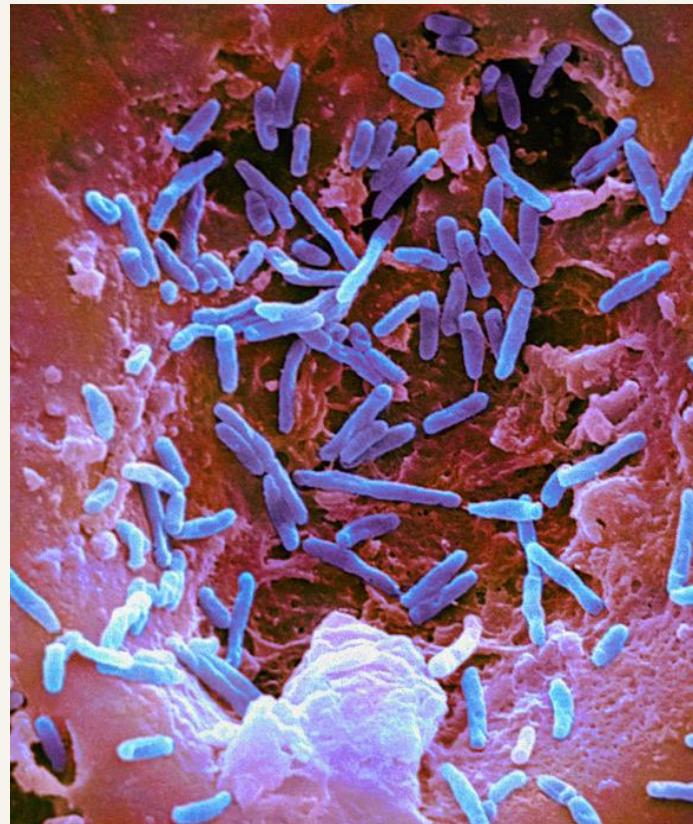

Exploring the Links Between Diseases and the Human Microbiome

By Jacqueline Daniel, Amanda Bell, Vivek Kanpa, and Zoe Morris

PROJECT IMPORTANCE

- Our microbiome plays a large role in our health
 - Well known and understood roles:
 - Digestion
 - Immunity
 - Lesser known and understood roles:
 - Neurotransmitter production
 - Micro-protein production
 - Health regulation
- Understanding how bacteria levels and diversity in our microbiome can inform treatment on many diseases and show novel connections



OUR QUESTIONS:

