

Standards,
Precautions &
Advances in
Ancient
Metagenomics

Lecture 5C: Microbial functional analysis

Irina M. Velsko



Overview

1. Why perform functional classification?
2. Gene-based classification
2. Species pan-genomes
3. Metagenomic gene content
4. HUMAnN3 (HMP Unified Metabolic Analysis Network 3)



What do you work on?

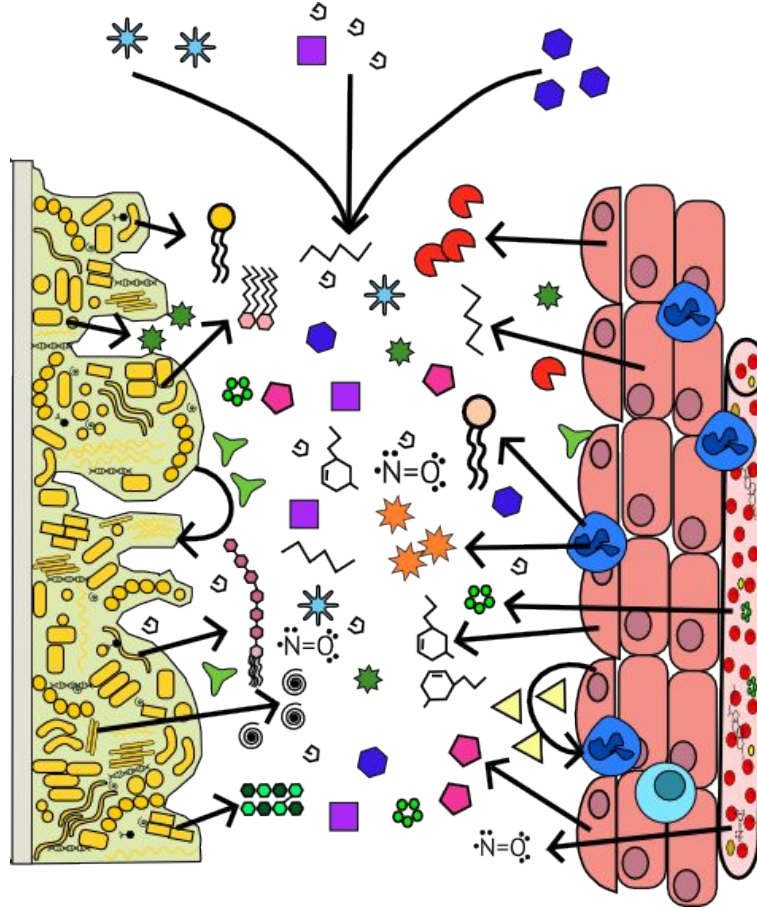




What is meant by function?



Microbial biofilm



Environment
(Blood vessel,
Gingival tissue)

Function ↑



Microbial functional analysis can tell you...

...what it produces

...what it can break down



Microbial functional analysis can tell you...

- ...what it produces
- ...what it can break down
- ...how can it kill you
- ...how can it be killed



Microbial functional analysis can tell you...

...what it produces

...what it can break down

...how can it kill you

...how can it be killed

...how can be used in ...

 bioremediation (i.e. breakdown oil, sequester heavy metals)

 medical treatment (i.e. species transplants)

 engineering (i.e. batteries, clean marble statues, produce plastics)

.

.

.

...how it interacts with the world and how that can be harmful or beneficial



Why study microbial function?

(What's infecting the cows?)

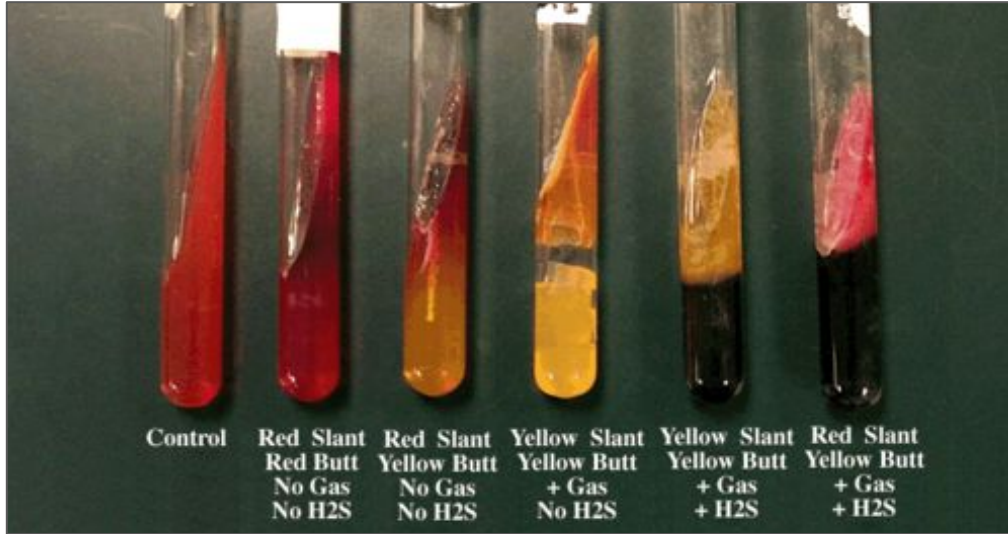
(What's spoiling the milk?)

(How to cure/prevent it?)



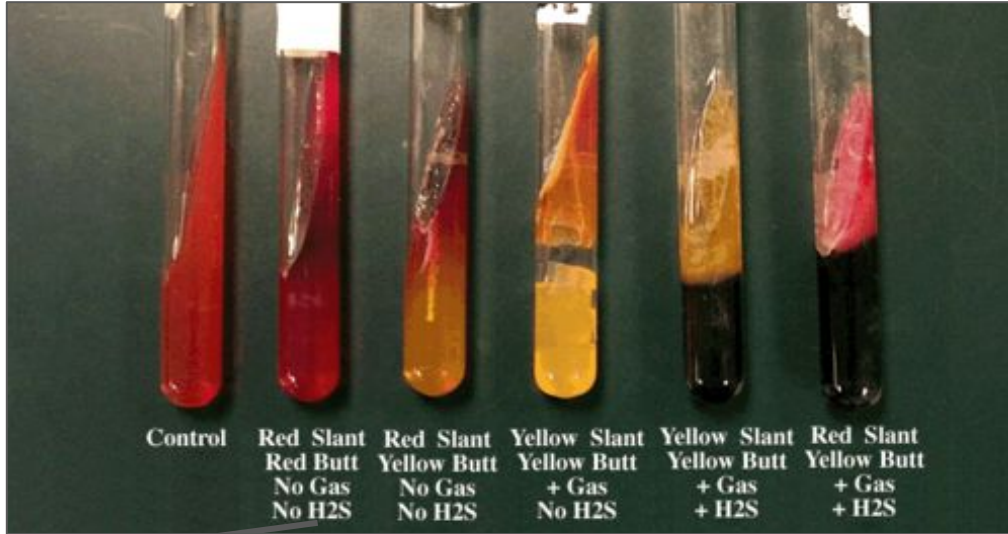
Species identification by function (old method)

Biochemical testing



Species identification by function (old method)

Biochemical testing



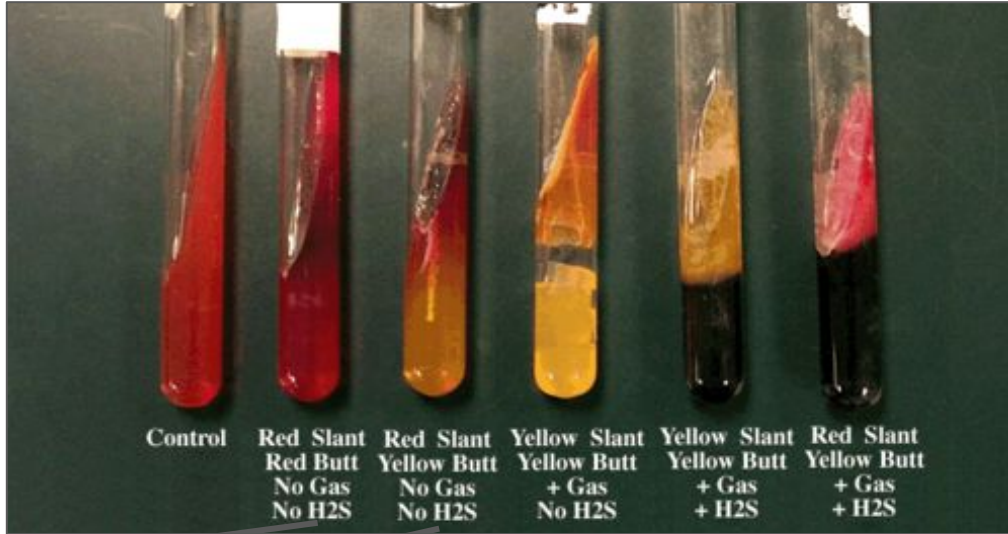
Species 1

Probably not
spoiling food



Species identification by function (old method)

Biochemical testing



Species 1

Probably not
spoiling food

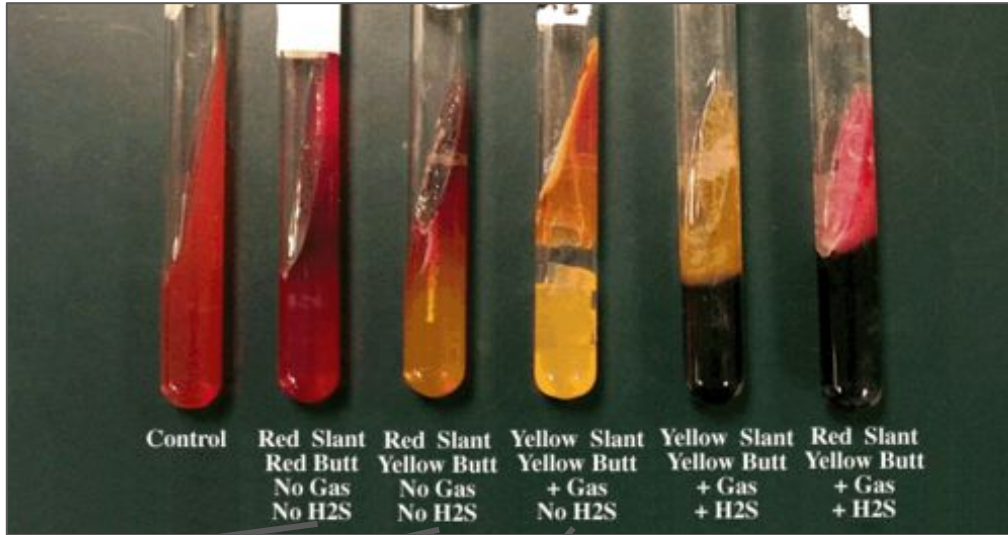
Species 2

Curdling (acid)



Species identification by function (old method)

Biochemical testing



Species 1

Probably not
spoiling food

Species 2

Curdling (acid)

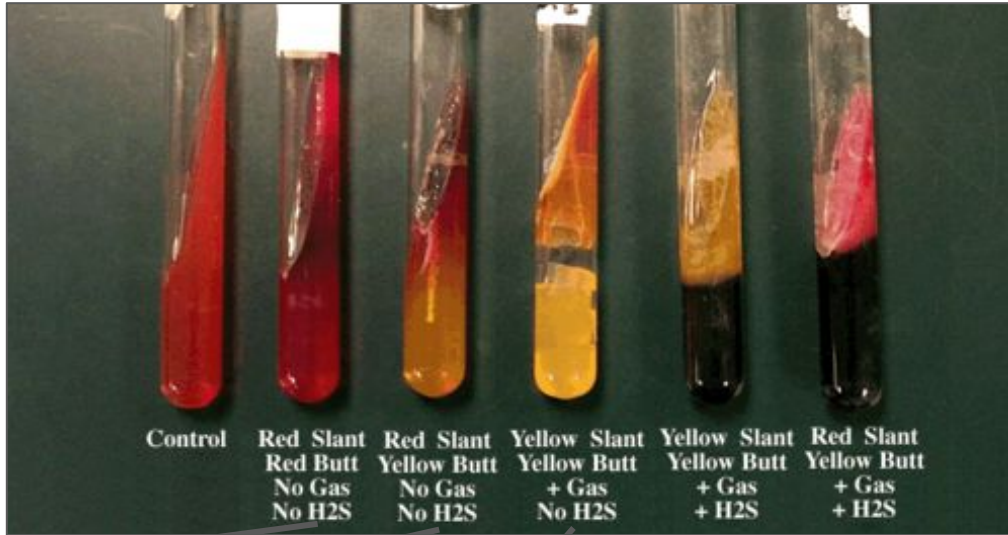
Species 3

Curdling (acid),
inflated jar



Species identification by function (old method)

