

Genome Information For Sequenced Organism

November 30, 2023

1 BY JAYANT SINGH PARIHAR

2 Project - Genome Information For Sequenced Organism

2.1 Import Libraries

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

2.2 Dataframe vs Series

```
[2]: data = pd.read_csv("eukaryotes.csv")
```

```
[3]: data
```

```
[3]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold

5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%		Replicons \
0	167.6760	64.5000		NaN
1	119.6680	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...	
2	412.9240	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...	
3	978.9720	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...	
4	823.7860	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...	
...
5126	12.0392	38.1000		NaN
5127	11.7086	38.1000		NaN
5128	11.8194	38.0000		NaN
5129	11.7081	38.1000		NaN
5130	11.8189	38.1000		NaN

	WGS	Scaffolds	CDS	Release Date \
0	AHAL01	7795	38554	2013-04-19T00:00:00Z
1	NaN	7	48350	2001-08-13T00:00:00Z
2	APN001	2187	57661	2011-08-12T00:00:00Z
3	ACUP02	1191	71525	2010-01-05T00:00:00Z
4	AEKE02	3224	36010	2010-12-10T00:00:00Z
...
5126	FXDE01	1925	0	2017-04-21T00:00:00Z
5127	FXCK01	590	0	2017-04-21T00:00:00Z
5128	FXC001	2013	0	2017-04-21T00:00:00Z
5129	FXCP01	1028	0	2017-04-21T00:00:00Z
5130	FXCG01	799	0	2017-04-21T00:00:00Z

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...

```

2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
5126                                     NaN
5127                                     NaN
5128                                     NaN
5129                                     NaN
5130                                     NaN

```

[5131 rows x 16 columns]

```
[4]: type(data)
```

```
[4]: pandas.core.frame.DataFrame
```

```
[5]: data["#Organism Name"]
```

```

[5]: 0      Emiliania huxleyi CCMP1516
      1      Arabidopsis thaliana
      2      Medicago truncatula
      3      Glycine max
      4      Solanum lycopersicum
      ...
      5126      Saccharomyces cerevisiae
      5127      Saccharomyces cerevisiae
      5128      Saccharomyces cerevisiae
      5129      Saccharomyces cerevisiae
      5130      Saccharomyces cerevisiae
      Name: #Organism Name, Length: 5131, dtype: object

```

```
[6]: type(data["#Organism Name"])
```

```
[6]: pandas.core.series.Series
```

2.3 Inspecting Data in Pandas

```
[7]: data.head()
```

```

[7]:      #Organism Name      Organism Groups      Strain \
0  Emiliania huxleyi CCMP1516  Eukaryota;Protists;Other Protists  CCMP1516
1    Arabidopsis thaliana      Eukaryota;Plants;Land Plants      NaN
2    Medicago truncatula      Eukaryota;Plants;Land Plants      A17
3      Glycine max      Eukaryota;Plants;Land Plants      NaN
4    Solanum lycopersicum      Eukaryota;Plants;Land Plants      NaN

      BioSample  BioProject      Assembly      Level  Size(Mb)      GC% \

```

0	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold	167.676	64.5000
1	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome	119.668	36.0528
2	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome	412.924	34.0470
3	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome	978.972	35.1208
4	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome	823.786	35.7097

		Replicons	WGS	Scaffolds \
0		NaN	AHAL01	7795
1	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN		7
2	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001		2187
3	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02		1191
4	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02		3224

	CDS	Release Date \
0	38554	2013-04-19T00:00:00Z
1	48350	2001-08-13T00:00:00Z
2	57661	2011-08-12T00:00:00Z
3	71525	2010-01-05T00:00:00Z
4	36010	2010-12-10T00:00:00Z

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...

```
[8]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5131 entries, 0 to 5130
Data columns (total 16 columns):
#   Column              Non-Null Count  Dtype
---  -
0   #Organism Name      5131 non-null   object
1   Organism Groups     5131 non-null   object
2   Strain              4162 non-null   object
3   BioSample           5109 non-null   object
4   BioProject          5131 non-null   object
5   Assembly            5131 non-null   object
6   Level              5131 non-null   object
```

```

7   Size(Mb)          5131 non-null    float64
8   GC%               5131 non-null    float64
9   Replicons         1068 non-null    object
10  WGS               4792 non-null    object
11  Scaffolds         5131 non-null    int64
12  CDS               5131 non-null    int64
13  Release Date      5077 non-null    object
14  GenBank FTP       5125 non-null    object
15  RefSeq FTP        859 non-null     object
dtypes: float64(2), int64(2), object(12)
memory usage: 641.5+ KB

```

```
[9]: data.describe()
```

```

[9]:          Size(Mb)          GC%          Scaffolds          CDS
count    5131.000000    5131.000000    5.131000e+03    5131.000000
mean       393.103324     42.036497    4.630382e+04    6205.094913
std       1206.196248     10.584974    2.427502e+05    11930.220757
min         0.662517      0.000000    1.000000e+00      0.000000
25%        20.279550     38.100000    1.960000e+02      0.000000
50%        39.491700     42.000000    1.591000e+03      0.000000
75%       235.581000     49.000000    1.034250e+04     8392.000000
max      32393.600000     71.400000    6.870110e+06   123467.000000

```

```
[10]: data.describe(include="O")
```

```

[10]:          #Organism Name          Organism Groups  Strain \
count                5131                5131    4162
unique                3458                19    3962
top    Saccharomyces cerevisiae  Eukaryota;Fungi;Ascomycetes  Race 77
freq                344                2177    14

          BioSample  BioProject          Assembly  Level \
count            5109            5131            5131    5131
unique            5047            3859            5131      4
top    SAMN02743421  PRJNA323691  GCA_000372725.1  Scaffold
freq              9            157              1    3350

          Replicons  WGS \
count            1068    4792
unique            1068    4790
top    chromosome 1:NC_003070.9/CP002684.1; chromosom...  ABBA01
freq              1      2

          Release Date \
count                5077
unique                1580
top    2016-09-12T00:00:00Z

```

```

freq                                161

                                GenBank FTP \
count                                5125
unique                              5125
top    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
freq                                1

                                RefSeq FTP
count                                859
unique                              859
top    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
freq                                1

```

2.4 Selecting Columns in Pandas

```
[11]: data.columns
```

```
[11]: Index(['#Organism Name', 'Organism Groups', 'Strain', 'BioSample',
            'BioProject', 'Assembly', 'Level', 'Size(Mb)', 'GC%', 'Replicons',
            'WGS', 'Scaffolds', 'CDS', 'Release Date', 'GenBank FTP', 'RefSeq FTP'],
            dtype='object')
```

```
[12]: # Sophisticated way of selecting a column, bracket notation
      data["BioSample"]
```

```
[12]: 0      SAMN02744062
      1      SAMN03081427
      2      SAMN02299339
      3      SAMN00002965
      4      SAMN02981290
      ...
      5126   SAMEA3110475
      5127   SAMEA3110493
      5128   SAMEA3110473
      5129   SAMEA3110476
      5130   SAMEA3110463
      Name: BioSample, Length: 5131, dtype: object
```

```
[13]: # lazy way of selecting, the dot notation
      data.BioSample
```

```
[13]: 0      SAMN02744062
      1      SAMN03081427
      2      SAMN02299339
      3      SAMN00002965
      4      SAMN02981290
```

```

...
5126 SAMEA3110475
5127 SAMEA3110493
5128 SAMEA3110473
5129 SAMEA3110476
5130 SAMEA3110463
Name: BioSample, Length: 5131, dtype: object

```

```
[14]: data["Organism Groups"]
```

```

[14]: 0      Eukaryota;Protists;Other Protists
      1      Eukaryota;Plants;Land Plants
      2      Eukaryota;Plants;Land Plants
      3      Eukaryota;Plants;Land Plants
      4      Eukaryota;Plants;Land Plants
      ...
5126      Eukaryota;Fungi;Ascomycetes
5127      Eukaryota;Fungi;Ascomycetes
5128      Eukaryota;Fungi;Ascomycetes
5129      Eukaryota;Fungi;Ascomycetes
5130      Eukaryota;Fungi;Ascomycetes
Name: Organism Groups, Length: 5131, dtype: object

```

```
[15]: data[["Release Date" , "#Organism Name", "Organism Groups"]]
```

```

[15]:      Release Date      #Organism Name \
0      2013-04-19T00:00:00Z  Emiliana huxleyi CCMP1516
1      2001-08-13T00:00:00Z      Arabidopsis thaliana
2      2011-08-12T00:00:00Z      Medicago truncatula
3      2010-01-05T00:00:00Z      Glycine max
4      2010-12-10T00:00:00Z      Solanum lycopersicum
...
5126  2017-04-21T00:00:00Z      Saccharomyces cerevisiae
5127  2017-04-21T00:00:00Z      Saccharomyces cerevisiae
5128  2017-04-21T00:00:00Z      Saccharomyces cerevisiae
5129  2017-04-21T00:00:00Z      Saccharomyces cerevisiae
5130  2017-04-21T00:00:00Z      Saccharomyces cerevisiae

      Organism Groups
0      Eukaryota;Protists;Other Protists
1      Eukaryota;Plants;Land Plants
2      Eukaryota;Plants;Land Plants
3      Eukaryota;Plants;Land Plants
4      Eukaryota;Plants;Land Plants
...
5126      Eukaryota;Fungi;Ascomycetes
5127      Eukaryota;Fungi;Ascomycetes

```