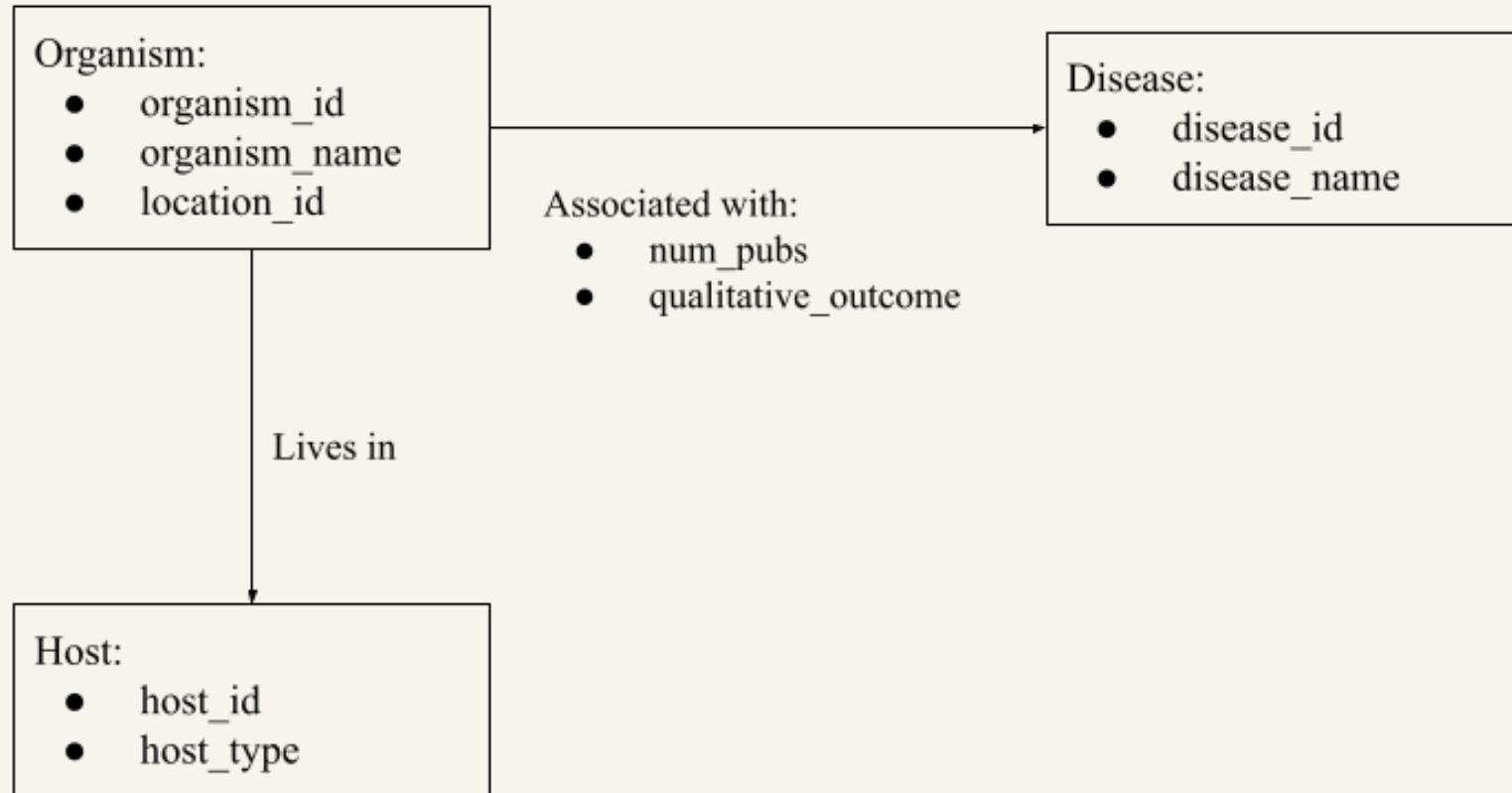
- What are the diseases that are highly associated with bacteria?
- Which bacterium is linked to the greatest number of diseases?
- Are there diseases that are linked to covid-19 through a type of bacteria?
- Which bacteria are linked to diabetes ?
 - Are there common bacteria among the different types of diabetes?
- What diseases are associated with bacteria linked to cancer?
- Which bacteria are linked to neurodegenerative diseases?
 - Alzheimer's Disease, Parkinson's Disease, ALS

THE DATASET

- Disbiome Database
 - Rows:
 - Represent a publication
 - Columns:
 - Disease name, organism name,
 - qualitative outcome, sample type