Biochemical testing



Species 1

Probably not
spoiling food

Species 2

Curdling (acid)

Species 3

Curdling (acid),
inflated jar

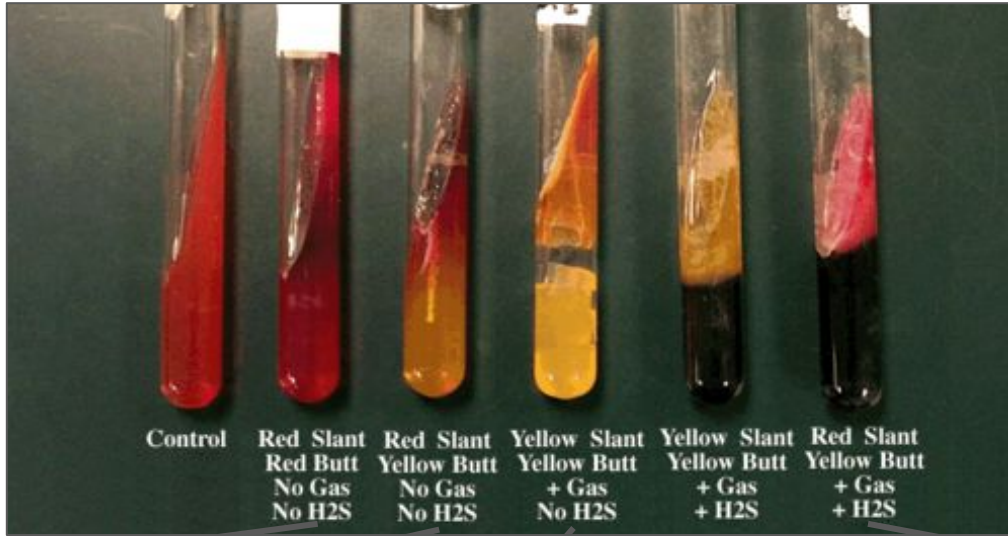
Species 4

Rotten egg
flavor, curdling,
inflated jar



Species identification by function (old method)

Biochemical testing



Species 1
Probably not
spoiling food

Species 2
Curdling (acid)

Species 3
Curdling (acid),
inflated jar

Species 4
Rotten egg
flavor, curdling,
inflated jar

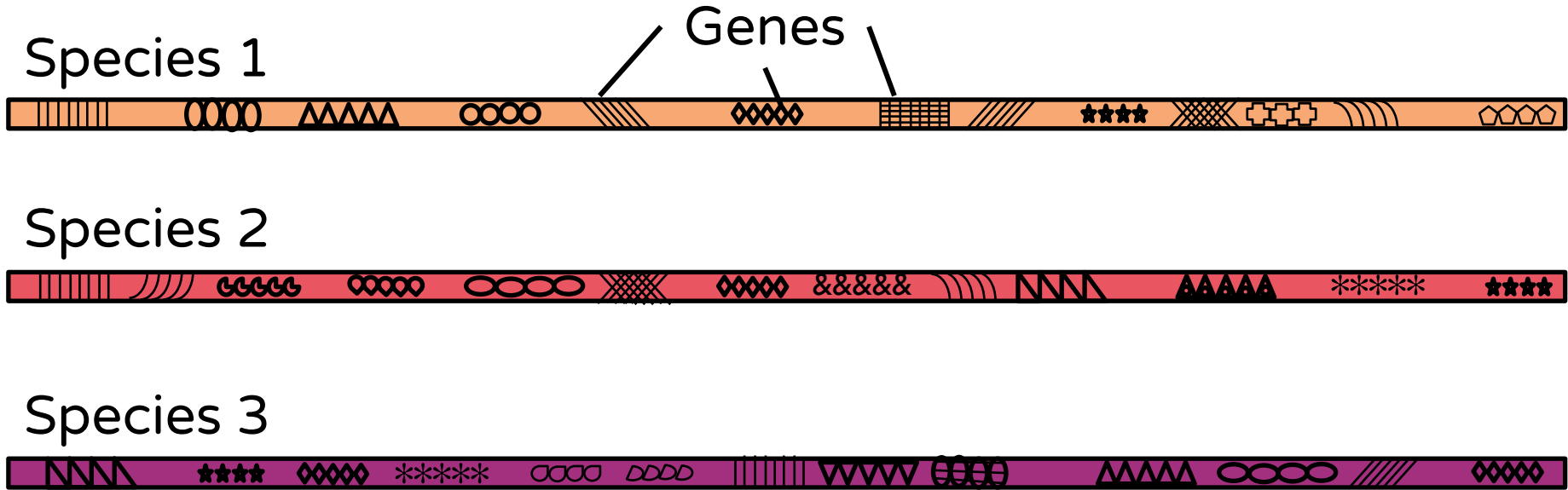
Species 5
Rotten egg
flavor,
inflated jar



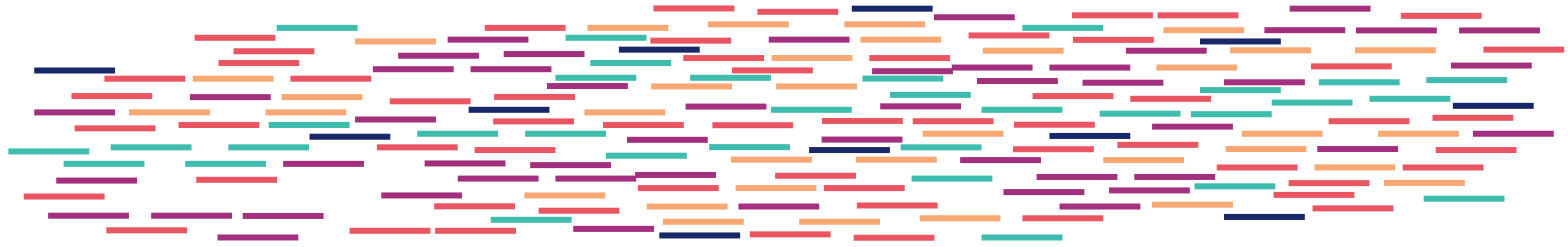
Where do we get functional
information in a(n ancient)
metagenome?



Genomes contain taxonomic and functional information



Metagenomes are classified against known species genomes



Species 1



Species 2

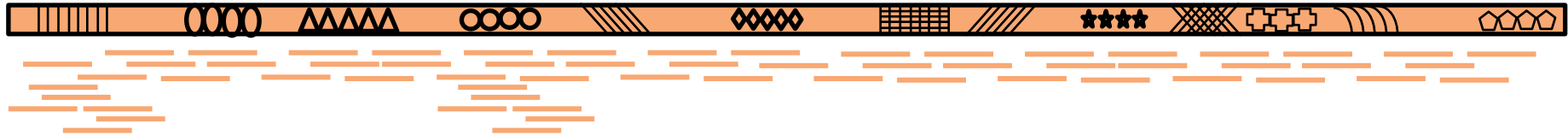


Species 3

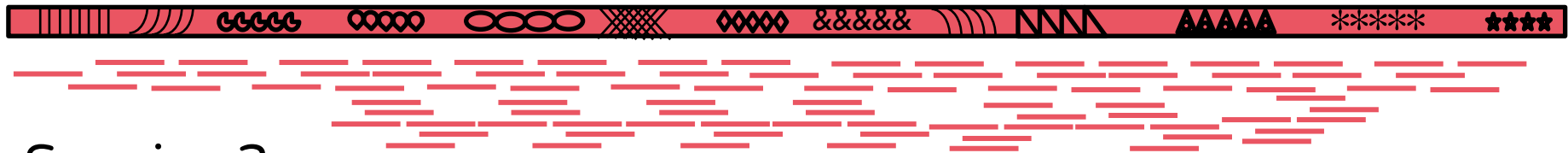


Metagenomes contain taxonomic and functional information

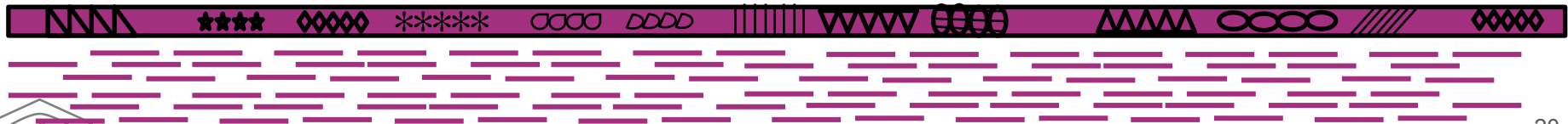
Species 1



Species 2

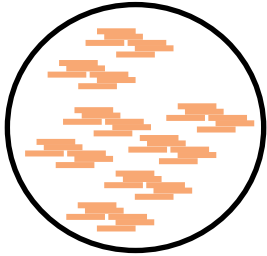


Species 3

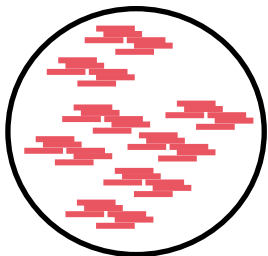


Many metagenome reads cannot be classified

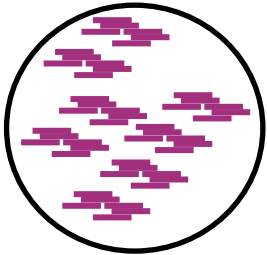
Species 1



Species 2



Species 3

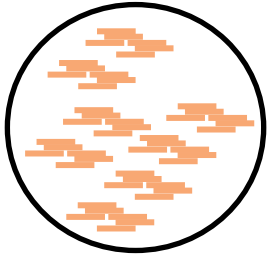


Species	Sa1	Sa2	Sa3
Species 1	34	18	27
Species 2	56	78	15
Species 3	16	55	63

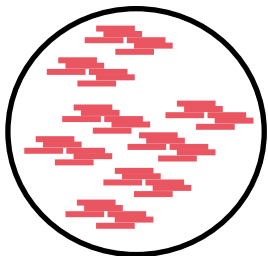


Many metagenome reads cannot be classified

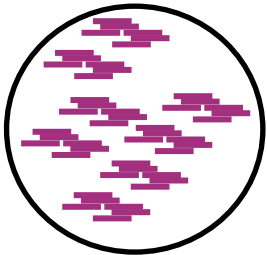
Species 1



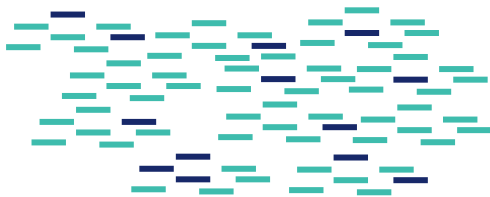
Species 2



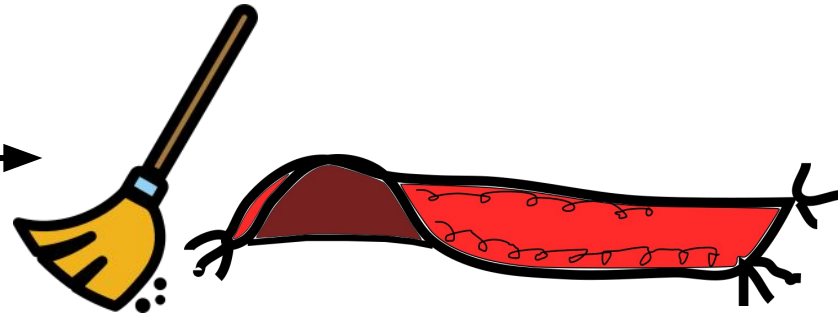
Species 3



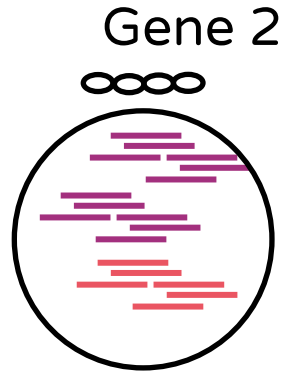
Not classified



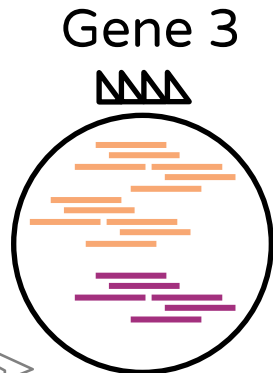
Species	Sa1	Sa2	Sa3
Species 1	34	18	27
Species 2	56	78	15
Species 3	16	55	63