```

5128      Eukaryota;Fungi;Ascomycetes
5129      Eukaryota;Fungi;Ascomycetes
5130      Eukaryota;Fungi;Ascomycetes

```

[5131 rows x 3 columns]

```
[16]: data._get_numeric_data()
```

```

[16]:      Size(Mb)      GC%  Scaffolds    CDS
0      167.6760  64.5000      7795  38554
1      119.6680  36.0528         7  48350
2      412.9240  34.0470      2187  57661
3      978.9720  35.1208      1191  71525
4      823.7860  35.7097      3224  36010
...
5126     12.0392  38.1000      1925     0
5127     11.7086  38.1000       590     0
5128     11.8194  38.0000      2013     0
5129     11.7081  38.1000      1028     0
5130     11.8189  38.1000       799     0

```

[5131 rows x 4 columns]

```
[17]: data.select_dtypes("object")
```

```

[17]:      #Organism Name      Organism Groups \
0      Emiliana huxleyi CCMP1516  Eukaryota;Protists;Other Protists
1      Arabidopsis thaliana      Eukaryota;Plants;Land Plants
2      Medicago truncatula      Eukaryota;Plants;Land Plants
3      Glycine max      Eukaryota;Plants;Land Plants
4      Solanum lycopersicum      Eukaryota;Plants;Land Plants
...
5126      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5127      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5128      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5129      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5130      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes

      Strain      BioSample  BioProject      Assembly      Level \
0      CCMP1516  SAMN02744062  PRJNA77753  GCA_000372725.1      Scaffold
1      NaN      SAMN03081427  PRJNA10719  GCA_000001735.1      Chromosome
2      A17      SAMN02299339  PRJNA10791  GCA_000219495.2      Chromosome
3      NaN      SAMN00002965  PRJNA19861  GCA_000004515.3      Chromosome
4      NaN      SAMN02981290      PRJNA119  GCA_000188115.2      Chromosome
...
5126      NaN      SAMEA3110475  PRJEB7675  GCA_900178465.1      Scaffold
5127      ZP848  SAMEA3110493  PRJEB7675  GCA_900178475.1      Scaffold

```


5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

		Replicons	WGS	\
0		NaN	AHAL01	
1	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN		
2	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001		
3	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02		
4	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02		
...		
5126		NaN	FXDE01	
5127		NaN	FXCK01	
5128		NaN	FXC001	
5129		NaN	FXCP01	
5130		NaN	FXCG01	

	Release Date	GenBank FTP	\
0	2013-04-19T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
1	2001-08-13T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
2	2011-08-12T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
3	2010-01-05T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
4	2010-12-10T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
...	
5126	2017-04-21T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
5127	2017-04-21T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
5128	2017-04-21T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
5129	2017-04-21T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
5130	2017-04-21T00:00:00Z	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
5126	NaN
5127	NaN
5128	NaN
5129	NaN
5130	NaN

[5131 rows x 12 columns]

```
[18]: data.select_dtypes("float")
```

```
[18]:      Size(Mb)      GC%
      0      167.6760  64.5000
      1      119.6680  36.0528
      2      412.9240  34.0470
      3      978.9720  35.1208
      4      823.7860  35.7097
      ...      ...      ...
    5126      12.0392  38.1000
    5127      11.7086  38.1000
    5128      11.8194  38.0000
    5129      11.7081  38.1000
    5130      11.8189  38.1000
```

[5131 rows x 2 columns]

```
[19]: data.select_dtypes("int")
```

```
[19]:      Scaffolds      CDS
      0      7795  38554
      1         7  48350
      2      2187  57661
      3      1191  71525
      4      3224  36010
      ...      ...      ...
    5126      1925      0
    5127       590      0
    5128      2013      0
    5129      1028      0
    5130       799      0
```

[5131 rows x 2 columns]

```
[20]: data.select_dtypes("number")
```

```
[20]:      Size(Mb)      GC%  Scaffolds      CDS
      0      167.6760  64.5000      7795  38554
      1      119.6680  36.0528         7  48350
      2      412.9240  34.0470      2187  57661
      3      978.9720  35.1208      1191  71525
      4      823.7860  35.7097      3224  36010
      ...      ...      ...      ...      ...
    5126      12.0392  38.1000      1925      0
    5127      11.7086  38.1000       590      0
    5128      11.8194  38.0000      2013      0
    5129      11.7081  38.1000      1028      0
    5130      11.8189  38.1000       799      0
```

[5131 rows x 4 columns]

```
[21]: data.select_dtypes(["object", "int"])
```

```
[21]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Replicons	WGS	Scaffolds \
0	NaN	AHAL01	7795
1	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN	7
2	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001	2187
3	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02	1191
4	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02	3224
...
5126	NaN	FXDE01	1925
5127	NaN	FXCK01	590
5128	NaN	FXC001	2013
5129	NaN	FXCP01	1028
5130	NaN	FXCG01	799

	CDS	Release Date \
0	38554	2013-04-19T00:00:00Z
1	48350	2001-08-13T00:00:00Z
2	57661	2011-08-12T00:00:00Z

```

3      71525  2010-01-05T00:00:00Z
4      36010  2010-12-10T00:00:00Z
...
5126      0  2017-04-21T00:00:00Z
5127      0  2017-04-21T00:00:00Z
5128      0  2017-04-21T00:00:00Z
5129      0  2017-04-21T00:00:00Z
5130      0  2017-04-21T00:00:00Z

```

GenBank FTP \

```

0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...
5126  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

RefSeq FTP

```

0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
5126      NaN
5127      NaN
5128      NaN
5129      NaN
5130      NaN

```

[5131 rows x 14 columns]

```
[22]: data[[col for col in data.columns if data[col].nunique() <50]]
```

```

[22]:
      Organism Groups      Level
0  Eukaryota;Protists;Other Protists  Scaffold
1      Eukaryota;Plants;Land Plants  Chromosome
2      Eukaryota;Plants;Land Plants  Chromosome
3      Eukaryota;Plants;Land Plants  Chromosome
4      Eukaryota;Plants;Land Plants  Chromosome
...
5126  Eukaryota;Fungi;Ascomycetes  Scaffold

```

```

5127      Eukaryota;Fungi;Ascomycetes      Scaffold
5128      Eukaryota;Fungi;Ascomycetes      Scaffold
5129      Eukaryota;Fungi;Ascomycetes      Scaffold
5130      Eukaryota;Fungi;Ascomycetes      Scaffold

```

[5131 rows x 2 columns]

2.5 Methods vs. Attributes

[23]: data

```

[23]:      #Organism Name      Organism Groups \
0      Emiliana huxleyi CCMP1516      Eukaryota;Protists;Other Protists
1      Arabidopsis thaliana      Eukaryota;Plants;Land Plants
2      Medicago truncatula      Eukaryota;Plants;Land Plants
3      Glycine max      Eukaryota;Plants;Land Plants
4      Solanum lycopersicum      Eukaryota;Plants;Land Plants
...      ...      ...
5126      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5127      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5128      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5129      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5130      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes

      Strain      BioSample      BioProject      Assembly      Level \
0      CCMP1516      SAMN02744062      PRJNA77753      GCA_000372725.1      Scaffold
1      NaN      SAMN03081427      PRJNA10719      GCA_000001735.1      Chromosome
2      A17      SAMN02299339      PRJNA10791      GCA_000219495.2      Chromosome
3      NaN      SAMN00002965      PRJNA19861      GCA_000004515.3      Chromosome
4      NaN      SAMN02981290      PRJNA119      GCA_000188115.2      Chromosome
...      ...      ...      ...      ...      ...
5126      NaN      SAMEA3110475      PRJEB7675      GCA_900178465.1      Scaffold
5127      ZP848      SAMEA3110493      PRJEB7675      GCA_900178475.1      Scaffold
5128      VKMY373      SAMEA3110473      PRJEB7675      GCA_900178485.1      Scaffold
5129      245      SAMEA3110476      PRJEB7675      GCA_900178495.1      Scaffold
5130      M2ON0800_1A      SAMEA3110463      PRJEB7675      GCA_900178505.1      Scaffold

      Size(Mb)      GC%      Replicons \
0      167.6760      64.5000      NaN
1      119.6680      36.0528      chromosome 1:NC_003070.9/CP002684.1; chromosom...
2      412.9240      34.0470      chromosome 1:NC_016407.2/CM001217.2; chromosom...
3      978.9720      35.1208      chromosome 1:NC_016088.2/CM000834.2; chromosom...
4      823.7860      35.7097      chromosome 1:NC_015438.2/CM001064.2; chromosom...
...      ...      ...      ...
5126      12.0392      38.1000      NaN
5127      11.7086      38.1000      NaN
5128      11.8194      38.0000      NaN

```

5129	11.7081	38.1000	NaN
5130	11.8189	38.1000	NaN

	WGS	Scaffolds	CDS	Release Date \
0	AHAL01	7795	38554	2013-04-19T00:00:00Z
1	NaN	7	48350	2001-08-13T00:00:00Z
2	APN001	2187	57661	2011-08-12T00:00:00Z
3	ACUP02	1191	71525	2010-01-05T00:00:00Z
4	AEKE02	3224	36010	2010-12-10T00:00:00Z
...
5126	FXDE01	1925	0	2017-04-21T00:00:00Z
5127	FXCK01	590	0	2017-04-21T00:00:00Z
5128	FXC001	2013	0	2017-04-21T00:00:00Z
5129	FXCP01	1028	0	2017-04-21T00:00:00Z
5130	FXCG01	799	0	2017-04-21T00:00:00Z

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
5126	NaN
5127	NaN
5128	NaN
5129	NaN
5130	NaN

[5131 rows x 16 columns]

