation Pipeline
  - ShotgunFunctionalizeR
- Binning
  - SOrt-ITEMS Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences
- Community composition, comparative metagenomics
  - MEGAN (MEtaGenome ANalyzer)
  - CARMA
  - GAAS (Genome relative Abundance and Average Size)
  - Galaxy
  - GINKGO
  - Metarep Suite of web based tools
  - Metastats compare clinical metagenomic samples from two treatment populations
  - RAMMCAP - Statistical metagenome comparison
  - ShotgunFunctionalizeR R-package for functional comparison
- Visualization
  - Invue API and software suite for large scale data visualization
- Online resources
  - My IMG/M tools for analyzing microbiome functional capability
  - MG-RAST - variety of comparative and visualization tools

HMP DACC Team

IGS

Jennifer Wortman
Michelle Gwinn Giglio
Heather Huot Creasy
Brandi Cantarel
Jonathan Crabtree
Joshua Orvis
Cesar Arze
Mark Mazaitis
Victor Felix
Catherine Jordan
Anup Mahurkar

LBL

Gary Andersen
Todd DeSantis
Navjeet Singh
Victor Markowitz
Amy Chen

JGI

Nikos Kyrpides
Konstantinos Liolios

Univ. of Colorado

Rob Knight
Dan Knights
Justin Kuczynski

Cornell University : Ruth Ley,

San Diego State: Scott Kelley

Argonne National Lab: Folker Meyer