Function from grouping reads by their gene of origin



Gene	Sa1	Sa2	Sa3
Gene 1	34	18	27
Gene 2	56	78	15
Gene 3	16	55	63



Not classified



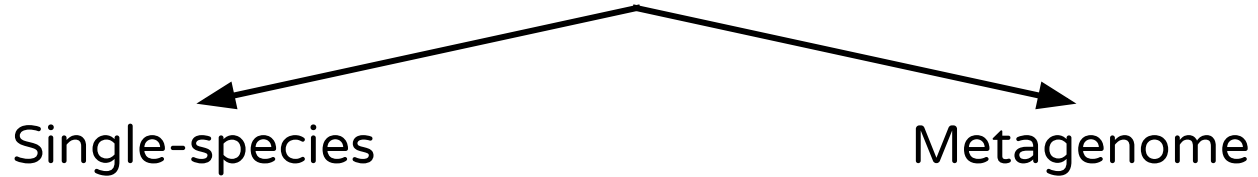
Most microbial isolates have genes of unknown function

Isolate	Total CDS	Hypothetical CDS	%
<i>Streptococcus gordonii</i> str. Challis substr. CH1	2061	290	14
<i>Streptococcus mutans</i> UA159	1859	144	7
<i>Escherichia coli</i> K-12 MG1655	3829	112	3
<i>Yersinia pestis</i> CO92	4098	350	8.5
<i>Mycobacterium tuberculosis</i> H37Rv	3946	473	12
<i>Staphylococcus aureus</i> NCTC 8325	2535	276	11
<i>Tannerella forsythia</i> 92A2	2665	453	17
<i>Desulfobulbus oralis</i> HOT-041	2373	415	17



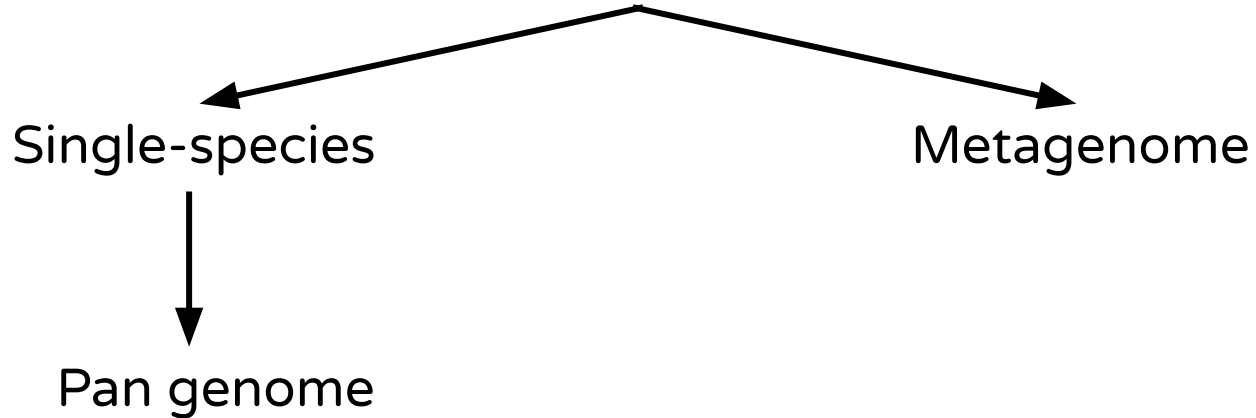
How can a gene table be used?

Gene	Sa1	Sa2	Sa3
Gene 1	34	18	27
Gene 2	56	78	15
Gene 3	16	55	63

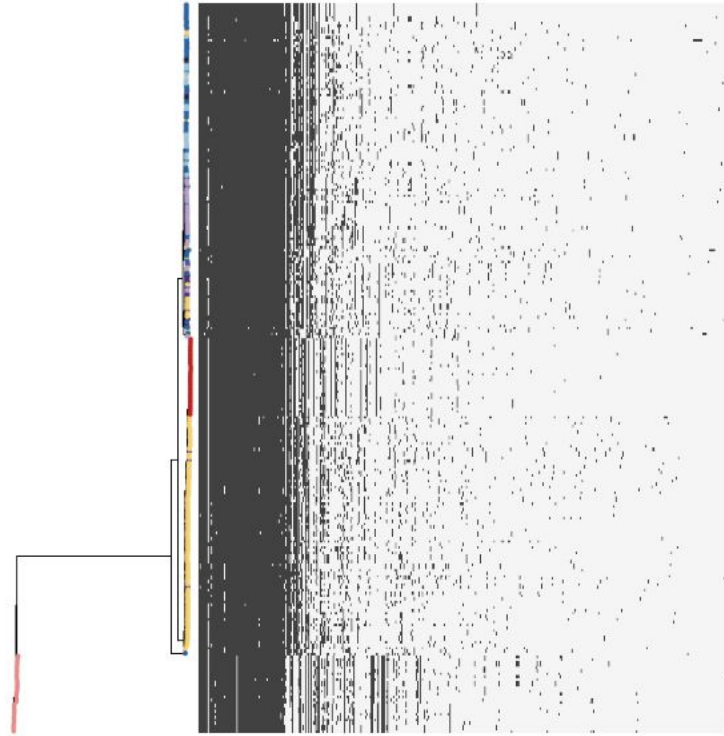


How can a gene table be used?

Gene	Sa1	Sa2	Sa3
Gene 1	34	18	27
Gene 2	56	78	15
Gene 3	16	55	63



Pan genome

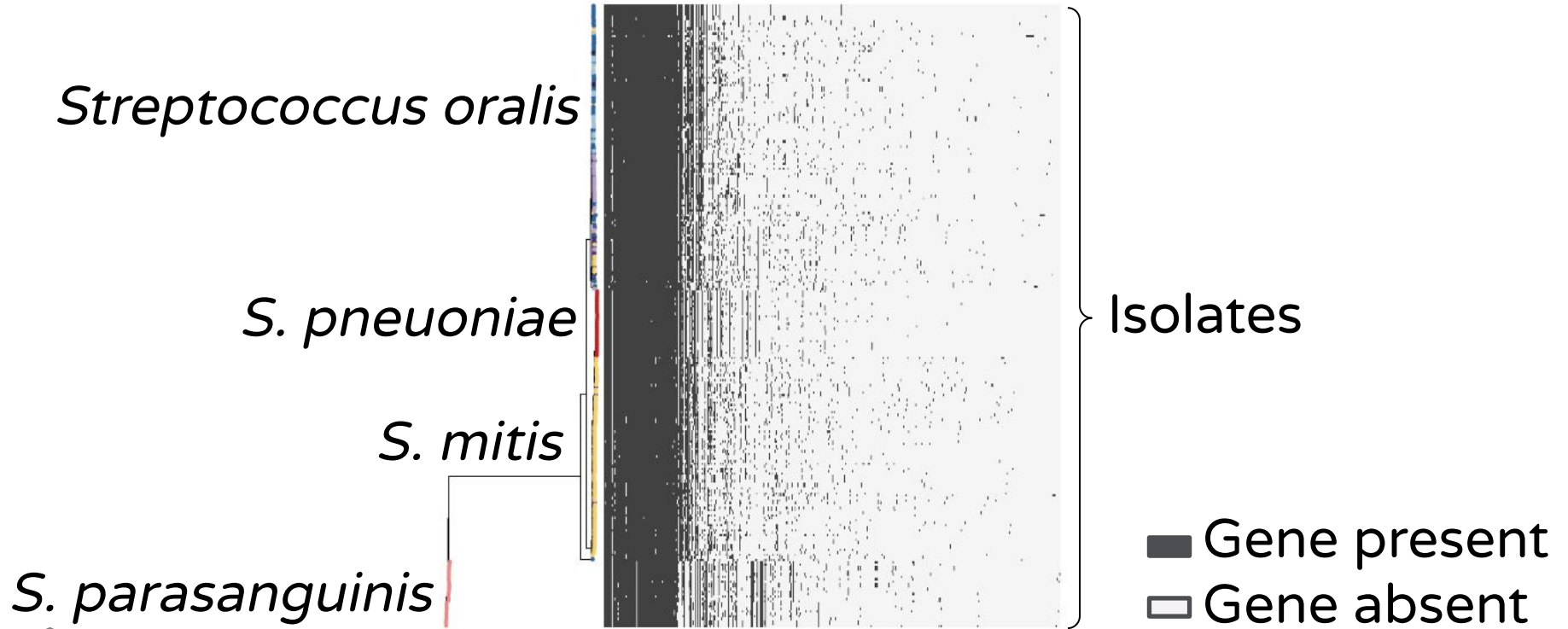


- Gene present
- Gene absent

Modified from Velsko, *et al.* 2019 *GBE*



Pan genome



Modified from Velsko, *et al.* 2019 *GBE*



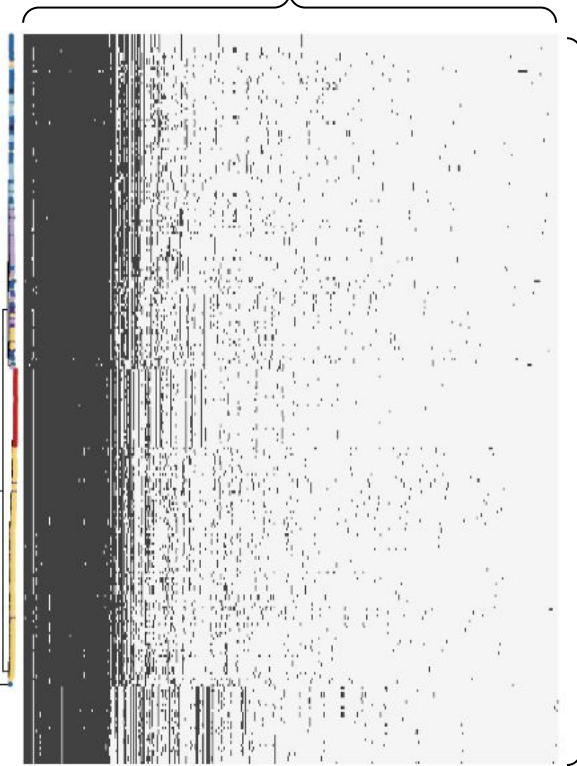
Pan genome Genes

Streptococcus oralis

S. pneumoniae

S. mitis

S. parasanguinis



Isolates

■ Gene present
□ Gene absent


Modified from Velsko, et al. 2019 *GBE*



Tools for pan genome analysis

The Pangenome pp 89–112 | [Cite as](#)