disease_name	disease_id	meddra_level	meddra_id	location_id	organism_name	control_value	control_name	organism_id	sample_name	organism_ncbi_id	subject_value	ratio	host_id	host_type
Cystic Fibrosis	14	preferred_term	10011762	10	Firmicutes		Healthy control	410	Faeces	1239.0			182	Human
Crohn's Disease	3	preferred_term	10011401	10	Faecalibacterium prausnitzii		Healthy control	6	Faeces	853.0			57	Human
Crohn's Disease	3	preferred_term	10011401	10	Ruminococcus albus		Healthy control	208	Faeces	1264.0			57	Human
Crohn's Disease	3	preferred_term	10011401	10	Bacteroides fragilis		Healthy control	240	Faeces	817.0			57	Human
► Crohn's Disease	3	preferred_term	10011401	10	Clostridioides difficile		Healthy control	194	Faeces	1496.0			57	Human
Systemic sclerosis	45	low_level_term	10042953	21	Fusobacterium		Healthy control	77	Cecum lavage sample	68766.0			120	Human
Lung cancer	237	preferred_term	10023774	10	Enterobacteriaceae		Healthy control	232	Faeces	543.0			1322	Human
Primary biliary cholangitis	165	preferred_term	10008604	10	Dialister		Healthy control	80	Faeces	39948.0			839	Human
Colon adenomas	77	preferred_term	10048832	10	Comamonas		Colorectal carcinoma patient	629	Colonoscopy aspirate	283.0			907	Human
Colorectal cancer	22	preferred_term	10061451	10	Escherichia coli		Healthy control	51	Faeces	562.0			936	Human
Chronic Obstructive Pul...	15	preferred_term	10009033	15	Fusobacterium		Healthy individuals with nor...	77	Bronchoalveolar lava...	68766.0			267	Human
Bacterial vaginosis	59	preferred_term	10004055	32	Lactobacillus		Healthy control	1	Vaginal swab	1591.0			287	Human
Bacterial vaginosis	59	preferred_term	10004055	32	Atopobium		Healthy control	321	Vaginal swab	1380.0			287	Human
Bacterial vaginosis	59	preferred_term	10004055	32	Sneathia		Healthy control	461	Vaginal swab	168808.0			287	Human
Bacterial vaginosis	59	preferred_term	10004055	32	Eggerthella		Healthy control	412	Vaginal swab	84111.0			287	Human
Bacterial vaginosis	59	preferred_term	10004055	32	Megasphaera		Healthy control	453	Vaginal swab	906.0			287	Human
Chronic kidney disease	142	preferred_term	10064848	10	Roseburia		Healthy control	14	Faeces	841.0			439	Human
Chronic kidney disease	142	preferred_term	10064848	10	Prevotella		Healthy control	44	Faeces	59823.0			439	Human
Irritable Bowel Syndrome	5	preferred_term	10023002	10	Butyrivibrio		Healthy control	224	Faeces	574607.0			457	Human

Data Representation



PRELIMINARY SQL AND NEO4J QUERIES:

```
create view disbiome_view as
select disease_name, disease_id,
       location_id, organism_name,
       control_name, organism_id,
       host_id, host_type,
       count(organism_name) as num_publications
from disbiome
group by disease_name, disease_id,
         location_id, organism_name,
         control_name, organism_id,
         host_id, host_type
order by num_publications DESC;
```

```
1 call apoc.load.jdbc("jdbc:mysql://localhost:3306/disbiome?serverTimezone=EST5EDT&user=ds4300user&password=ds4300password",
                      "disbiome_view") yield row
2 merge (o:Organism {organism_id: row.organism_id,
                     organism:row.organism_name,location:row.location_id})
3 merge (d:Disease {disease:row.disease_name, disease_id:row.disease_id})
4 merge (h:Host {host_id:row.host_id, host_type:row.host_type,})
5 merge (o)-[:Assoc {publications:row.num_publications,
                     outcome:row.qualitative_outcome}]->(d)
6 merge (o)-[:lives_in]->(h)
7 return o,d,h
```