```
[24]: data.shape
```

```
[24]: (5131, 16)
```

```
[25]: data.drop(columns="#Organism Name")
```

```
[25]:
```

	Organism Groups	Strain	BioSample	\
0	Eukaryota;Protists;Other Protists	CCMP1516	SAMN02744062	
1	Eukaryota;Plants;Land Plants	NaN	SAMN03081427	
2	Eukaryota;Plants;Land Plants	A17	SAMN02299339	
3	Eukaryota;Plants;Land Plants	NaN	SAMN00002965	
4	Eukaryota;Plants;Land Plants	NaN	SAMN02981290	
...	
5126	Eukaryota;Fungi;Ascomycetes	NaN	SAMEA3110475	
5127	Eukaryota;Fungi;Ascomycetes	ZP848	SAMEA3110493	
5128	Eukaryota;Fungi;Ascomycetes	VKMY373	SAMEA3110473	
5129	Eukaryota;Fungi;Ascomycetes	245	SAMEA3110476	
5130	Eukaryota;Fungi;Ascomycetes	M2ON0800_1A	SAMEA3110463	

	BioProject	Assembly	Level	Size(Mb)	GC%	\
0	PRJNA77753	GCA_000372725.1	Scaffold	167.6760	64.5000	
1	PRJNA10719	GCA_000001735.1	Chromosome	119.6680	36.0528	
2	PRJNA10791	GCA_000219495.2	Chromosome	412.9240	34.0470	
3	PRJNA19861	GCA_000004515.3	Chromosome	978.9720	35.1208	
4	PRJNA119	GCA_000188115.2	Chromosome	823.7860	35.7097	
...	
5126	PRJEB7675	GCA_900178465.1	Scaffold	12.0392	38.1000	
5127	PRJEB7675	GCA_900178475.1	Scaffold	11.7086	38.1000	
5128	PRJEB7675	GCA_900178485.1	Scaffold	11.8194	38.0000	
5129	PRJEB7675	GCA_900178495.1	Scaffold	11.7081	38.1000	
5130	PRJEB7675	GCA_900178505.1	Scaffold	11.8189	38.1000	

	Replicons	WGS	Scaffolds	\
0	NaN	AHAL01	7795	
1	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN	7	
2	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001	2187	
3	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02	1191	
4	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02	3224	
...	
5126	NaN	FXDE01	1925	
5127	NaN	FXCK01	590	
5128	NaN	FXC001	2013	
5129	NaN	FXCP01	1028	
5130	NaN	FXCG01	799	

	CDS	Release Date	\
0	38554	2013-04-19T00:00:00Z	
1	48350	2001-08-13T00:00:00Z	
2	57661	2011-08-12T00:00:00Z	
3	71525	2010-01-05T00:00:00Z	
4	36010	2010-12-10T00:00:00Z	

```

...
5126      0  2017-04-21T00:00:00Z
5127      0  2017-04-21T00:00:00Z
5128      0  2017-04-21T00:00:00Z
5129      0  2017-04-21T00:00:00Z
5130      0  2017-04-21T00:00:00Z

```

```

                                GenBank FTP  \
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...
5126   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

```

                                RefSeq FTP
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
5126                                     NaN
5127                                     NaN
5128                                     NaN
5129                                     NaN
5130                                     NaN

```

[5131 rows x 15 columns]

```
[26]: data.drop(columns="#Organism Name").shape
```

```
[26]: (5131, 15)
```

```
[27]: data.describe()
```

```

[27]:
      count      Size(Mb)      GC%      Scaffolds      CDS
count    5131.000000    5131.000000    5.131000e+03    5131.000000
mean       393.103324     42.036497    4.630382e+04    6205.094913
std       1206.196248     10.584974    2.427502e+05    11930.220757
min         0.662517      0.000000    1.000000e+00      0.000000
25%        20.279550     38.100000    1.960000e+02      0.000000

```


50%	39.491700	42.000000	1.591000e+03	0.000000
75%	235.581000	49.000000	1.034250e+04	8392.000000
max	32393.600000	71.400000	6.870110e+06	123467.000000

```
[28]: data["Release Date"]
```

```
[28]: 0      2013-04-19T00:00:00Z
      1      2001-08-13T00:00:00Z
      2      2011-08-12T00:00:00Z
      3      2010-01-05T00:00:00Z
      4      2010-12-10T00:00:00Z
      ...
      5126     2017-04-21T00:00:00Z
      5127     2017-04-21T00:00:00Z
      5128     2017-04-21T00:00:00Z
      5129     2017-04-21T00:00:00Z
      5130     2017-04-21T00:00:00Z
      Name: Release Date, Length: 5131, dtype: object
```

```
[29]: data["Release Date"].value_counts()
```

```
[29]: Release Date
      2016-09-12T00:00:00Z    161
      2016-03-01T00:00:00Z     99
      2015-11-29T00:00:00Z     81
      2017-09-26T00:00:00Z     63
      2016-06-20T00:00:00Z     54
      ...
      2017-04-23T00:00:00Z      1
      2014-12-12T00:00:00Z      1
      2017-05-11T00:00:00Z      1
      2012-11-29T00:00:00Z      1
      2015-05-12T00:00:00Z      1
      Name: count, Length: 1580, dtype: int64
```

```
[30]: data["Release Date"].value_counts(normalize=True)
```

```
[30]: Release Date
      2016-09-12T00:00:00Z    0.031712
      2016-03-01T00:00:00Z    0.019500
      2015-11-29T00:00:00Z    0.015954
      2017-09-26T00:00:00Z    0.012409
      2016-06-20T00:00:00Z    0.010636
      ...
      2017-04-23T00:00:00Z    0.000197
      2014-12-12T00:00:00Z    0.000197
      2017-05-11T00:00:00Z    0.000197
```

```
2012-11-29T00:00:00Z    0.000197
2015-05-12T00:00:00Z    0.000197
Name: proportion, Length: 1580, dtype: float64
```

2.6 Intuition or Common sense in Pandas

```
[31]: year = data["Release Date"]
```

```
[32]: type(year)
```

```
[32]: pandas.core.series.Series
```

```
[33]: type(data)
```

```
[33]: pandas.core.frame.DataFrame
```

```
[34]: data.shape
```

```
[34]: (5131, 16)
```

```
[35]: year.shape
```

```
[35]: (5131,)
```

```
[36]: data.columns
```

```
[36]: Index(['#Organism Name', 'Organism Groups', 'Strain', 'BioSample',
          'BioProject', 'Assembly', 'Level', 'Size(Mb)', 'GC%', 'Replicons',
          'WGS', 'Scaffolds', 'CDS', 'Release Date', 'GenBank FTP', 'RefSeq FTP'],
          dtype='object')
```

```
[37]: data["Organism Groups"]
```

```
[37]: 0      Eukaryota;Protists;Other Protists
      1      Eukaryota;Plants;Land Plants
      2      Eukaryota;Plants;Land Plants
      3      Eukaryota;Plants;Land Plants
      4      Eukaryota;Plants;Land Plants
      ...
      5126   Eukaryota;Fungi;Ascomycetes
      5127   Eukaryota;Fungi;Ascomycetes
      5128   Eukaryota;Fungi;Ascomycetes
      5129   Eukaryota;Fungi;Ascomycetes
      5130   Eukaryota;Fungi;Ascomycetes
      Name: Organism Groups, Length: 5131, dtype: object
```

```
[38]: data["Organism Groups"].value_counts()
```

```
[38]: Organism Groups
      Eukaryota;Fungi;Ascomycetes          2177
      Eukaryota;Fungi;Basidiomycetes       508
      Eukaryota;Plants;Land Plants         430
      Eukaryota;Animals;Insects            390
      Eukaryota;Animals;Mammals            288
      Eukaryota;Protists;Other Protists    274
      Eukaryota;Protists;Apicomplexans    191
      Eukaryota;Fungi;Other Fungi         159
      Eukaryota;Animals;Other Animals      132
      Eukaryota;Animals;Roundworms        129
      Eukaryota;Animals;Fishes            125
      Eukaryota;Animals;Birds             113
      Eukaryota;Protists;Kinetoplasts      74
      Eukaryota;Plants;Green Algae         47
      Eukaryota;Animals;Flatworms         44
      Eukaryota;Animals;Reptiles           24
      Eukaryota;Other;Other                20
      Eukaryota;Animals;Amphibians         5
      Eukaryota;Plants;Other Plants        1
      Name: count, dtype: int64
```

```
[39]: data.value_counts()
```

```
[39]: #Organism Name          Organism Groups          Strain
      BioSample      BioProject      Assembly      Level      Size(Mb)      GC%
      Replicons
      WGS      Scaffolds      CDS      Release Date      GenBank FTP
      RefSeq FTP
      Acinonyx jubatus          Eukaryota;Animals;Mammals      AJU 981
      Chewbacca SAMN04127588 PRJNA305718 GCA_001443585.1 Scaffold      2372.55000
      41.4000 mitochondrion MT:NC_005212.1/AF344830.1
      LLWD01 14383      27284 2015-11-13T00:00:00Z
      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/443/585/GCA_001443585.1_aciJub1
      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/443/585/GCF_001443585.1_aciJub1
      1
      Odobenus rosmarus divergens          Eukaryota;Animals;Mammals      Ivan
      SAMN01180787 PRJNA167474 GCA_000321225.1 Scaffold      2400.15000 41.7000
      mitochondrion MT:NC_004029.2/AJ428576.2
      ANOP01 3893      31370 2012-12-17T00:00:00Z
      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/321/225/GCA_000321225.1_Oros_1.0
      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/321/225/GCF_000321225.1_Oros_1.0
      1
      Ogataea parapolyomorpha DL-1          Eukaryota;Fungi;Ascomycetes      DL-1
      SAMN02981294 PRJNA60503 GCA_000187245.3 Chromosome 8.87459 47.8672
      chromosome I:NC_027866.1/CM002300.1; chromosome II:NC_027865.1/CM002301.1;
      chromosome III:NC_027864.1/CM002302.1; chromosome IV:NC_027863.1/CM002303.1;
```