A Review of Pangenome Tools and Recent Studies

[G. S. Vernikos](#) 

Chapter | [Open Access](#) | [First Online: 01 May 2020](#)

16k Accesses | **9** Citations

Abstract

With the advance of sequencing technologies, the landscape of genomic analysis has been transformed, by moving from single strain to species (or even higher taxa)-wide genomic resolution, toward the direction of capturing the “totality” of life diversity; from this scientific advance and curiosity, the concept of “pangenome” was born. Herein we will review, from practical and technical implementation, existing projects of pangenome analysis, with the aim of providing the reader with a snapshot of useful tools should they need to embark on such a pangenomic journey.

https://link.springer.com/chapter/10.1007/978-3-030-38281-0_4



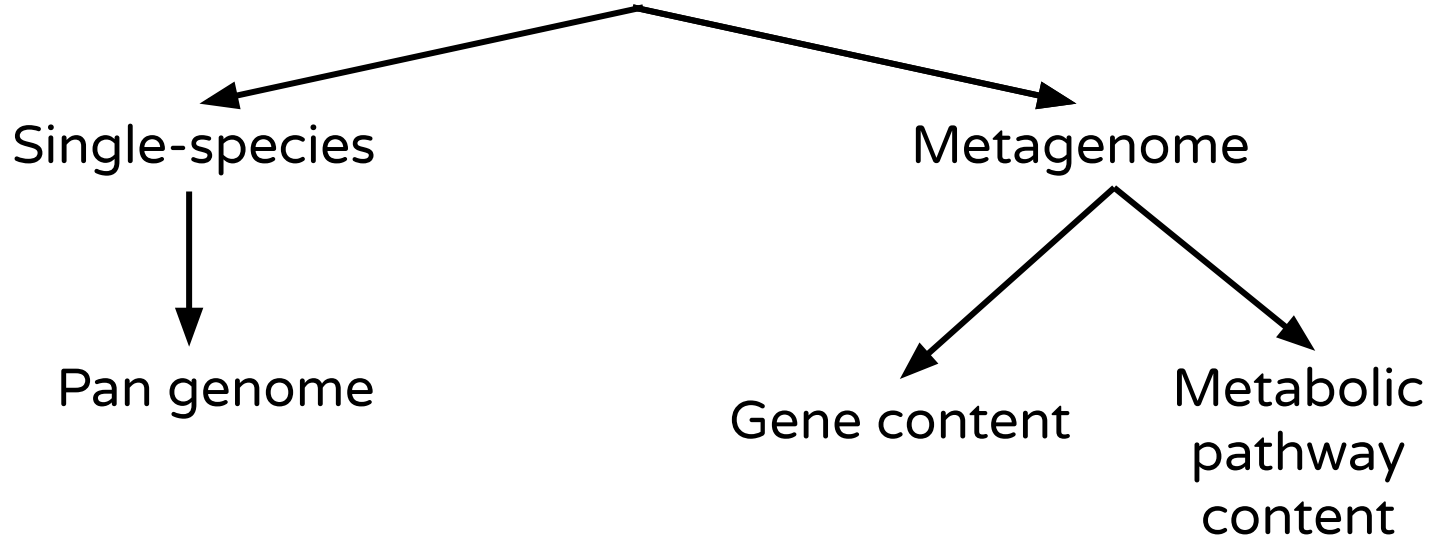
Note for pan genome analysis

- All genomes must be annotated in the same way
- Do not use NCBI annotations (too many different annotation programs)
- Download fasta files and annotate them yourself
- Use whatever program you like best (roary, etc)



How can a gene table be used?

Gene	Sa1	Sa2	Sa3
Gene 1	34	18	27
Gene 2	56	78	15
Gene 3	16	55	63



(some) Tools for gene content analysis

- AADDER
 - MEGAN tool (never published)
 - DNA alignment
 - Like MALT (high memory, long time)



(some) Tools for gene content analysis

- AADDER
 - MEGAN tool (never published)
 - DNA alignment
 - Like MALT (high memory, long time)
- Paladin (<https://doi.org/10.1093/bioinformatics/btx021>)
 - Protein alignment
 - (not actively maintained)




(some) Tools for gene content analysis

- AADDER
 - MEGAN tool (never published)
 - DNA alignment
 - Like MALT (high memory, long time)
- Paladin (<https://doi.org/10.1093/bioinformatics/btx021>)
 - Protein alignment
 - (not actively maintained)
- HUMAnN3 (<https://doi.org/10.7554/eLife.65088>)
 - DNA and protein alignment
 - BioBakery (MetaPhlan, LefSe, MaAsLin2, etc)

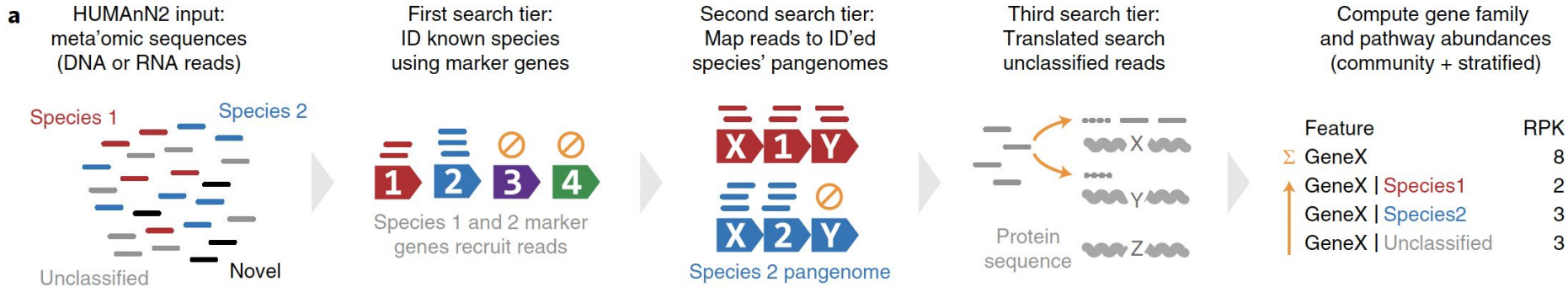


(some) Tools for gene content analysis

- AADDER
 - MEGAN tool (never published)
 - DNA alignment
 - Like MALT (high memory, long time)
- Paladin (<https://doi.org/10.1093/bioinformatics/btx021>)
 - Protein alignment
 - (not actively maintained)
- HUMAnN3 (<https://doi.org/10.7554/eLife.65088>) 
 - DNA and protein alignment
 - BioBakery (MetaPhlan, LefSe, MaAsLin2, etc)

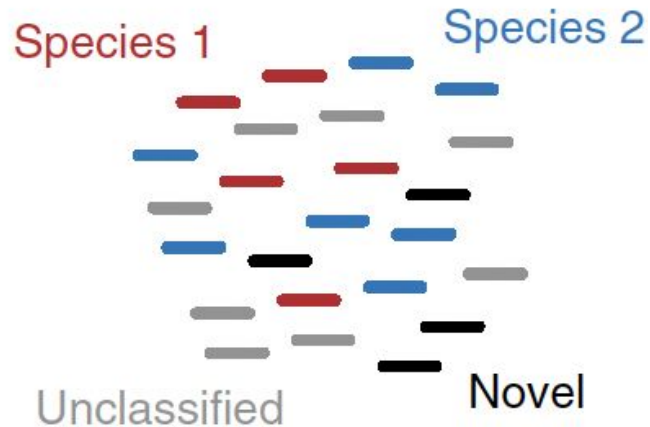


Metagenome gene content with HUMAnN3



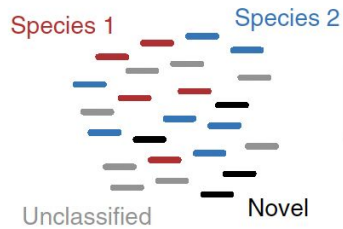
Metagenome sample

a HUMAnN2 input:
meta'omic sequences
(DNA or RNA reads)



Align all reads to marker gene database

a HUMAnN2 input:
meta'omic sequences
(DNA or RNA reads)



First search tier:
ID known species
using marker genes

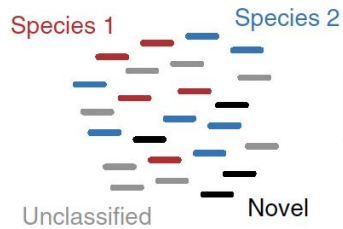


Species 1 and 2 marker
genes recruit reads



Align all reads to species pan genomes

a HUMAnN2 input:
meta'omic sequences
(DNA or RNA reads)



First search tier:
ID known species
using marker genes



Second search tier:
Map reads to ID'ed
species' pangenomes

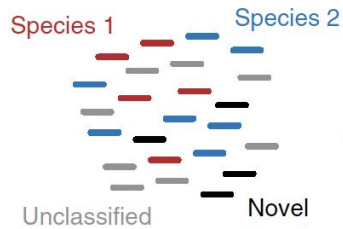


Species 2 pangenome



Translate only unaligned reads and align to protein database

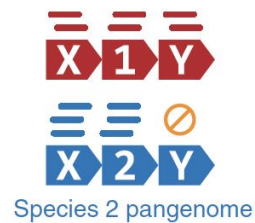
a HUMAnN2 input:
meta'omic sequences
(DNA or RNA reads)



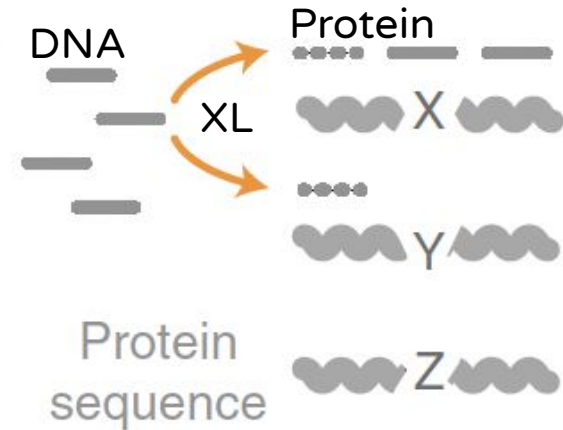
First search tier:
ID known species
using marker genes



Second search tier:
Map reads to ID'ed
species' pangenomes

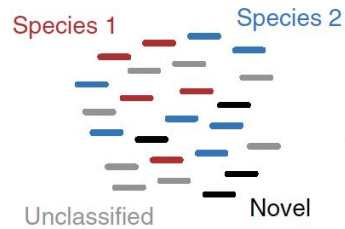


Third search tier:
Translated search
unclassified reads



Translate only unaligned reads and align to protein database

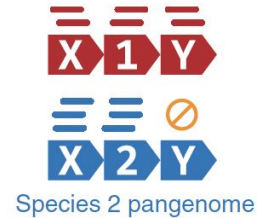
a HUMAnN2 input:
meta'omic sequences
(DNA or RNA reads)



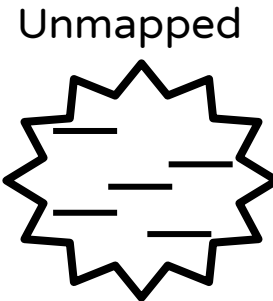
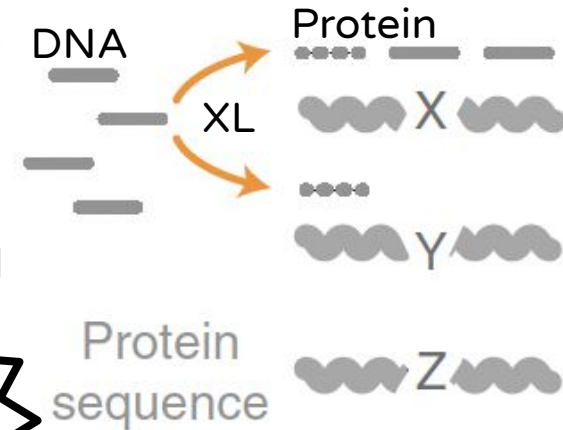
First search tier:
ID known species
using marker genes