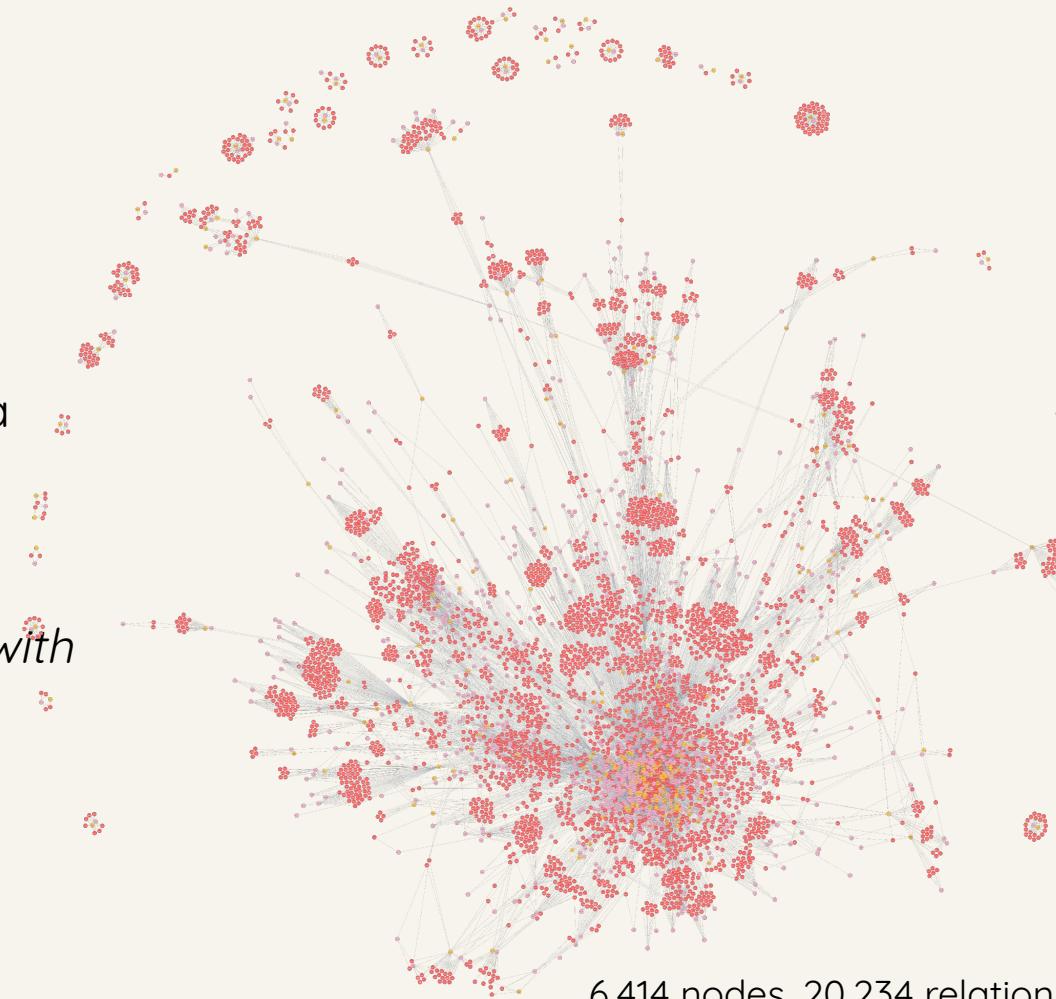
Graph:

Nodes:

- Yellow: Disease
- Red: Organism/Bacteria
- Pink: Host

Relationships:

- A bacteria is *associated with* a Disease
- A bacteria *lives in* a host



Which bacteria are most associated with diseases?

organism_name	num_publications	sample_name
Prevotella	242	Vaginal fluid
Streptococcus	212	Saliva
Lactobacillus	170	Faeces
Bacteroides	159	Faeces
Bifidobacterium	145	Faeces
Veillonella	137	Faeces
Faecalibacterium	134	Faeces
Roseburia	121	Faeces
Fusobacterium	114	Buccal swab
Ruminococcus	108	Faeces
Blautia	106	Faeces
Faecalibacteriu...	101	Faeces
Lachnospiraceae	101	swab left an...
Enterococcus	99	Faeces
Parabacteroides	94	Faeces
Coprococcus	93	Faeces

Which diseases have been studied most with regards to bacteria?

disease_name	num_publications
Crohn's Disease	481
Colorectal cancer	377
Parkinson's Disease	296
periodontitis	288
Obesity	268
Type 2 Diabetes	246
Caries	243
Ulcerative Colitis	220
HIV infection	206
Cirrhosis	189
Cystic Fibrosis	173
Irritable Bowel Syn...	167
Autism spectrum di...	165
Chronic Obstructiv...	155
Psoriasis	147
Autism	136

Which diseases are most linked to specific gut bacteria?