chromosome V:NC_027862.1/CM002304.1; chromosome VI:NC_027861.1/CM002305.1;
 chromosome VII:NC_027860.1/CM002306.1
 AEOI02 7 10650 2011-01-31T00:00:00Z ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/187/245/GCA_000187245.3_Hansenula_2 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/187/245/GCF_000187245.1_Hansenula_2 1
 Oncorhynchus kisutch Eukaryota;Animals;Fishes
 150728-3 SAMN05991263 PRJNA378663 GCA_002021735.1 Chromosome
 2369.93000 43.6225 linkage group LG1:NC_034174.1/CM007718.1; linkage group
 LG2:NC_034175.1/CM007719.1; linkage group LG3:NC_034176.1/CM007720.1; linkage
 group LG4:NC_034177.1/CM007721.1; linkage group LG5:NC_034178.1/CM007722.1;
 linkage group LG6:NC_034179.1/CM007723.1; linkage group
 LG7:NC_034180.1/CM007724.1; linkage group LG8:NC_034181.1/CM007725.1; linkage
 group LG9:NC_034182.1/CM007726.1; linkage group LG10:NC_034183.1/CM007727.1;
 linkage group LG11:NC_034184.1/CM007728.1; linkage group
 LG12:NC_034185.1/CM007729.1; linkage group LG13:NC_034186.1/CM007730.1; linkage
 group LG14:NC_034187.1/CM007731.1; linkage group LG15:NC_034188.1/CM007732.1;
 linkage group LG16:NC_034189.1/CM007733.1; linkage group
 LG17:NC_034190.1/CM007734.1; linkage group LG18:NC_034191.1/CM007735.1; linkage
 group LG19:NC_034192.1/CM007736.1; linkage group LG20:NC_034193.1/CM007737.1;
 linkage group LG21:NC_034194.1/CM007738.1; linkage group
 LG22:NC_034195.1/CM007739.1; linkage group LG23:NC_034196.1/CM007740.1; linkage
 group LG24:NC_034197.1/CM007741.1; linkage group LG25:NC_034198.1/CM007742.1;
 linkage group LG26:NC_034199.1/CM007743.1; linkage group
 LG27:NC_034200.1/CM007744.1; linkage group LG28:NC_034201.1/CM007745.1; linkage
 group LG29:NC_034202.1/CM007746.1; linkage group LG30:NC_034203.1/CM007747.1;
 mitochondrion MT:NC_009263.1/EF126369.1 MPKV01 22813 57592
 2017-03-06T00:00:00Z
 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/021/735/GCA_002021735.1_Okis_V1
 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/021/735/GCF_002021735.1_Okis_V1
 1
 Oncorhynchus mykiss Eukaryota;Animals;Fishes Swanson
 SAMN05449231 PRJNA390662 GCA_002163495.1 Chromosome 2179.00000 43.4875
 chromosome 1:NC_035077.1/CM007935.1; chromosome 2:NC_035078.1/CM007936.1;
 chromosome 3:NC_035079.1/CM007937.1; chromosome 4:NC_035080.1/CM007938.1;
 chromosome 5:NC_035081.1/CM007939.1; chromosome 6:NC_035082.1/CM007940.1;
 chromosome 7:NC_035083.1/CM007941.1; chromosome 8:NC_035084.1/CM007942.1;
 chromosome 9:NC_035085.1/CM007943.1; chromosome 10:NC_035086.1/CM007944.1;
 chromosome 11:NC_035087.1/CM007945.1; chromosome 12:NC_035088.1/CM007946.1;
 chromosome 13:NC_035089.1/CM007947.1; chromosome 14:NC_035090.1/CM007948.1;
 chromosome 15:NC_035091.1/CM007949.1; chromosome 16:NC_035092.1/CM007950.1;
 chromosome 17:NC_035093.1/CM007951.1; chromosome 18:NC_035094.1/CM007952.1;
 chromosome 19:NC_035095.1/CM007953.1; chromosome 20:NC_035096.1/CM007954.1;
 chromosome 21:NC_035097.1/CM007955.1; chromosome 22:NC_035098.1/CM007956.1;
 chromosome 23:NC_035099.1/CM007957.1; chromosome 24:NC_035100.1/CM007958.1;
 chromosome 25:NC_035101.1/CM007959.1; chromosome 26:NC_035102.1/CM007960.1;
 chromosome 27:NC_035103.1/CM007961.1; chromosome 28:NC_035104.1/CM007962.1;
 chromosome 29:NC_035105.1/CM007963.1; mitochondrion MT:NC_001717.1/L29771.1

MSJN01 139800 71266 2017-05-31T00:00:00Z
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/163/495/GCA_002163495.1_Omyk_1.0
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/163/495/GCF_002163495.1_Omyk_1.0
1

..

Fusarium graminearum PH-1 Eukaryota;Fungi;Ascomycetes PH-1;
NRRL 31084 SAMN02953593 PRJNA13839 GCA_000240135.3 Chromosome 36.45800
48.3278 chromosome 1:NC_026474.1/CM000574.1; chromosome
2:NC_026475.1/CM000575.1; chromosome 3:NC_026476.1/CM000576.1; chromosome
4:NC_026477.1/CM000577.1
AACM02 31 13313 2003-05-09T00:00:00Z ftp://ftp.ncbi.nlm.nih.gov/genom
es/all/GCA/000/240/135/GCA_000240135.3_ASM24013v3 ftp://ftp.ncbi.nlm.nih.gov/g
enomes/all/GCF/000/240/135/GCF_000240135.3_ASM24013v3 1
Fusarium oxysporum f. sp. lycopersici 4287 Eukaryota;Fungi;Ascomycetes 4287
SAMN02953675 PRJNA18813 GCA_000149955.2 Chromosome 61.38690 48.4264
chromosome 1:NC_030986.1/CM000589.1; chromosome 2:NC_030987.1/CM000590.1;
chromosome 3:NC_030988.1/CM000591.1; chromosome 4:NC_030989.1/CM000592.1;
chromosome 5:NC_030990.1/CM000593.1; chromosome 6:NC_030991.1/CM000594.1;
chromosome 7:NC_030992.1/CM000595.1; chromosome 8:NC_030993.1/CM000596.1;
chromosome 9:NC_030994.1/CM000597.1; chromosome 10:NC_030995.1/CM000598.1;
chromosome 11:NC_030996.1/CM000599.1; chromosome 12:NC_030997.1/CM000600.1;
chromosome 13:NC_030998.1/CM000601.1; chromosome 14:NC_030999.1/CM000602.2;
chromosome 15:NC_031000.1/CM000603.1
AAXH01 114 27347 2007-04-10T00:00:00Z ftp://ftp.ncbi.nlm.nih.gov/genom
es/all/GCA/000/149/955/GCA_000149955.2_ASM14995v2 ftp://ftp.ncbi.nlm.nih.gov/g
enomes/all/GCF/000/149/955/GCF_000149955.1_ASM14995v2 1
Fusarium pseudograminearum CS3096 Eukaryota;Fungi;Ascomycetes CS3096
SAMN02981337 PRJNA66583 GCA_000303195.2 Chromosome 36.97330 47.7506
chromosome 1:NC_031951.1/CM003198.1; chromosome 2:NC_031952.1/CM003199.1;
chromosome 3:NC_031953.1/CM003200.1; chromosome 4:NC_031954.1/CM003201.1
AFNW01 281 12397 2012-10-05T00:00:00Z
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/303/195/GCA_000303195.2_FP7
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/303/195/GCF_000303195.2_FP7
1
Fusarium verticillioides 7600 Eukaryota;Fungi;Ascomycetes 7600
SAMN02953630 PRJNA15553 GCA_000149555.1 Chromosome 41.84490 48.6857
chromosome 1:NC_031675.1/CM000578.1; chromosome 2:NC_031676.1/CM000579.1;
chromosome 3:NC_031677.1/CM000580.1; chromosome 4:NC_031678.1/CM000581.1;
chromosome 5:NC_031679.1/CM000582.1; chromosome 6:NC_031680.1/CM000583.1;
chromosome 7:NC_031681.1/CM000584.1; chromosome 8:NC_031682.1/CM000585.1;
chromosome 9:NC_031683.1/CM000586.1; chromosome 10:NC_031684.1/CM000587.1;
chromosome 11:NC_031685.1/CM000588.1; mitochondrion MT:NC_016687.1/JN041210.1
AAIM02 37 20574 2005-10-18T00:00:00Z ftp://ftp.ncbi.nlm.nih.gov/genom
es/all/GCA/000/149/555/GCA_000149555.1_ASM14955v1 ftp://ftp.ncbi.nlm.nih.gov/g
enomes/all/GCF/000/149/555/GCF_000149555.1_ASM14955v1 1
Zymoseptoria tritici IP0323 Eukaryota;Fungi;Ascomycetes IP0323
SAMN02746078 PRJNA19047 GCA_000219625.1 Chromosome 39.68630 52.1184

```

chromosome 1:NC_018218.1/CM001196.1; chromosome 2:NC_018217.1/CM001197.1;
chromosome 3:NC_018216.1/CM001198.1; chromosome 4:NC_018215.1/CM001199.1;
chromosome 5:NC_018214.1/CM001200.1; chromosome 6:NC_018213.1/CM001201.1;
chromosome 7:NC_018212.1/CM001202.1; chromosome 8:NC_018211.1/CM001203.1;
chromosome 9:NC_018210.1/CM001204.1; chromosome 10:NC_018209.1/CM001205.1;
chromosome 11:NC_018208.1/CM001206.1; chromosome 12:NC_018207.1/CM001207.1;
chromosome 13:NC_018206.1/CM001208.1; chromosome 14:NC_018205.1/CM001209.1;
chromosome 15:NC_018204.1/CM001210.1; chromosome 16:NC_018203.1/CM001211.1;
chromosome 17:NC_018202.1/CM001212.1; chromosome 18:NC_018201.1/CM001213.1;
chromosome 19:NC_018200.1/CM001214.1; chromosome 20:NC_018199.1/CM001215.1;
chromosome 21:NC_018198.1/CM001216.1
ACPE01 21 21926 2011-06-22T00:00:00Z ftp://ftp.ncbi.nlm.nih.gov/genom
es/all/GCA/000/219/625/GCA_000219625.1_MYCGR_v2.0 ftp://ftp.ncbi.nlm.nih.gov/g
enomes/all/GCF/000/219/625/GCF_000219625.1_MYCGR_v2.0 1
Name: count, Length: 214, dtype: int64