Second search tier:
Map reads to ID'ed
species' pangenomes

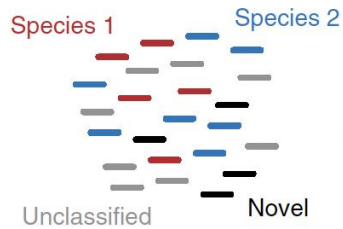


Third search tier:
Translated search
unclassified reads



Normalization

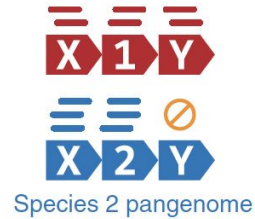
a HUMAnN2 input:
meta'omic sequences
(DNA or RNA reads)



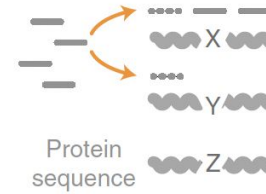
First search tier:
ID known species
using marker genes



Second search tier:
Map reads to ID'ed
species' pangenomes



Third search tier:
Translated search
unclassified reads



Compute gene family
and pathway abundances
(community + stratified)

Feature	RPK
Σ GeneX	8
GeneX Species1	2
GeneX Species2	3
GeneX Unclassified	3



Determine pathways from gene content

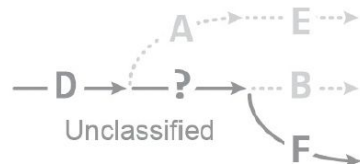
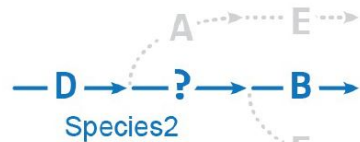
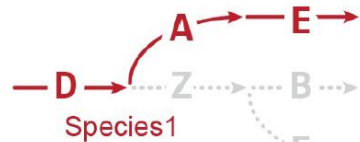
b HUMAnN2 gene family & pathway quantification

Gene abundance estimation

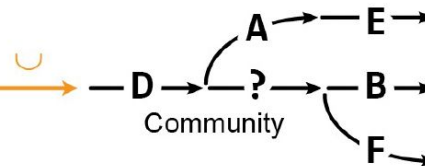
Feature	RPK
GeneA	2
GeneA Species1	2
GeneB	3
GeneB Species2	3
Σ GeneD	8
GeneD Species1	2
GeneD Species2	3
GeneD unclassified	3
GeneE	2
GeneE Species1	2
GeneF	5
GeneF unclassified	5

Process mapping results to estimate per-species and community total gene family abundance, weighting by 1) alignment quality, 2) gene length, and 3) gene coverage

Per-species and community-level metabolic network reconstruction



Map genes to metabolic reactions; identify a parsimonious set of pathways that explain each species' observed reactions



Quantify pathway abundance and coverage by 1) optimizing over alternative subpathways and 2) imputing abundance for conspicuously depleted ("?) reactions



What's a metabolic pathway ?

(and why do we want it if we have gene counts?)



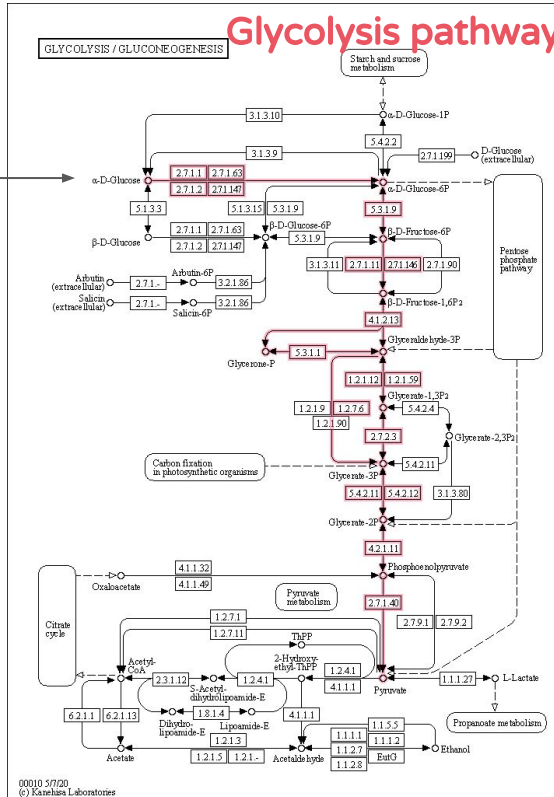
Pathways are proteins processing molecules in sequence



Pathways are proteins processing molecules in sequence

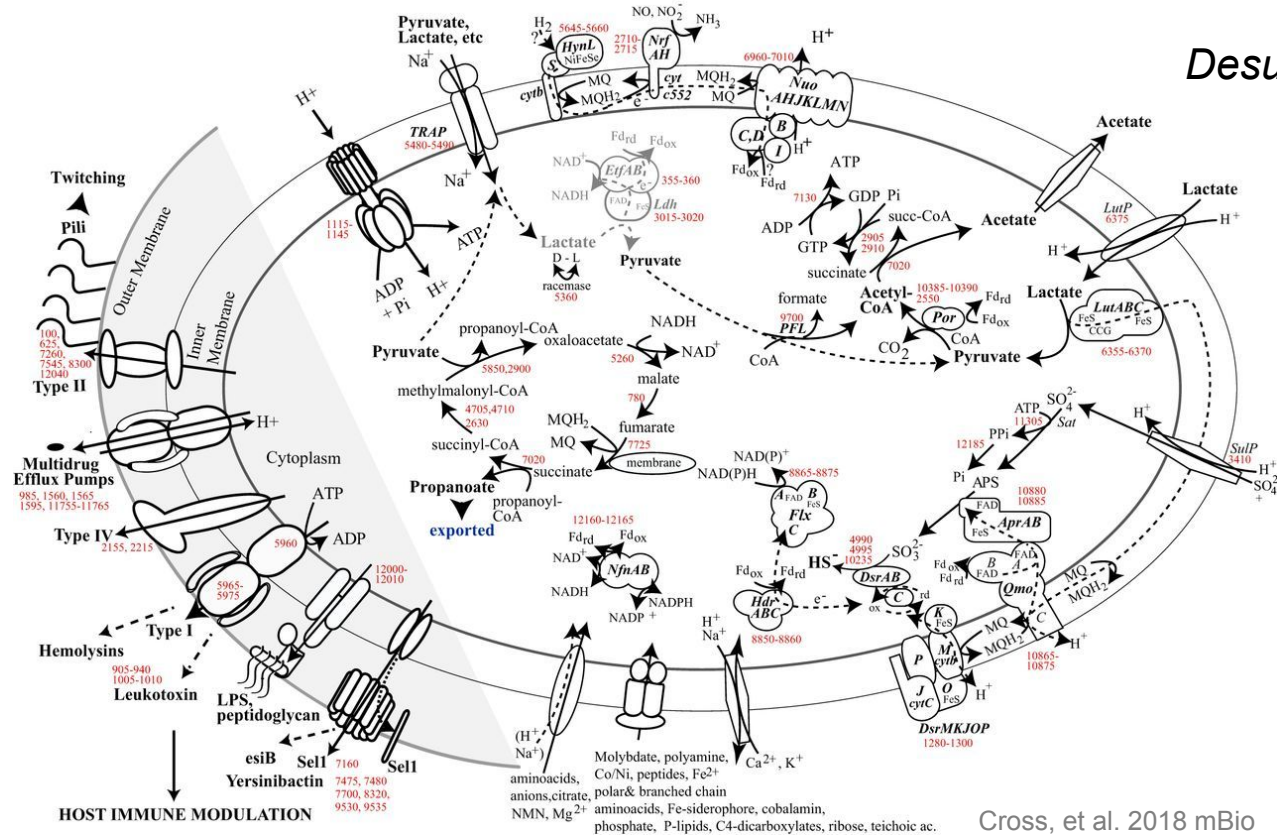
KEGG map00010 Central Carbohydrate Metabolism

Start
D-Glucose



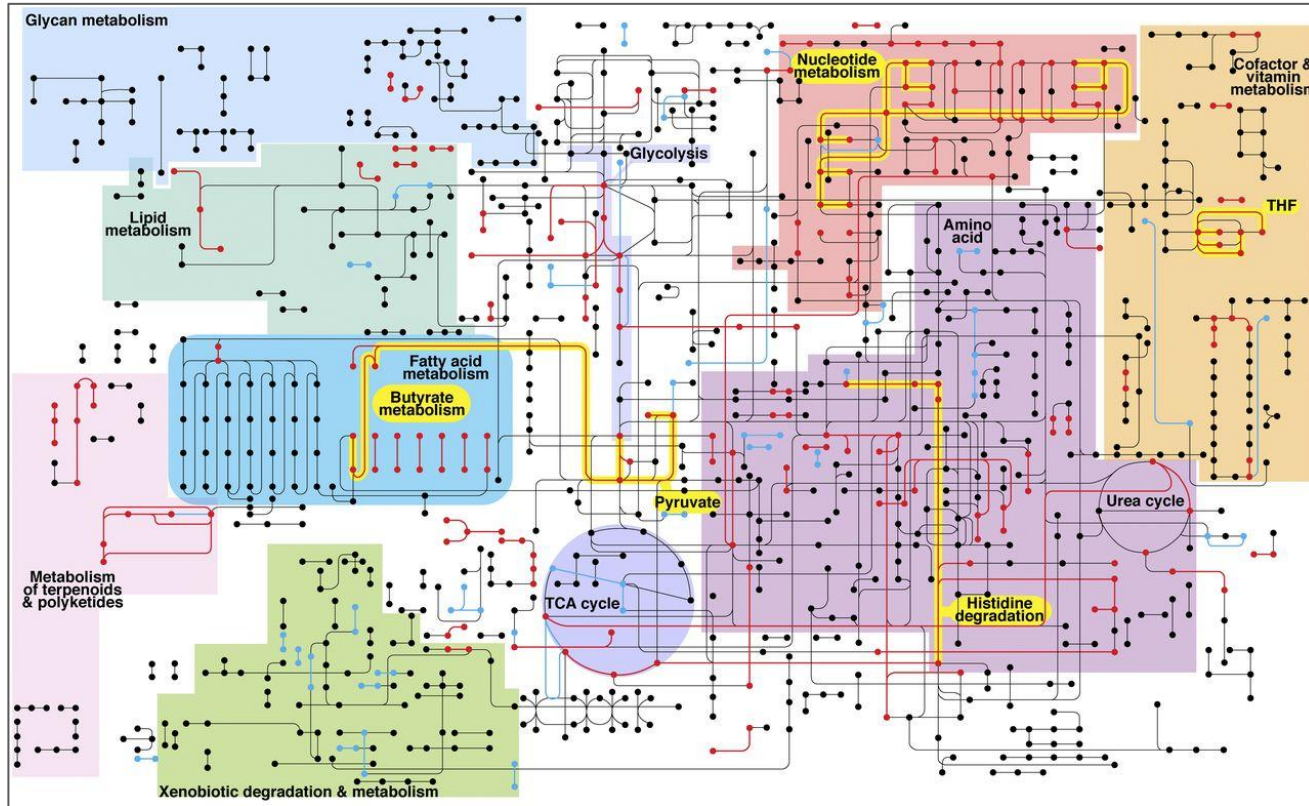
Pathways describe species activity range