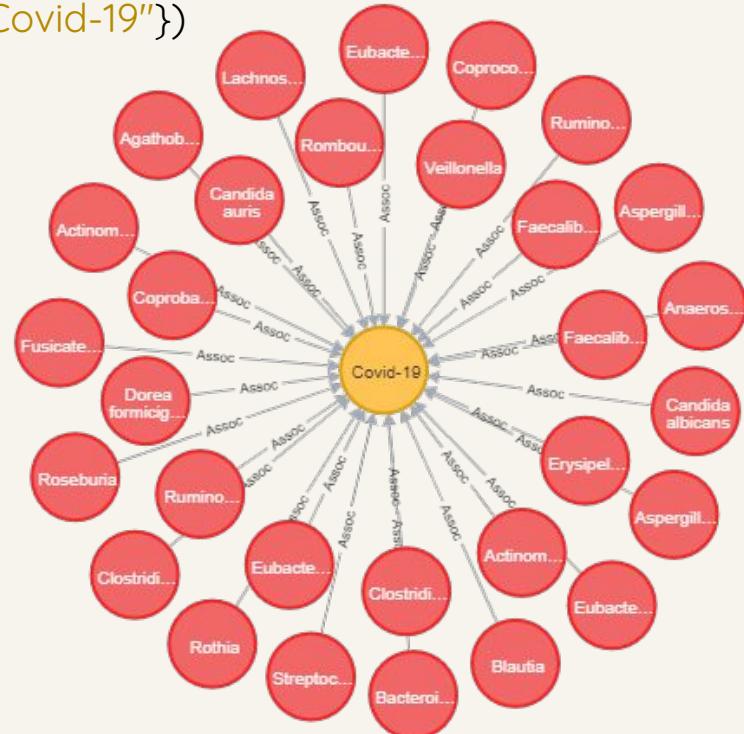
```
match
(o:Organism)-[a:Assoc]->(d:
Disease)
where a.publications > 5
return o.organism as
organism,
sum(a.publications) as
npubs, d.disease as disease,
a.outcome as outcome
order by npubs desc
```

"organism"	"npubs"	"disease"	"outcome"
"Fusobacterium"	17	"Colorectal cancer"	"Elevated"
"Fusobacterium nucleatum"	14	"Colorectal cancer"	"Elevated"
"Faecalibacterium prausnitzii"	13	"Crohn's Disease"	"Reduced"
"Porphyromonas gingivalis"	13	"periodontitis"	"Elevated"
"Tannerella forsythia"	12	"periodontitis"	"Elevated"
"Streptococcus"	10	"Colorectal cancer"	"Elevated"
"Megasphaera"	10	"Bacterial vaginosis"	"Elevated"
"Bacteroides"	9	"Colorectal cancer"	"Reduced"
"Roseburia"	9	"Crohn's Disease"	"Reduced"
"Prevotella"	9	"Bacterial vaginosis"	"Elevated"
"Bifidobacterium"	9	"Obesity"	"Reduced"
"Lactobacillus"	9	"Obesity"	"Reduced"

Which bacteria are linked to Covid-19?

```
match (o:Organism)-[a:Assoc]->(d:Disease {disease:"Covid-19"})  
return o, a, d
```

"organism_name"	"disease_name"
"Streptococcus"	"Covid-19"
"Bacteroides nordii"	"Covid-19"
"Blautia"	"Covid-19"
"Actinomyces"	"Covid-19"
"Eubacterium ventriosum"	"Covid-19"
"Clostridium ramosum"	"Covid-19"
"Aspergillus niger"	"Covid-19"
"Faecalibacterium prausnitzii"	"Covid-19"
"Candida albicans"	"Covid-19"