```

```
[40]: year.sort_values()
```

```

[40]: 88      1994-11-22T00:00:00Z
      17      1998-06-29T00:00:00Z
      26      1998-12-17T00:00:00Z
      11      1999-11-08T00:00:00Z
      33      2001-02-05T00:00:00Z
      ...
      3119      NaN
      3156      NaN
      3209      NaN
      3357      NaN
      4976      NaN
Name: Release Date, Length: 5131, dtype: object

```

```
[41]: data.sort_values(by="Strain")
```

```

[41]:      #Organism Name      Organism Groups \
369      Maylandia zebra      Eukaryota;Animals;Fishes
234      Ictidomys tridecemlineatus      Eukaryota;Animals;Mammals
3452      Pyrenophora teres f. teres 0-1      Eukaryota;Fungi;Ascomycetes
3270      Oreochromis niloticus      Eukaryota;Animals;Fishes
2944      Phytophthora kernoviae 00238/432      Eukaryota;Protists;Other Protists
...
5087      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5088      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5089      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5090      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5126      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes

      Strain      BioSample      BioProject      Assembly \

```

369	#2_Kocher and #3_Kocher	SAMN03892560	PRJNA60369	GCA_000238955.3
234	#75	SAMN00210855	PRJNA61725	GCA_000236235.1
3452	0-1	SAMN00016826	PRJNA50389	GCA_000166005.1
3270	000638D3DF	SAMN00117560	PRJNA59571	GCA_000188235.2
2944	00238/432	SAMN02981496	PRJNA184368	GCA_000333075.2
...
5087	NaN	SAMEA3249804	PRJEB8455	GCA_900178035.1
5088	NaN	SAMEA3249826	PRJEB8455	GCA_900178045.1
5089	NaN	SAMEA3249812	PRJEB8455	GCA_900178065.1
5090	NaN	SAMEA3249800	PRJEB8455	GCA_900178075.1
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1

	Level	Size(Mb)	GC% \
369	Scaffold	859.8420	41.4002
234	Scaffold	2478.3900	40.5000
3452	Scaffold	33.5833	50.9000
3270	Chromosome	927.6960	39.0824
2944	Scaffold	42.5238	49.7000
...
5087	Scaffold	13.8535	41.9000
5088	Scaffold	11.5139	38.2000
5089	Scaffold	11.5691	38.1000
5090	Scaffold	12.5459	40.0000
5126	Scaffold	12.0392	38.1000

		Replicons	WGS	Scaffolds \
369	mitochondrion MT:NC_027944.1/KT166981.1; mitoc...	AGTA03		3555
234		NaN	AGTP01	12483
3452		NaN	AEEY01	6412
3270	linkage group LG1:CM001444.1; linkage group LG...	AERX01		5910
2944		NaN	AOFI02	2303
...	
5087		NaN	FXDY01	8158
5088		NaN	FXDN01	1441
5089		NaN	FXLH01	1144
5090		NaN	FXDS01	6246
5126		NaN	FXDE01	1925

	CDS	Release Date \
369	44136	2011-12-01T00:00:00Z
234	38592	2011-11-25T00:00:00Z
3452	11799	2010-11-03T00:00:00Z
3270	0	2011-01-28T00:00:00Z
2944	0	2013-01-24T00:00:00Z
...
5087	0	2017-04-25T00:00:00Z
5088	0	2017-04-25T00:00:00Z

```

5089      0  2017-04-26T00:00:00Z
5090      0  2017-04-25T00:00:00Z
5126      0  2017-04-21T00:00:00Z

```

GenBank FTP \

```

369  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
234  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3452 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3270 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2944 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...
5087 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5088 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5089 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5090 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5126 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

RefSeq FTP

```

369  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
234  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3452 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3270 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2944                                     NaN
...
5087                                     NaN
5088                                     NaN
5089                                     NaN
5090                                     NaN
5126                                     NaN

```

[5131 rows x 16 columns]

2.7 Method Chaining in Pandas

```
[42]: data["Strain"]
```

```

[42]: 0      CCMP1516
      1      NaN
      2      A17
      3      NaN
      4      NaN
      ...
      5126     NaN
      5127     ZP848
      5128     VKMY373
      5129      245
      5130  M2ON0800_1A

```


Name: Strain, Length: 5131, dtype: object

```
[43]: data["Strain"].value_counts()
```

```
[43]: Strain
Race 77      14
CHM1      11
EU1       7
DAOM 197198w 7
Clinical   7
..
CBS 133217   1
CBS 341.76   1
CBS 650.92   1
CBS 111550   1
M2ON0800_1A  1
Name: count, Length: 3962, dtype: int64
```

```
[44]: data["Strain"].value_counts().sort_index()
```

```
[44]: Strain
#2_Kocher and #3_Kocher  1
#75                      1
0-1                     1
000638D3DF              1
00238/432               1
..
ya'a_city_454_Pm        1
ya'a_city_454_Px        1
yakQH1                  1
ysfri_paol              1
zdho120                 1
Name: count, Length: 3962, dtype: int64
```

```
[45]: data["Strain"].sort_index()
```

```
[45]: 0      CCMP1516
1      NaN
2      A17
3      NaN
4      NaN
...
5126   NaN
5127   ZP848
5128   VKMY373
5129   245
5130   M2ON0800_1A
```

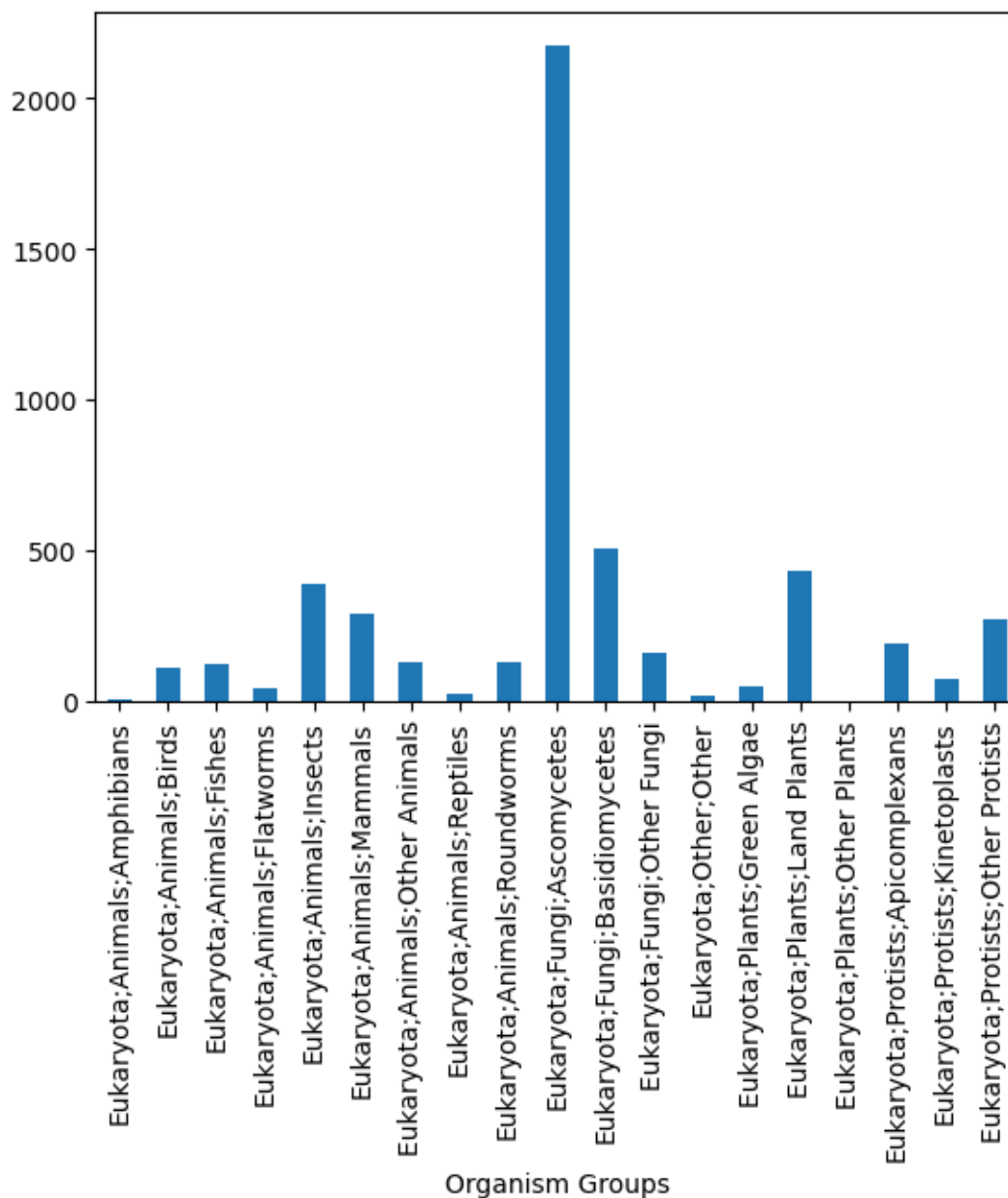
Name: Strain, Length: 5131, dtype: object