Desulfobulbus oralis



Cross, et al. 2018 mBio
<https://doi.org/10.1128/mBio.02061-17>

Pathways describe metagenome activity range

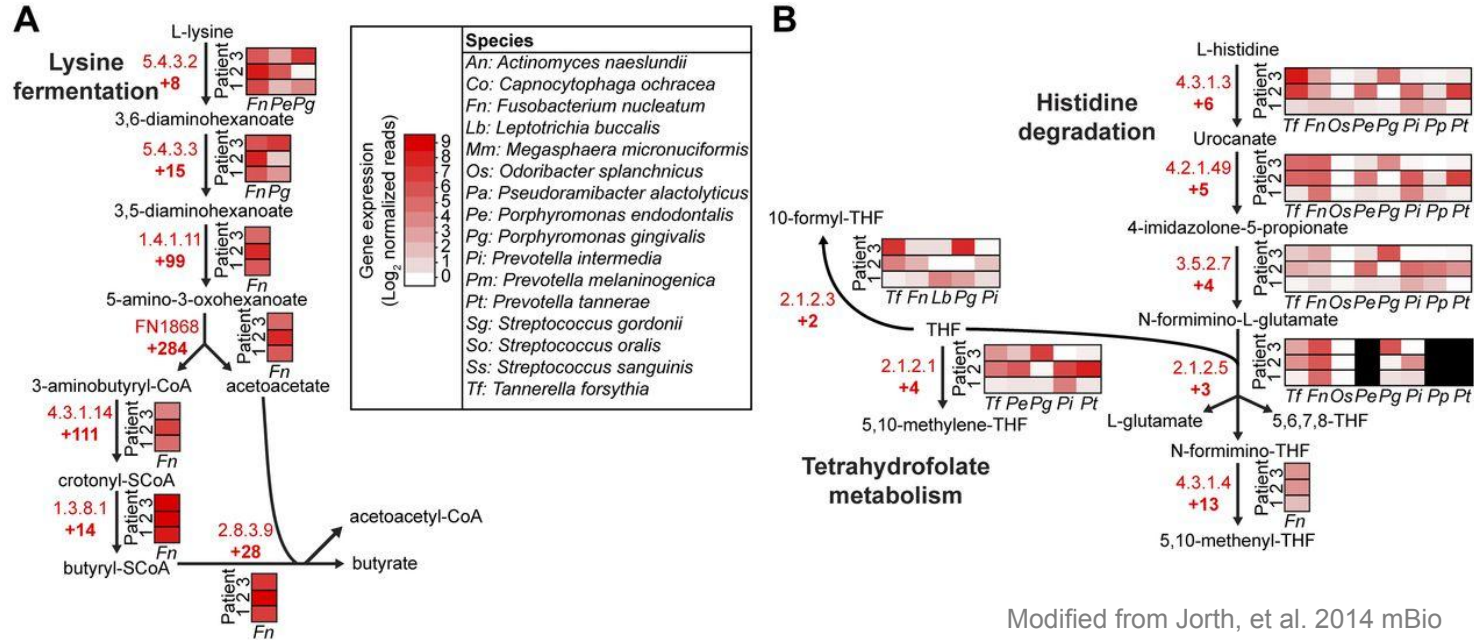


Yellow pathways upregulated in periodontitis-associated microbiomes compared to health

Jorth, et al. 2014 mBio
<https://doi.org/10.1128/mBio.01012-14>



Pathways can be broken out into the species that have genes in the pathway



Modified from Jorth, et al. 2014 mBio
<https://doi.org/10.1128/mBio.01012-14>



c Data products

Each meta'ome profiled produces **3 main outputs**:

Gene family
abundance profile
(stratified by species)

Pathway
abundance profile
(stratified by species)

Pathway
coverage profile
(stratified by species)



Gene families vs pathways

- Genes:
 - identify **specific genes** that are enriched for further investigation of these specific genes (independent of the pathways it's involved in)
- Pathways:
 - identify a **series of genes** that act on a metabolite in sequence with a defined end product



c Data products

Each meta'ome profiled produces **3 main outputs**:

Gene family abundance profile
(stratified by species)

Pathway abundance profile
(stratified by species)

Pathway coverage profile
(stratified by species)

Gene	Sa1	Sa2	Sa3
Unmapped	34.9	18.3	27
Gene 1	1.2	4.5	1.0
Gene 1 Species1	0.9	2.1	6.7
Gene 1 Species2	3.3	0.7	4.6
Gene 1 Species3	0.4	0.5	0.2
Gene 1 unclassified	20.3	10.0	15
Gene 2	4.0	0.6	0.8
Gene 2 Species 1	0.9	0.01	0.02



Gene content stratified output

UniRef90_A0A024DHT2	0	4.04336	0	0	0	0	0	0
UniRef90_A0A024DHT2 g__Streptococcus.s__Streptococcus_gordonii							0.859866	
UniRef90_A0A024DHT2 g__Streptococcus.s__Streptococcus_oralis							3.18349	0
UniRef90_A0A024DHT2 g__Streptococcus.s__Streptococcus_sp_oral_taxon_056							0	0
UniRef90_A0A024DIJ1	0	3.88208	0	0	0	0	0	0
UniRef90_A0A024DIJ1 g__Streptococcus.s__Streptococcus_oralis							3.88208	0
UniRef90_A0A024DIQ4	0	0	0	0	0	0	0	0
UniRef90_A0A024DIQ4 unclassified							0	0
UniRef90_A0A024DIT2	0	0	0	0	0	0	0	0
UniRef90_A0A024DIT2 unclassified							0	0
UniRef90_A0A024GWJ5	0	1.04735	6.9544	0	0	1.0487	0	1.11773
UniRef90_A0A024GWJ5 unclassified							1.04735	6.9544

A0A024DHT2: Lantibiotic ABC transporter permease

A0A024DIJ1: Membrane protein

etc



Normalize to copies per million (CPM)



Group genes based on classification system

- UniProt
- MetaCyc reaction
- KEGG orthologs
- GeneOntology (GO)
- Enzyme commission (EC)
- Pfam
- EGGNog



Group genes based on classification system

- UniProt
- MetaCyc reaction
- KEGG orthologs
- GeneOntology (GO)
- Enzyme commission (EC)
- Pfam
- EGGNog

ortholog - homologous genes that have diverged from each other as a consequence of speciation

paralog - homologous genes that have diverged from each other as a consequence of genetic duplication



Group genes based on classification system

- UniProt (<https://www.uniprot.org>)
- MetaCyc reaction (<https://metacyc.org>)
- KEGG orthologs (<https://www.genome.jp/kegg/ko.html>)
- GeneOntology (GO) (<http://geneontology.org>)
- Enzyme commission (EC) (<https://enzyme.expasy.org>)
- Pfam (<http://pfam.xfam.org>)
- EGGNog (<http://eggnog5.embl.de/#/app/home>)



Add names to genes

1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	47.309340000000006	24.494439999999997	81.568559999999999	71.619740000000001				
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Anaerolineaceae_unclassified.s__Anaerolineaceae_bacterium_oral_taxon_439		10.1386	6.24735	10.6777			
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Bacteroidetes_unclassified.s__Bacteroidetes_bacterium_oral_taxon_272		2.11437	0.0	0.0			
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Bacteroidetes_unclassified.s__Bacteroidetes_oral_taxon_274	5.20685	2.68581	1.7635	0.0	13.3631		
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Capnocytophaga.s__Capnocytophaga_gingivalis	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Capnocytophaga.s__Capnocytophaga_leadbetteri	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Capnocytophaga.s__Capnocytophaga_ochracea	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Capnocytophaga.s__Capnocytophaga_sp_oral_taxon_332	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Capnocytophaga.s__Capnocytophaga_sp_oral_taxon_338	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Capnocytophaga.s__Capnocytophaga_sp_oral_taxon_878	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Capnocytophaga.s__Capnocytophaga_sputigena	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Desulfomicrobium.s__Desulfomicrobium_orale	0.0	0.0	45.4766	33.5371	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Fretibacterium.s__Fretibacterium_fastidiosum	8.66264	3.31647	4.89466	3.71567	3.02542	7.33439	
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Methanobrevibacter.s__Methanobrevibacter_oralis	0.0	0.0	8.24078	13.0389	0.0	0.0	
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Porphyromonas.s__Porphyromonas_catoniae	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Porphyromonas.s__Porphyromonas_gingivalis	3.92668	0.0	1.91889	0.0	0.0	0.0	2.11321
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Porphyromonas.s__Porphyromonas_somerae	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Porphyromonas.s__Porphyromonas_sp_oral_taxon_278	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Prevotella.s__Prevotella_loescheii	0.0	1.49891	0.0	2.40256	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Prevotella.s__Prevotella_saccharolytica	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Selenomonas.s__Selenomonas_noxia	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Selenomonas.s__Selenomonas_sp_oral_taxon_920	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Tannerella.s__Tannerella_forsythia	17.2602	10.7459	8.59643	6.31757	16.4405	10.9016	11.8978
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Tannerella.s__Tannerella_sp_oral_taxon_808	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	g__Tannerella.s__Tannerella_sp_oral_taxon_HOT_286	0.0	0.0	0.0	0.0	0.0	2.11521	0.0
1.1.1.271-RXN:	(expasy)	GDP-L-fucose	synthase	[1.1.1.271]	unclassified	0.0	0.0	0.0	0.0	0.0	2.96799	0.0



pathways table processing



Pathway abundance table

ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis	150.801	101.093	75.9777	53.2045	124.48	157.241	96.7058	29.4937	0	0	0	0	0	0	31.6389	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Aggregatibacter.s__Aggregatibacter_aphrophilus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Aggregatibacter.s__Aggregatibacter_segnsis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Bacteroidetes_unclassified.s__Bacteroidetes_oral_taxon_274	0	0	0	0	0	0	0	0	0	0	0	0	0	18.2989	14.9218	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Capnocytophaga.s__Capnocytophaga_gingivalis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Capnocytophaga.s__Capnocytophaga_granulosa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Capnocytophaga.s__Capnocytophaga_leadbetteri	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Capnocytophaga.s__Capnocytophaga_ochracea	0	0	0	0	0	0	0	0	0	2.85281	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Capnocytophaga.s__Capnocytophaga_sp_oral_taxon_863	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Capnocytophaga.s__Capnocytophaga_sp_oral_taxon_878	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Capnocytophaga.s__Capnocytophaga_sputigena	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Fusobacterium.s__Fusobacterium_nucleatum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Haemophilus.s__Haemophilus_paraainfluenzae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Haemophilus.s__Haemophilus_sp_HMSC71H05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Morococcus.s__Morococcus_cerebrosus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Prevotella.s__Prevotella_loescheii	0	0	0	0	3.04216	0	0	0	0	7.46696	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Streptococcus.s__Streptococcus_mitis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Streptococcus.s__Streptococcus_oralis	0	0	0	0	0	0	0	0	0	12.4256	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Streptococcus.s__Streptococcus_sp_HMSC070B10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Streptococcus.s__Streptococcus_sp_NLAE_zl_C503	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Streptococcus.s__Streptococcus_sp_SK643	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Streptococcus.s__Streptococcus_sp_oral_taxon_058	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis g__Tannerella.s__Tannerella_forsythia	49.2569	22.7856	23.131	16.2346	36.2448	29.9096	29.8649	0	0	0	0	0	0	0	0	0	0
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis unclassified	0	11.1506	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Most reads for genes assigned to this pathway come from *Tannerella forsythia*