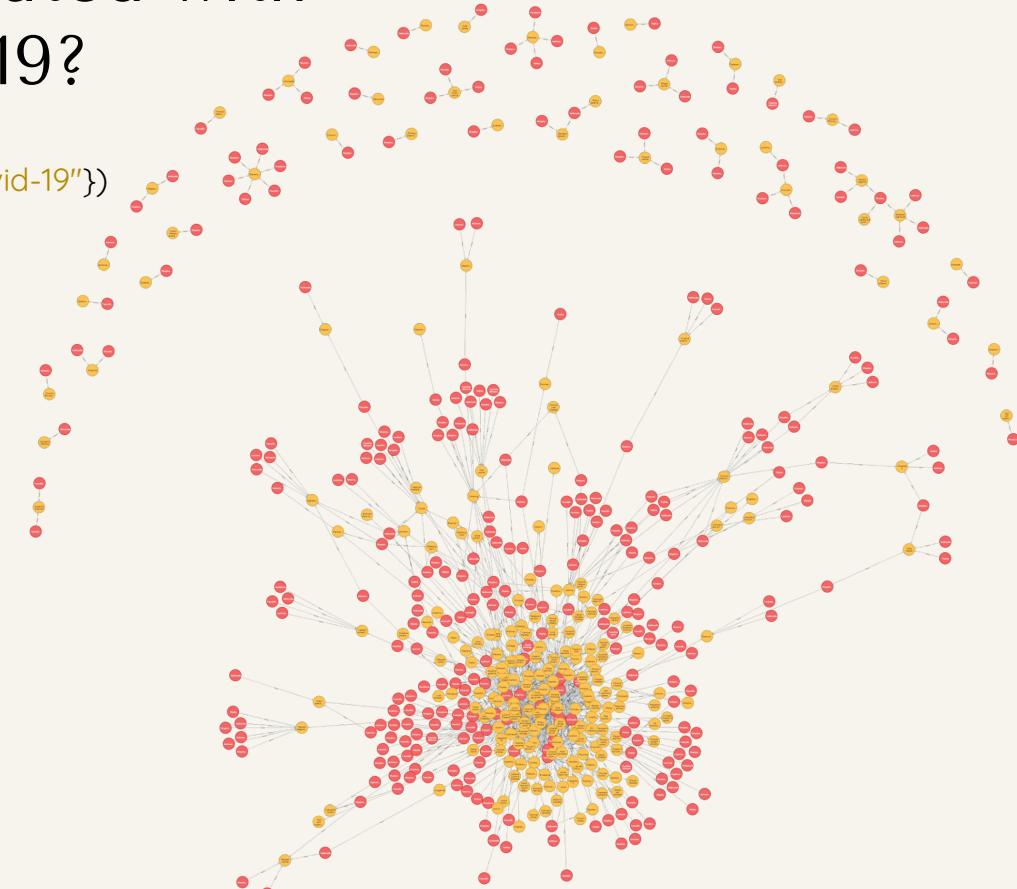


30 nodes, 31 relationships

What diseases are associated with bacteria linked to Covid-19?

```
call {
  match (o:Organism)-[a:Assoc]->(d:Disease {disease:"Covid-19"})
  return collect(o.organism) as bacteria
}
match (o:Organism)-[a:Assoc]->(d:Disease)
where o.organism in bacteria
return o, a, d
```

"disease_name"
"Cystic Fibrosis"
"Crohn's Disease"
"Systemic sclerosis"
"Colon adenomas"
"Colorectal cancer"
"Chronic Obstructive Pulmonary Disease"
"Bacterial vaginosis"
"Chronic kidney disease"



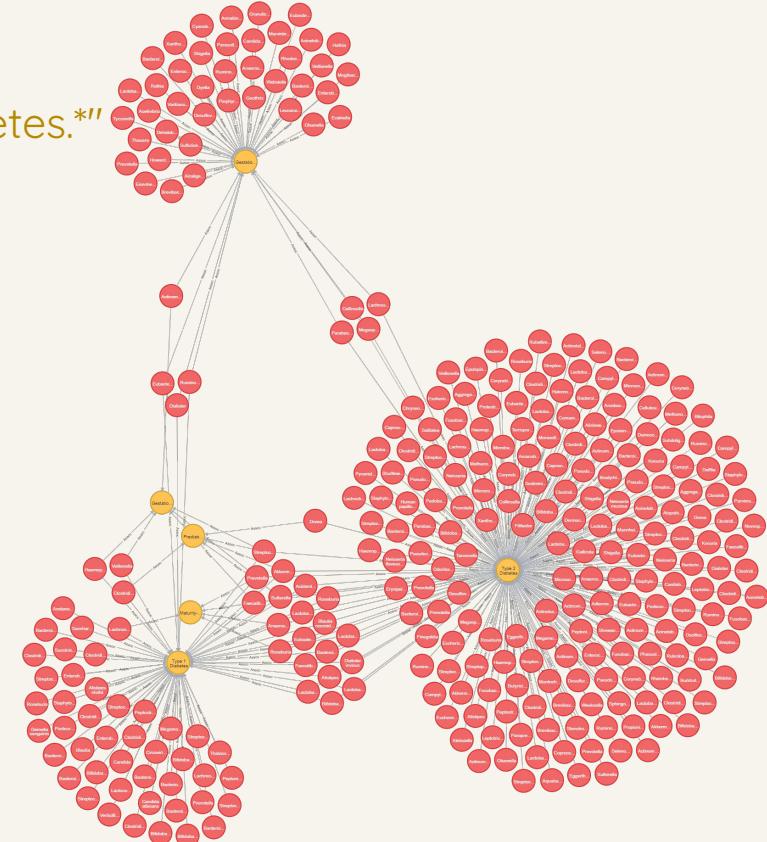
Which bacteria are linked to the different types of diabetes? Do they all share any common bacteria?

```
match (o:Organism)-[a:Assoc]->(d:Disease)
```

```
where d.disease =~ ".*diabetes.*" or d.disease =~ ".*Diabetes.*"
```

```
return o, a, d
```