```
[46]: data["Strain"].sort_index().value_counts()
```

```
[46]: Strain
      Race 77      14
      CHM1      11
      EU1       7
      DAOM 197198w  7
      Clinical    7
      ..
      CBS 133217    1
      CBS 341.76    1
      CBS 650.92    1
      CBS 111550    1
      M2ON0800_1A    1
      Name: count, Length: 3962, dtype: int64
```

```
[47]: data["Organism Groups"].value_counts().sort_index().plot.bar()
```

```
[47]: <Axes: xlabel='Organism Groups'>
```



2.8 loc [] and iloc []

[48]: data

```
[48]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants

4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%	Replicons \
0	167.6760	64.5000	NaN
1	119.6680	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...
2	412.9240	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...
3	978.9720	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...
4	823.7860	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...
...
5126	12.0392	38.1000	NaN
5127	11.7086	38.1000	NaN
5128	11.8194	38.0000	NaN
5129	11.7081	38.1000	NaN
5130	11.8189	38.1000	NaN

	WGS	Scaffolds	CDS	Release Date \
0	AHAL01	7795	38554	2013-04-19T00:00:00Z
1	NaN	7	48350	2001-08-13T00:00:00Z
2	APN001	2187	57661	2011-08-12T00:00:00Z
3	ACUP02	1191	71525	2010-01-05T00:00:00Z
4	AEKE02	3224	36010	2010-12-10T00:00:00Z
...
5126	FXDE01	1925	0	2017-04-21T00:00:00Z
5127	FXCK01	590	0	2017-04-21T00:00:00Z
5128	FXC001	2013	0	2017-04-21T00:00:00Z
5129	FXCP01	1028	0	2017-04-21T00:00:00Z
5130	FXCG01	799	0	2017-04-21T00:00:00Z

```

                                GenBank FTP \
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...
5126   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

```

                                RefSeq FTP
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
5126   NaN
5127   NaN
5128   NaN
5129   NaN
5130   NaN

```

[5131 rows x 16 columns]

```
[49]: data.iloc[4]
```

```

[49]: #Organism Name          Solanum lycopersicum
Organism Groups              Eukaryota;Plants;Land Plants
Strain                       NaN
BioSample                    SAMN02981290
BioProject                   PRJNA119
Assembly                     GCA_000188115.2
Level                        Chromosome
Size(Mb)                     823.786
GC%                          35.7097
Replicons                    chromosome 1:NC_015438.2/CM001064.2; chromosom...
WGS                           AEKE02
Scaffolds                    3224
CDS                          36010
Release Date                  2010-12-10T00:00:00Z
GenBank FTP                   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
RefSeq FTP                    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Name: 4, dtype: object

```

```
[50]: data.iloc[:5]
```

```
[50]:
```

	#Organism Name	Organism Groups	Strain \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists	CCMP1516
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants	NaN
2	Medicago truncatula	Eukaryota;Plants;Land Plants	A17
3	Glycine max	Eukaryota;Plants;Land Plants	NaN
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants	NaN

	BioSample	BioProject	Assembly	Level	Size(Mb)	GC% \
0	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold	167.676	64.5000
1	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome	119.668	36.0528
2	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome	412.924	34.0470
3	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome	978.972	35.1208
4	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome	823.786	35.7097

	Replicons	WGS	Scaffolds \
0	NaN	AHAL01	7795
1	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN	7
2	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001	2187
3	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02	1191
4	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02	3224

	CDS	Release Date \
0	38554	2013-04-19T00:00:00Z
1	48350	2001-08-13T00:00:00Z
2	57661	2011-08-12T00:00:00Z
3	71525	2010-01-05T00:00:00Z
4	36010	2010-12-10T00:00:00Z

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...

```
[51]: #[start:stop:step]
data.iloc[1:10]
```

[51]:

	#Organism Name	Organism Groups	Strain \
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants	NaN
2	Medicago truncatula	Eukaryota;Plants;Land Plants	A17
3	Glycine max	Eukaryota;Plants;Land Plants	NaN
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants	NaN
5	Hordeum vulgare	Eukaryota;Plants;Land Plants	NaN
6	Oryza sativa Japonica Group	Eukaryota;Plants;Land Plants	NaN
7	Triticum aestivum	Eukaryota;Plants;Land Plants	NaN
8	Zea mays	Eukaryota;Plants;Land Plants	NaN
9	Pneumocystis carinii B80	Eukaryota;Fungi;Ascomycetes	B80

	BioSample	BioProject	Assembly	Level	Size(Mb) \
1	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome	119.66800
2	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome	412.92400
3	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome	978.97200
4	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome	823.78600
5	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold	9788.86000
6	SAMD00000397	PRJNA12269	GCA_001433935.1	Chromosome	374.42300
7	SAMN07284949	PRJNA392179	GCA_002220415.2	Contig	15344.70000
8	SAMN04296295	PRJNA10769	GCA_000005005.6	Chromosome	2135.08000
9	SAMN02380717	PRJNA223511	GCA_001477545.1	Contig	7.66146

	GC%	Replicons	WGS \
1	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN
2	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001
3	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02
4	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02
5	0.0000		NaN FJWB02
6	43.5769	chromosome 1:NC_029256.1/AP014957.1; chromosom...	NaN
7	46.1000		NaN NMPL02
8	46.9109	chromosome 1:NC_024459.2/CM007647.1; chromosom...	LPUQ01
9	27.8000		NaN LFBV01

	Scaffolds	CDS	Release Date \
1	7	48350	2001-08-13T00:00:00Z
2	2187	57661	2011-08-12T00:00:00Z
3	1191	71525	2010-01-05T00:00:00Z
4	3224	36010	2010-12-10T00:00:00Z
5	72295	0	2016-08-25T00:00:00Z
6	58	41070	2015-10-10T00:00:00Z
7	279439	0	2017-07-13T00:00:00Z
8	598	58411	2010-01-29T00:00:00Z
9	62	3646	2015-12-22T00:00:00Z

	GenBank FTP \
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...

```

3 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
6 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
7 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
8 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
9 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...

```

RefSeq FTP

```

1 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
5
6 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...
7
8 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
9 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...

```

```
[52]: data.iloc[1:10:2]
```

```

[52]:      #Organism Name      Organism Groups Strain \
1      Arabidopsis thaliana  Eukaryota;Plants;Land Plants  NaN
3      Glycine max          Eukaryota;Plants;Land Plants  NaN
5      Hordeum vulgare      Eukaryota;Plants;Land Plants  NaN
7      Triticum aestivum    Eukaryota;Plants;Land Plants  NaN
9  Pneumocystis carinii B80  Eukaryota;Fungi;Ascomycetes  B80

```

```

      BioSample  BioProject      Assembly      Level      Size(Mb) \
1  SAMN03081427  PRJNA10719  GCA_000001735.1  Chromosome  119.66800
3  SAMN00002965  PRJNA19861  GCA_000004515.3  Chromosome  978.97200
5  SAMEA3903302  PRJEB13020  GCA_900075435.2  Scaffold    9788.86000
7  SAMN07284949  PRJNA392179  GCA_002220415.2  Contig     15344.70000
9  SAMN02380717  PRJNA223511  GCA_001477545.1  Contig       7.66146

```

```

      GC%      Replicons      WGS \
1  36.0528  chromosome 1:NC_003070.9/CP002684.1; chromosom...  NaN
3  35.1208  chromosome 1:NC_016088.2/CM000834.2; chromosom...  ACUP02
5    0.0000                                     NaN  FJWB02
7  46.1000                                     NaN  NMPL02
9  27.8000                                     NaN  LFVZ01

```

```

      Scaffolds      CDS      Release Date \
1           7  48350  2001-08-13T00:00:00Z
3        1191  71525  2010-01-05T00:00:00Z
5        72295      0  2016-08-25T00:00:00Z
7       279439      0  2017-07-13T00:00:00Z

```


9 62 3646 2015-12-22T00:00:00Z

```
GenBank FTP \
1 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
7 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
9 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
```