Let's look at one approach to analyze
HUMANn output



On your VM

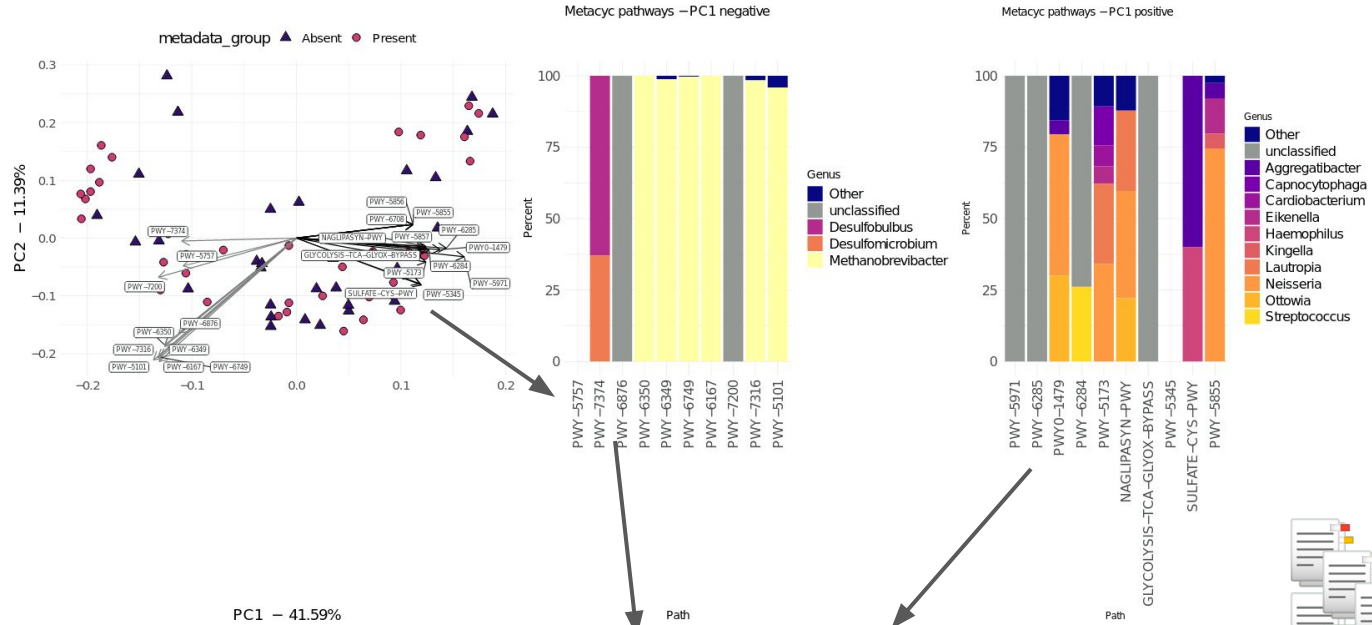
```
$ conda activate phylogenomics-functional  
$ rstudio /vol/volume/5c-functional-genomics/WSS_function_HUMAN3_GFs.Rmd
```



Now what?
-how to interpret
output

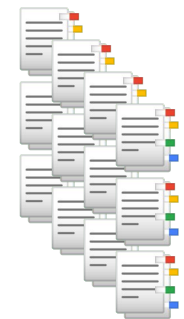


Now what?



PC1 – 41.59%

- 14: PWY-6350: archaeidylinositol biosynthesis
- 15: PWY-6349: CDP-archaeol biosynthesis
- 16: PWY-6749: CMP-legionaminate biosynthesis I
- 17: PWY-6167: flavin biosynthesis II (archaea)
- 18: PWY-7200: superpathway of pyrimidine deoxyribonucleoside salvage



Removed slides



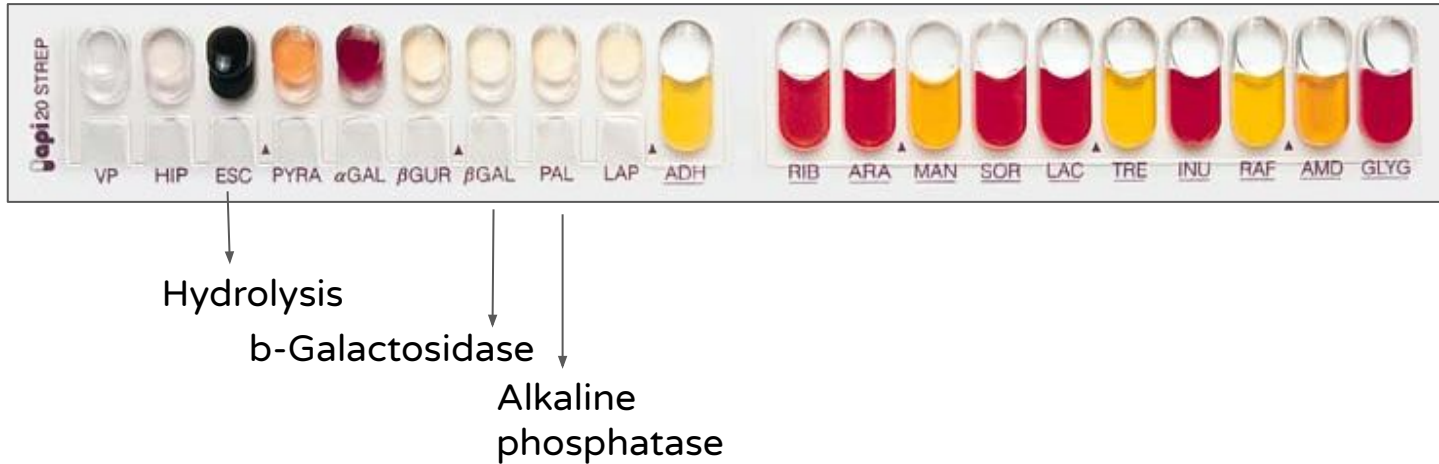
Species identification by function (new method)

Biochemical testing

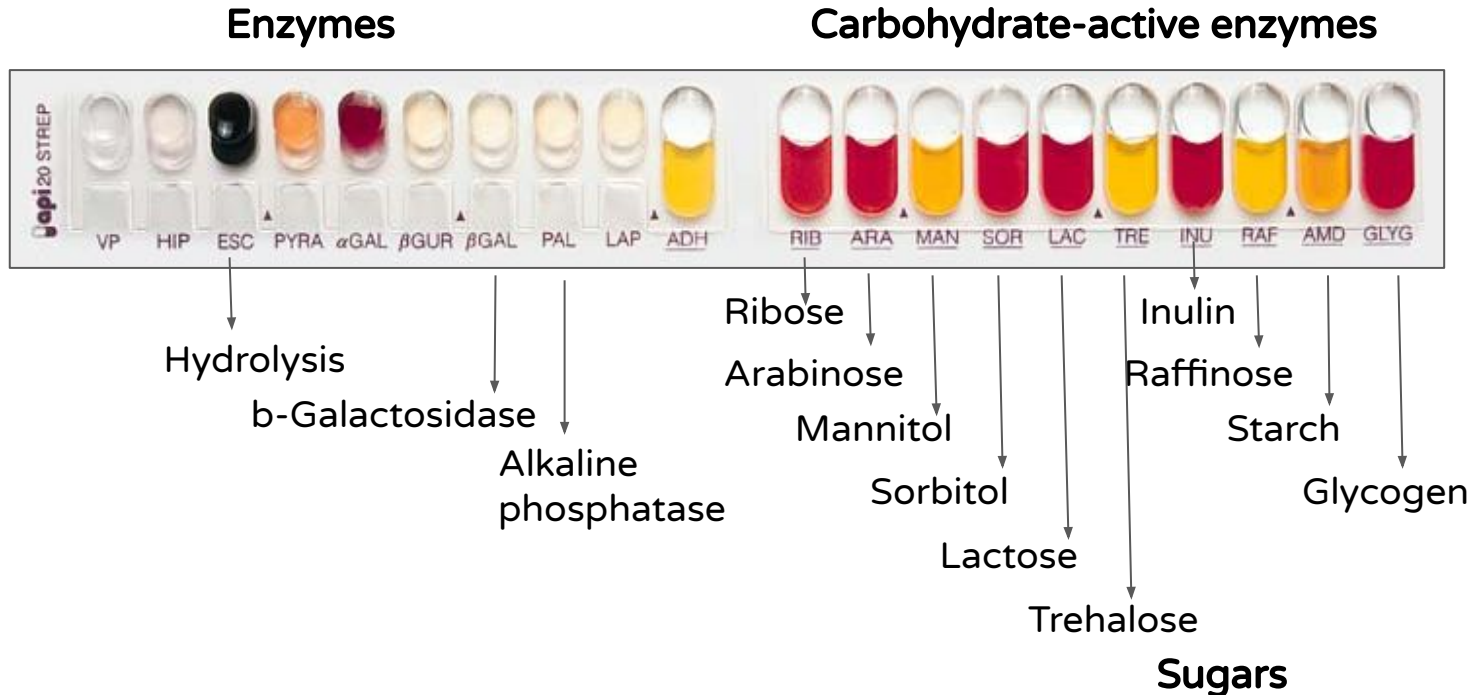


Species identification by function (new method)

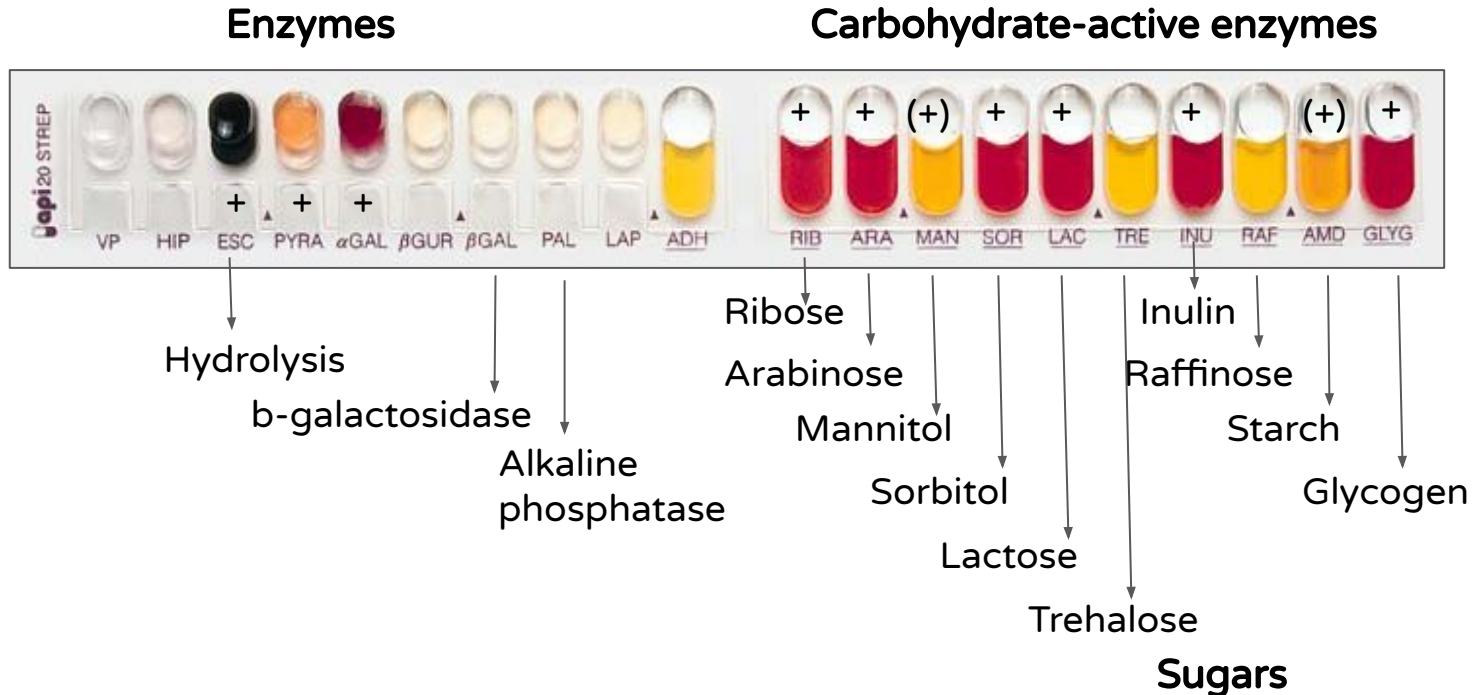
Enzymes



Species identification by function (new method)



Species identification by function (new method)



Species identification by function (new method)