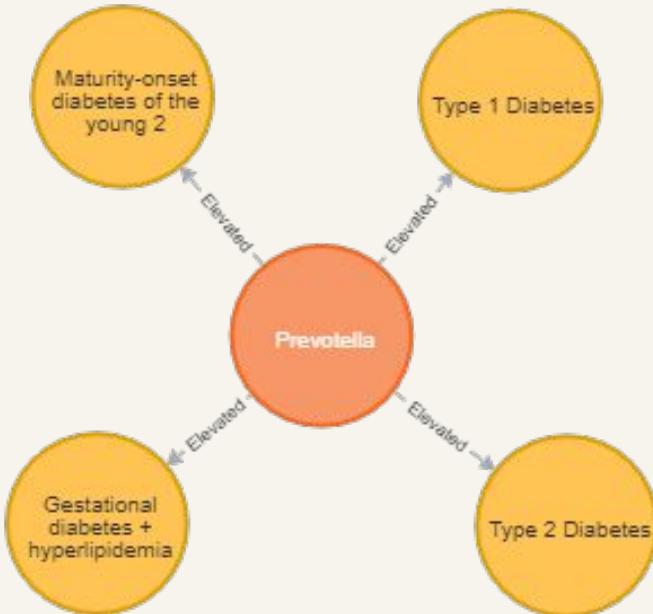
"organism"	"disease"
"Olsenella"	"Type 2 Diabetes"
"Ruminococcus gnavus"	"Type 2 Diabetes"
"Phascolarctobacterium"	"Type 2 Diabetes"
"Streptococcus"	"Type 2 Diabetes"
"Dorea"	"Type 2 Diabetes"
"Ruminococcaceae"	"Type 2 Diabetes"
"Verticillium"	"Type 1 Diabetes"
"Acidaminococcus"	"Type 1 Diabetes"
"Roseburia faecis"	"Type 1 Diabetes"
"Bacteroides vulgatus"	"Type 1 Diabetes"
"Akkermansia muciniphila"	"Type 1 Diabetes"
"Bifidobacterium longum"	"Type 1 Diabetes"



319 nodes, 356 relationships

What is the greatest number of diabetes types that share common relationships with bacteria?