```
RefSeq FTP
1 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
5 NaN
7 NaN
9 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...
```

[53]: data.iloc[1:10:3]

[53]:

	#Organism Name	Organism Groups	Strain	BioSample	\
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants	NaN	SAMN03081427	
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants	NaN	SAMN02981290	
7	Triticum aestivum	Eukaryota;Plants;Land Plants	NaN	SAMN07284949	

	BioProject	Assembly	Level	Size(Mb)	GC%	\
1	PRJNA10719	GCA_000001735.1	Chromosome	119.668	36.0528	
4	PRJNA119	GCA_000188115.2	Chromosome	823.786	35.7097	
7	PRJNA392179	GCA_002220415.2	Contig	15344.700	46.1000	

	Replicons	WGS	Scaffolds	\
1	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN	7	
4	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02	3224	
7		NaN NMPL02	279439	

	CDS	Release Date	\
1	48350	2001-08-13T00:00:00Z	
4	36010	2010-12-10T00:00:00Z	
7	0	2017-07-13T00:00:00Z	

```
GenBank FTP \
1 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
7 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
```

```
RefSeq FTP
1 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
7 NaN
```

```
[54]: data.iloc[1:]
```

```
[54]:
```

	#Organism Name	Organism Groups	Strain \
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants	NaN
2	Medicago truncatula	Eukaryota;Plants;Land Plants	A17
3	Glycine max	Eukaryota;Plants;Land Plants	NaN
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants	NaN
5	Hordeum vulgare	Eukaryota;Plants;Land Plants	NaN
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	NaN
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	ZP848
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	VKMY373
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	245
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	M2ON0800_1A

	BioSample	BioProject	Assembly	Level	Size(Mb) \
1	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome	119.6680
2	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome	412.9240
3	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome	978.9720
4	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome	823.7860
5	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold	9788.8600
...
5126	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold	12.0392
5127	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold	11.7086
5128	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold	11.8194
5129	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold	11.7081
5130	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold	11.8189

	GC%	Replicons	WGS \
1	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN
2	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001
3	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02
4	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02
5	0.0000		NaN FJWB02
...
5126	38.1000		NaN FXDE01
5127	38.1000		NaN FXCK01
5128	38.0000		NaN FXC001
5129	38.1000		NaN FXCP01
5130	38.1000		NaN FXCG01

	Scaffolds	CDS	Release Date \
1	7	48350	2001-08-13T00:00:00Z
2	2187	57661	2011-08-12T00:00:00Z
3	1191	71525	2010-01-05T00:00:00Z
4	3224	36010	2010-12-10T00:00:00Z
5	72295	0	2016-08-25T00:00:00Z

```

...      ...      ...
5126      1925      0      2017-04-21T00:00:00Z
5127      590      0      2017-04-21T00:00:00Z
5128      2013      0      2017-04-21T00:00:00Z
5129      1028      0      2017-04-21T00:00:00Z
5130      799      0      2017-04-21T00:00:00Z

```

GenBank FTP \

```

1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

```

...
5126      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

RefSeq FTP

```

1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...

```

```

5      NaN
...
5126      NaN
5127      NaN
5128      NaN
5129      NaN
5130      NaN

```

[5130 rows x 16 columns]

```
[55]: data.iloc[-10:]
```

```

[55]:      #Organism Name      Organism Groups      Strain \
5121      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      VPDN_Fino
5122      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      CBS1171
5123      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      KS11
5124      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      YA3
5125      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      460
5126      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      NaN
5127      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      ZP848
5128      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      VKMY373
5129      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes      245

```

5130 *Saccharomyces cerevisiae* Eukaryota;Fungi;Ascomycetes M2ON0800_1A

	BioSample	BioProject	Assembly	Level	Size(Mb)	GC%	\
5121	SAMEA3110500	PRJEB7675	GCA_900178415.1	Scaffold	11.5745	38.0	
5122	SAMEA3110486	PRJEB7675	GCA_900178425.1	Scaffold	12.1625	38.0	
5123	SAMEA3110474	PRJEB7675	GCA_900178435.1	Scaffold	12.0671	38.0	
5124	SAMEA3110484	PRJEB7675	GCA_900178445.1	Scaffold	11.6276	38.1	
5125	SAMEA3110483	PRJEB7675	GCA_900178455.1	Scaffold	12.5711	38.0	
5126	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold	12.0392	38.1	
5127	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold	11.7086	38.1	
5128	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold	11.8194	38.0	
5129	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold	11.7081	38.1	
5130	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold	11.8189	38.1	

	Replicons	WGS	Scaffolds	CDS	Release Date	\
5121	NaN	FXBZ01	924	0	2017-04-21T00:00:00Z	
5122	NaN	FXDD01	2430	0	2017-04-21T00:00:00Z	
5123	NaN	FXDH01	2718	0	2017-04-21T00:00:00Z	
5124	NaN	FXCA01	490	0	2017-04-21T00:00:00Z	
5125	NaN	FXCV01	3297	0	2017-04-21T00:00:00Z	
5126	NaN	FXDE01	1925	0	2017-04-21T00:00:00Z	
5127	NaN	FXCK01	590	0	2017-04-21T00:00:00Z	
5128	NaN	FXC001	2013	0	2017-04-21T00:00:00Z	
5129	NaN	FXCP01	1028	0	2017-04-21T00:00:00Z	
5130	NaN	FXCG01	799	0	2017-04-21T00:00:00Z	

	GenBank	FTP	RefSeq	FTP
5121	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5122	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5123	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5124	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5125	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			NaN

```
[56]: #[start:stop:step]
data.iloc[:,2:6]
```

```
[56]:
```

	Strain	BioSample	BioProject	Assembly
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2

```

...
5126      NaN  SAMEA3110475  PRJEB7675  GCA_900178465.1
5127      ZP848  SAMEA3110493  PRJEB7675  GCA_900178475.1
5128      VKMY373  SAMEA3110473  PRJEB7675  GCA_900178485.1
5129      245  SAMEA3110476  PRJEB7675  GCA_900178495.1
5130  M2ON0800_1A  SAMEA3110463  PRJEB7675  GCA_900178505.1

```

[5131 rows x 4 columns]

```
[57]: data.iloc[3:10,5:-2]
```

```

[57]:      Assembly      Level      Size(Mb)      GC%  \
3  GCA_000004515.3  Chromosome    978.97200  35.1208
4  GCA_000188115.2  Chromosome    823.78600  35.7097
5  GCA_900075435.2    Scaffold    9788.86000   0.0000
6  GCA_001433935.1  Chromosome    374.42300  43.5769
7  GCA_002220415.2    Contig   15344.70000  46.1000
8  GCA_000005005.6  Chromosome    2135.08000  46.9109
9  GCA_001477545.1    Contig      7.66146  27.8000

      Replicons      WGS  Scaffolds  \
3  chromosome 1:NC_016088.2/CM000834.2; chromosom...  ACUP02      1191
4  chromosome 1:NC_015438.2/CM001064.2; chromosom...  AEKE02      3224
5                                     NaN  FJWB02      72295
6  chromosome 1:NC_029256.1/AP014957.1; chromosom...    NaN        58
7                                     NaN  NMPL02     279439
8  chromosome 1:NC_024459.2/CM007647.1; chromosom...  LPUQ01        598
9                                     NaN  LFVZ01        62

      CDS      Release Date
3  71525  2010-01-05T00:00:00Z
4  36010  2010-12-10T00:00:00Z
5      0  2016-08-25T00:00:00Z
6  41070  2015-10-10T00:00:00Z
7      0  2017-07-13T00:00:00Z
8  58411  2010-01-29T00:00:00Z
9   3646  2015-12-22T00:00:00Z

```

```
[58]: data.iloc[:1000]
```

```

[58]:      #Organism Name      Organism Groups  \
0  Emiliana huxleyi CCMP1516  Eukaryota;Protists;Other Protists
1  Arabidopsis thaliana      Eukaryota;Plants;Land Plants
2  Medicago truncatula      Eukaryota;Plants;Land Plants
3  Glycine max      Eukaryota;Plants;Land Plants
4  Solanum lycopersicum      Eukaryota;Plants;Land Plants
..

```

995	Nelumbo nucifera	Eukaryota;Plants;Land Plants
996	Halyomorpha halys	Eukaryota;Animals;Insects
997	Falco cherrug	Eukaryota;Animals;Birds
998	Centruroides sculpturatus	Eukaryota;Animals;Other Animals
999	Trichogramma pretiosum	Eukaryota;Animals;Insects

	Strain	BioSample	BioProject	Assembly \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2
..
995	NaN	SAMN02054413	PRJNA168000	GCA_000365185.2
996	HHAL.00	SAMN02737379	PRJNA298780	GCA_000696795.1
997	NaN	SAMN02981426	PRJNA217049	GCA_000337975.1
998	CEXI.00-Female	SAMN02617800	PRJNA422877	GCA_000671375.2
999	Unisexual culture from Peru	SAMN02439301	PRJNA297592	GCA_000599845.3

	Level	Size(Mb)	GC% \
0	Scaffold	167.676	64.5000
1	Chromosome	119.668	36.0528
2	Chromosome	412.924	34.0470
3	Chromosome	978.972	35.1208
4	Chromosome	823.786	35.7097
..
995	Scaffold	804.648	39.4997
996	Scaffold	1150.110	33.0999
997	Scaffold	1174.810	41.8000
998	Scaffold	925.475	31.4000
999	Scaffold	187.642	39.9000

	Replicons	WGS	Scaffolds \
0	NaN	AHAL01	7795
1	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN	7
2	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001	2187
3	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02	1191
4	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02	3224
..
995	chloroplast Chloroplast:NC_025339.1/KF009944.1	AQOG01	3603
996	mitochondrion MT:NC_013272.1/FJ685650.1	JMPT01	12168
997	NaN	AKMU01	5863
998	NaN	AXZI02	8338
999	NaN	JARR02	925

	CDS	Release Date \
0	38554	2013-04-19T00:00:00Z

```

1    48350  2001-08-13T00:00:00Z
2    57661  2011-08-12T00:00:00Z
3    71525  2010-01-05T00:00:00Z
4    36010  2010-12-10T00:00:00Z
..      ...
995   38191  2013-04-15T00:00:00Z
996   24660  2014-05-30T00:00:00Z
997   21235  2013-02-05T00:00:00Z
998   35529  2014-05-02T00:00:00Z
999      0   2014-03-24T00:00:00Z

```

GenBank FTP \

```

0    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
..      ...
995   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
996   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
997   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
998   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
999   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...

```

RefSeq FTP

```

0    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4    ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
..      ...
995   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
996   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
997   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
998   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
999   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...

```

[1000 rows x 16 columns]

```
[59]: data.loc[:1000, "#Organism Name"]
```

```

[59]: 0    Emiliana huxleyi CCMP1516
      1    Arabidopsis thaliana
      2    Medicago truncatula
      3    Glycine max
      4    Solanum lycopersicum

```

...