API 20 STREP V7.0	VP	HIP	ESC	PYRA	AGAL	BGUR	BGAL	PAL	LAP	ADH	RIB	ARA	MAN	SOR	LAC	TRE	INU	RAF	AMD	GLYC	HEM	
<i>Abiotrophia defectiva</i>	25	0	15	99	100	0	100	0	92	0	0	0	0	0	98	100	5	92	99	0	0	
<i>Aerococcus urinae</i>	3	99	24	12	0	52	41	50	92	28	28	0	32	13	56	64	1	1	40	0	0	
<i>Aerococcus viridans 1</i>	13	50	96	54	33	16	37	1	5	1	83	33	85	70	83	99	33	41	70	33	1	
<i>Aerococcus viridans 2</i>	15	70	50	76	10	20	25	1	5	5	25	1	35	2	70	89	1	5	24	1	5	
<i>Aerococcus viridans 3</i>	22	88	99	40	85	48	14	14	1	1	8	2	82	5	91	99	37	99	14	1	1	
<i>Alloiococcus otitis</i>	0	25	0	100	0	3	100	1	90	0	0	0	0	0	0	20	0	0	0	0	0	
<i>Enterococcus avium</i>	99	60	99	94	15	0	24	1	99	0	99	40	100	95	95	99	1	40	15	0	1	
<i>Enterococcus durans</i>	100	43	100	97	32	2	76	1	91	100	99	15	2	0	84	76	0	0	56	0	18	
<i>Enterococcus faecalis</i>	99	46	99	97	1	0	21	4	99	92	98	1	98	92	92	100	0	1	96	2	1	
<i>Enterococcus faecium</i> *	94	43	99	95	42	1	89	1	97	93	85	70	78	18	84	98	15	10	60	3	1	
<i>Gardnerella vaginalis</i>	0	95	0	1	0	1	53	0	99	0	46	6	1	0	1	0	0	0	73	53	0	
<i>Gemella haemolysans</i>	25	0	0	70	0	0	1	84	40	1	1	0	20	10	5	2	0	0	10	5	1	
<i>Gemella morbillorum</i>	3	0	0	35	0	0	10	35	86	4	5	0	1	0	1	11	3	1	16	5	0	
<i>Globicatella sanguinis</i>	4	40	98	40	52	16	100	0	9	0	76	95	71	47	76	100	71	95	100	90	0	
<i>Granulicatella adiacens</i>	0	0	10	80	0	25	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	
<i>Lactococcus lactis ssp cremoris</i>	98	25	41	1	23	0	18	4	88	0	27	0	17	0	97	30	0	15	25	0	0	
<i>Lactococcus lactis ssp lactis</i>	90	40	99	35	3	0	35	3	96	95	95	15	45	1	72	87	4	5	90	3	1	
<i>Leuconostoc ssp</i>	91	1	60	5	55	0	65	2	70	10	37	35	29	4	35	65	0	42	11	0	0	
<i>Listeria spp</i>	97	79	98	0	0	0	0	0	85	0	6	0	0	0	49	92	1	1	72	0	26	
<i>Streptococcus agalactiae</i> **	100	99	1	1	4	79	1	96	99	99	98	0	1	1	50	87	0	1	35	4	75	
<i>Streptococcus anginosus</i>	100	0	100	0	44	0	1	99	100	100	0	0	33	0	99	88	0	44	97	0	37	
<i>Streptococcus bovis I</i>	99	1	100	1	34	2	1	0	100	0	0	1	97	1	100	100	65	98	98	98	1	
<i>Streptococcus bovis II 1</i>	100	0	1	0	58	0	0	0	100	0	0	0	0	0	90	0	0	97	97	97	0	
<i>Streptococcus bovis II 2</i>	100	2	100	0	89	97	99	0	100	0	0	0	0	0	100	100	0	72	31	5	0	
<i>Streptococcus bovis II 3</i>	99	1	100	0	99	0	6	0	100	0	0	0	0	0	100	6	6	100	93	0	0	
<i>Streptococcus bovis II 4</i>	98	1	100	0	97	2	10	0	100	1	1	32	1	1	98	40	84	99	99	97	0	
<i>Streptococcus canis</i>	0	1	25	4	95	1	80	100	100	100	100	0	0	0	99	1	0	1	99	0	100	
<i>Streptococcus constellatus</i>	100	1	27	0	0	0	5	99	100	100	0	0	0	0	10	72	0	0	12	0	61	
<i>Streptococcus dys.ssp dysgalactiae</i>	0	0	1	1	1	1	99	0	100	99	100	99	0	1	50	86	100	0	1	99	30	2
<i>Streptococcus dys.ssp equisimilis</i>	0	1	25	1	1	1	99	1	99	100	97	1	1	1	45	99	0	1	98	40	94	
<i>Streptococcus equi ssp equi</i>	1	0	1	0	0	100	0	100	100	100	0	0	0	0	0	1	0	0	100	100	100	
<i>Streptococcus equi ssp zoeepidemicus</i>	0	1	15	0	0	100	1	99	100	99	85	0	0	99	100	0	0	0	99	99	99	
<i>Streptococcus equinus</i>	100	0	95	0	28	0	1	1	100	0	0	0	30	0	25	7	25	15	17	10	0	
<i>Streptococcus group L</i>	1	75	1	0	0	100	1	100	100	100	100	0	0	0	75	100	0	0	100	98	94	
<i>Streptococcus intermedius</i>	100	0	87	0	0	0	44	99	100	100	0	0	0	0	99	99	3	3	99	0	40	
<i>Streptococcus mitis 1</i>	1	0	3	1	21	0	25	35	99	19	14	1	0	1	94	7	3	26	67	5	0	
<i>Streptococcus mitis 2</i>	0	0	3	0	31	0	35	50	100	99	1	0	1	0	100	1	1	31	84	0	0	
<i>Streptococcus mutans</i>	99	0	99	1	64	0	1	1	100	18	0	0	99	90	90	100	81	81	1	0	1	
<i>Streptococcus oralis</i>	0	0	1	1	50	0	46	72	100	5	1	0	1	0	99	32	1	72	96	0	0	

Treatment
-Antibiotic
-Duration



Species identification by function (new method)

API 20 STREP V7.0	VP	HIP	ESC	PYRA	AGAL	BGUR	BGAL	PAL	LAP	ADH	RIB	ARA	MAN	SOR	LAC	TRE	INU	RAF	AMD	GLYC	HEM	
<i>Abiotrophia defectiva</i>	25	0	15	99	100	0	100	0	92	0	0	0	0	0	98	100	5	92	99	0	0	
<i>Aerococcus urinae</i>	3	99	24	12	0	52	41	50	92	28	28	0	32	13	56	64	1	1	40	0	0	
<i>Aerococcus viridans 1</i>	13	50	96	54	33	16	37	1	5	1	83	33	85	70	83	99	33	41	70	33	1	
<i>Aerococcus viridans 2</i>	15	70	50	76	10	20	25	1	5	5	25	1	35	2	70	89	1	5	24	1	5	
<i>Aerococcus viridans 3</i>	22	88	99	40	85	48	14	14	1	1	8	2	82	5	91	99	37	99	14	1	1	
<i>Alloiococcus otitis</i>	0	25	0	100	0	3	100	1	90	0	0	0	0	0	0	20	0	0	0	0	0	
<i>Enterococcus avium</i>	99	60	99	94	15	0	24	1	99	0	99	40	100	95	95	99	1	40	15	0	1	
<i>Enterococcus durans</i>	100	43	100	97	32	2	76	1	91	100	99	15	2	0	84	76	0	0	56	0	18	
<i>Enterococcus faecalis</i>	99	46	99	97	1	0	21	4	99	92	98	1	98	92	92	100	0	1	96	2	1	
<i>Enterococcus faecium *</i>	94	43	99	95	42	1	89	1	97	93	85	70	78	18	84	98	15	10	60	3	1	
<i>Gardnerella vaginalis</i>	0	95	0	1	0	1	53	0	99	0	46	6	1	0	1	0	0	0	73	53	0	
<i>Gemella haemolysans</i>	25	0	0	70	0	0	1	84	40	1	1	0	20	10	5	2	0	0	10	5	1	
<i>Gemella morbillorum</i>	3	0	0	35	0	0	10	35	86	4	5	0	1	0	1	11	3	1	16	5	0	
<i>Globicatella sanguinis</i>	4	40	98	40	52	16	100	0	9	0	76	95	71	47	76	100	71	95	100	90	0	
<i>Granulicatella adiacens</i>	0	0	10	80	0	25	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	
<i>Lactococcus lactis ssp cremoris</i>	98	25	41	1	23	0	18	4	88	0	27	0	17	0	97	30	0	15	25	0	0	
<i>Lactococcus lactis ssp lactis</i>	90	40	99	35	3	0	35	3	96	95	95	15	45	1	72	87	4	5	90	3	1	
<i>Leuconostoc ssp</i>	91	1	60	5	55	0	65	2	70	10	37	35	29	4	35	65	0	42	11	0	0	
<i>Listeria spp</i>	97	79	98	0	0	0	0	0	85	0	6	0	0	0	49	92	1	1	72	0	26	
<i>Streptococcus agalactiae **</i>	100	99	1	1	4	79	1	96	99	99	98	0	1	1	50	87	0	1	35	4	75	
<i>Streptococcus anginosus</i>	100	0	100	0	44	0	1	99	100	100	0	0	33	0	99	88	0	44	97	0	37	
<i>Streptococcus bovis I</i>	99	1	100	1	34	2	1	0	100	0	0	1	97	1	100	100	65	98	98	98	1	
<i>Streptococcus bovis II 1</i>	100	0	1	0	58	0	0	0	100	0	0	0	0	0	90	0	0	97	97	97	0	
<i>Streptococcus bovis II 2</i>	100	2	100	0	89	97	99	0	100	0	0	0	0	0	100	100	0	72	31	5	0	
<i>Streptococcus bovis II 3</i>	99	1	100	0	99	0	6	0	100	0	0	0	0	0	100	6	6	100	93	0	0	
<i>Streptococcus bovis II 4</i>	98	1	100	0	97	2	10	0	100	1	1	32	1	1	98	40	84	99	99	97	0	
<i>Streptococcus canis</i>	0	1	25	4	95	1	80	100	100	100	100	0	0	0	99	1	0	1	99	0	100	
<i>Streptococcus constellatus</i>	100	1	27	0	0	0	5	99	100	100	0	0	0	0	10	72	0	0	12	0	61	
<i>Streptococcus dys.ssp dysgalactiae</i>	0	0	1	1	1	1	99	0	100	99	100	99	0	1	50	86	100	0	1	99	30	2
<i>Streptococcus dys.ssp equisimilis</i>	0	1	25	1	1	1	99	1	99	100	97	97	1	1	1	45	99	0	1	98	40	94
<i>Streptococcus equi ssp equi</i>	1	0	1	0	0	0	100	0	100	100	100	0	0	0	0	1	0	0	100	100	100	
<i>Streptococcus equi ssp zoeepidemicus</i>	0	1	15	0	0	0	100	1	99	100	99	85	0	0	99	100	0	0	0	99	99	99
<i>Streptococcus equinus</i>	100	0	95	0	28	0	1	1	100	0	0	0	30	0	25	7	25	15	17	10	0	
<i>Streptococcus group L</i>	1	75	1	0	0	0	100	1	100	100	100	0	0	0	75	100	0	0	100	98	94	
<i>Streptococcus intermedius</i>	100	0	87	0	0	0	44	99	100	100	0	0	0	0	99	99	3	3	99	0	40	
<i>Streptococcus mitis 1</i>	1	0	3	1	21	0	25	35	99	19	14	1	0	1	94	7	3	26	67	5	0	
<i>Streptococcus mitis 2</i>	0	0	3	0	31	0	35	50	100	99	1	0	1	0	100	1	1	31	84	0	0	
<i>Streptococcus mutans</i>	99	0	99	1	64	0	1	1	100	18	0	0	99	90	90	100	81	81	1	0	1	
<i>Streptococcus oralis</i>	0	0	1	1	50	0	46	72	100	5	1	0	1	0	99	32	1	72	96	0	0	

Faster than sequencing*
but not as precise