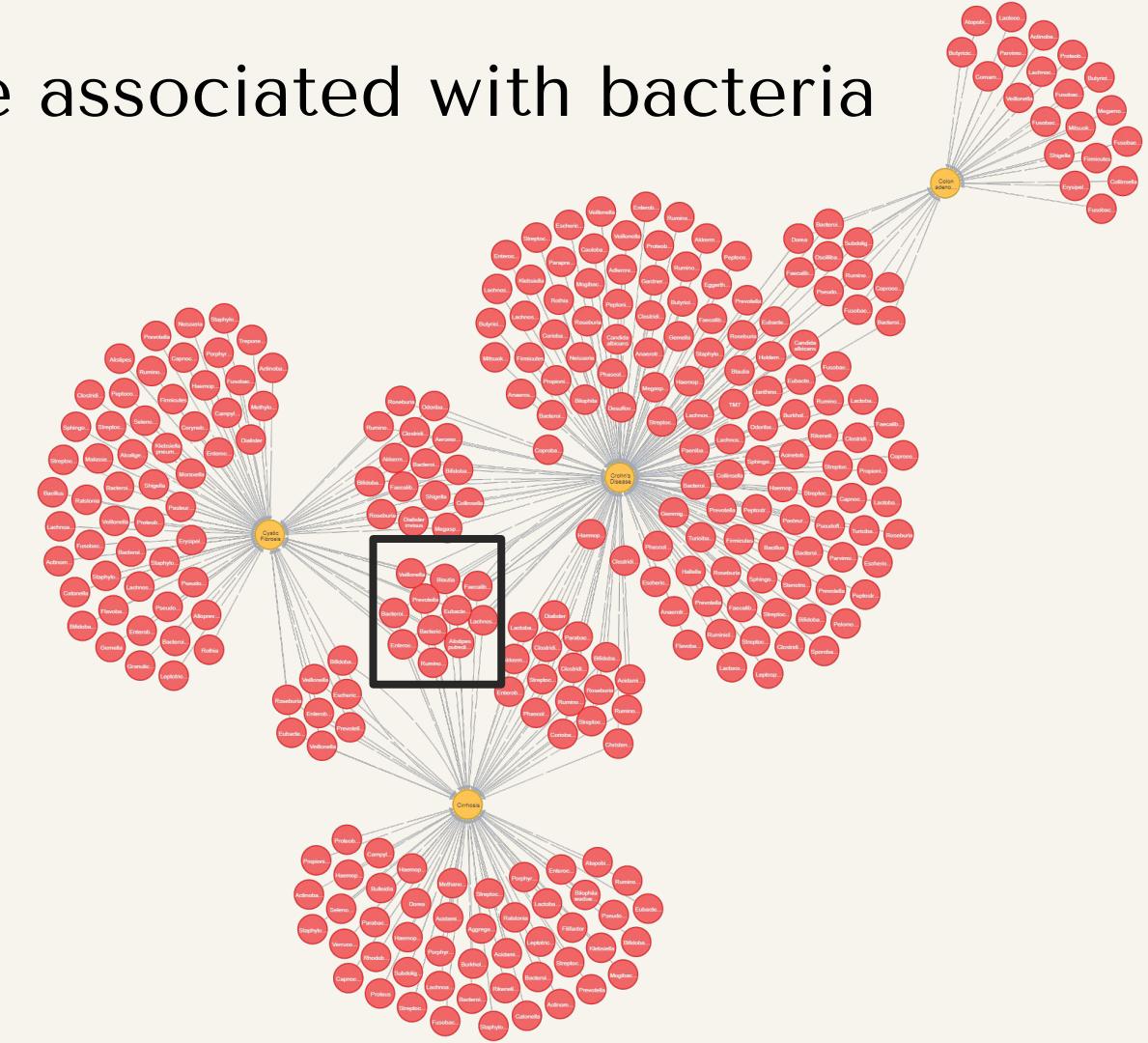
```
match (o:Organism)-[a]->(d:Disease{disease:"Type 2 Diabetes"})  
match(o)-[a2]->(d2:Disease{disease:"Type 1 Diabetes"})  
match(o)-[a3:Assoc]->(d3:Disease{disease:"Maturity-onset diabetes of the young 2"})  
match(o)-[a4:Assoc]->(d4:Disease{disease:"Gestational diabetes + hyperlipidemia"})  
return o, a, d, d2, a2, d3, a3, d4, a4
```



What diseases are associated with bacteria linked to cancer?

- Cystic fibrosis*
- Crohn's Disease
- Colon adenomas*
- Cirrhosis*

```
call {
  match (o:Organism)-[a:Assoc]->(d:Disease)
  where d.disease =~ ".*cancer.*" or d.disease
  =~ ".Cancer.*"
  return collect(o.organism) as bacteria
}
match (o:Organism)-[a:Assoc]->(d:Disease)
where o.organism in bacteria
return o, a, d
```



Further Research

- What microorganisms are implicated in multiple cancer types?
- Neurodegenerative diseases
- Look at other groups of diseases
 - Autoimmune, metabolic, cardiovascular, etc.
- Antibiotic and antiviral targets for reducing microbiological load
- Treatment to restore gut microbiome and targets for disease mitigation

Thank you!

Any
Questions?