```

996          Halyomorpha halys
997          Falco cherrug
998    Centruroides sculpturatus
999          Trichogramma pretiosum
1000          Latrodectus hesperus
Name: #Organism Name, Length: 1001, dtype: object

```

```
[60]: data.loc[:100,["Organism Groups", "#Organism Name", "CDS"]]
```

```

[60]:
      Organism Groups      #Organism Name      CDS
0    Eukaryota;Protists;Other Protists    Emiliania huxleyi CCMP1516  38554
1      Eukaryota;Plants;Land Plants      Arabidopsis thaliana  48350
2      Eukaryota;Plants;Land Plants      Medicago truncatula  57661
3      Eukaryota;Plants;Land Plants      Glycine max          71525
4      Eukaryota;Plants;Land Plants      Solanum lycopersicum  36010
..
96      Eukaryota;Fungi;Ascomycetes    Yarrowia lipolytica CLIB122   6472
97      Eukaryota;Fungi;Ascomycetes      Kluyveromyces lactis    5085
98      Eukaryota;Fungi;Ascomycetes    Debaryomyces hansenii CBS767   6290
99      Eukaryota;Animals;Fishes      Oreochromis niloticus   58088
100     Eukaryota;Animals;Fishes      Ictalurus punctatus    47978

[101 rows x 3 columns]

```

```
[61]: data.loc[100,"CDS"]
```

```
[61]: 47978
```

```

[62]: # Change/ update a value in a dataframe
      data.loc[100,"CDS"]=3.83

```

```
[63]: data.loc[100,"CDS"]
```

```
[63]: 3.83
```

2.9 Filter with conditional logic using loc []

```

[64]: for i in data[["#Organism Name", "Organism Groups", "Release Date",
    ↪ "BioProject", "Level"]]:
      print(i.upper())
      print(data[i].unique())
      print()

```

```
#ORGANISM NAME
```

```

['Emiliania huxleyi CCMP1516' 'Arabidopsis thaliana' 'Medicago truncatula'
... 'Saccharomyces cerevisiae M5' 'Saccharomyces cerevisiae Lalvin L2056'
'Saccharomyces cerevisiae NRRL Y-12632']

```


ORGANISM GROUPS

```
['Eukaryota;Protists;Other Protists' 'Eukaryota;Plants;Land Plants'
'Eukaryota;Fungi;Ascomycetes' 'Eukaryota;Fungi;Basidiomycetes'
'Eukaryota;Protists;Kinetoplasts' 'Eukaryota;Protists;Apicomplexans'
'Eukaryota;Fungi;Other Fungi' 'Eukaryota;Animals;Roundworms'
'Eukaryota;Animals;Insects' 'Eukaryota;Animals;Fishes'
'Eukaryota;Animals;Other Animals' 'Eukaryota;Animals;Mammals'
'Eukaryota;Other;Other' 'Eukaryota;Animals;Amphibians'
'Eukaryota;Animals;Birds' 'Eukaryota;Plants;Green Algae'
'Eukaryota;Animals;Flatworms' 'Eukaryota;Animals;Reptiles'
'Eukaryota;Plants;Other Plants']
```

RELEASE DATE

```
['2013-04-19T00:00:00Z' '2001-08-13T00:00:00Z' '2011-08-12T00:00:00Z' ...
'2016-04-06T00:00:00Z' '2018-02-10T00:00:00Z' '2015-05-12T00:00:00Z']
```

BIOPROJECT

```
['PRJNA77753' 'PRJNA10719' 'PRJNA10791' ... 'PRJEB6529' 'PRJEB8455'
'PRJEB7675']
```

LEVEL

```
['Scaffold' 'Chromosome' 'Contig' 'Complete']
```

```
[65]: Scaffold = data["BioProject"] == "Scaffold"
```

```
[66]: Scaffold.value_counts()
```

```
[66]: BioProject
False    5131
Name: count, dtype: int64
```

```
[67]: data
```

```
[67]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

Strain	BioSample	BioProject	Assembly	Level \
--------	-----------	------------	----------	---------

0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%	Replicons \
0	167.6760	64.5000	NaN
1	119.6680	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...
2	412.9240	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...
3	978.9720	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...
4	823.7860	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...
...
5126	12.0392	38.1000	NaN
5127	11.7086	38.1000	NaN
5128	11.8194	38.0000	NaN
5129	11.7081	38.1000	NaN
5130	11.8189	38.1000	NaN

	WGS	Scaffolds	CDS	Release Date \
0	AHAL01	7795	38554.0	2013-04-19T00:00:00Z
1	NaN	7	48350.0	2001-08-13T00:00:00Z
2	APN001	2187	57661.0	2011-08-12T00:00:00Z
3	ACUP02	1191	71525.0	2010-01-05T00:00:00Z
4	AEKE02	3224	36010.0	2010-12-10T00:00:00Z
...
5126	FXDE01	1925	0.0	2017-04-21T00:00:00Z
5127	FXCK01	590	0.0	2017-04-21T00:00:00Z
5128	FXC001	2013	0.0	2017-04-21T00:00:00Z
5129	FXCP01	1028	0.0	2017-04-21T00:00:00Z
5130	FXCG01	799	0.0	2017-04-21T00:00:00Z

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

5128 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

                                RefSeq FTP
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
5126                                     NaN
5127                                     NaN
5128                                     NaN
5129                                     NaN
5130                                     NaN

```

[5131 rows x 16 columns]

```
[68]: data.iloc[:10, 7]
```

```

[68]: 0      167.67600
      1      119.66800
      2      412.92400
      3      978.97200
      4      823.78600
      5     9788.86000
      6      374.42300
      7    15344.70000
      8     2135.08000
      9         7.66146
      Name: Size(Mb), dtype: float64

```

```
[69]: data.loc[data['#Organism Name'] == 'Arabidopsis thaliana', ['Size(Mb)']]
```

```

[69]:      Size(Mb)
      1      119.6680
    3334    118.8910
    3336     96.5002
    3747     98.0662
    3748     96.2565
    3936     96.6940
    3938    116.8460
    4064    127.4190
    4065    244.5830

```

```
[70]: data.loc[data["Level"] == "Scaffold"]
```

[70]:

#Organism Name			Organism Groups \		
0	Emiliana huxleyi	CCMP1516	Eukaryota;Protists;Other Protists		
5	Hordeum vulgare		Eukaryota;Plants;Land Plants		
14	Aspergillus nidulans	FGSC A4	Eukaryota;Fungi;Ascomycetes		
18	Trypanosoma cruzi		Eukaryota;Protists;Kinetoplasts		
19	Giardia lamblia	ATCC 50803	Eukaryota;Protists;Other Protists		
...		
5126	Saccharomyces cerevisiae		Eukaryota;Fungi;Ascomycetes		
5127	Saccharomyces cerevisiae		Eukaryota;Fungi;Ascomycetes		
5128	Saccharomyces cerevisiae		Eukaryota;Fungi;Ascomycetes		
5129	Saccharomyces cerevisiae		Eukaryota;Fungi;Ascomycetes		
5130	Saccharomyces cerevisiae		Eukaryota;Fungi;Ascomycetes		
	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
5	NaN	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold
14	FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2	Scaffold
18	CL Brener	SAMN02953627	PRJNA11755	GCA_000209065.1	Scaffold
19	WB C6	SAMN02952905	PRJNA1439	GCA_000002435.1	Scaffold
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold
	Size(Mb)	GC%		Replicons	WGS \
0	167.6760	64.5000		NaN	AHAL01
5	9788.8600	0.0000		NaN	FJWB02
14	30.2760	50.2721	mitochondrion	MT:NC_017896.1/JQ435097.1	AACD01
18	89.9375	51.7000		NaN	AAHK01
19	11.2136	49.2000		NaN	AACB02
...
5126	12.0392	38.1000		NaN	FXDE01
5127	11.7086	38.1000		NaN	FXCK01
5128	11.8194	38.0000		NaN	FXC001
5129	11.7081	38.1000		NaN	FXCP01
5130	11.8189	38.1000		NaN	FXCG01
	Scaffolds	CDS	Release Date \		
0	7795	38554.0	2013-04-19T00:00:00Z		
5	72295	0.0	2016-08-25T00:00:00Z		
14	91	9561.0	2003-04-07T00:00:00Z		
18	29495	19607.0	2005-07-14T00:00:00Z		
19	92	6502.0	2003-03-26T00:00:00Z		
...		
5126	1925	0.0	2017-04-21T00:00:00Z		

5127	590	0.0	2017-04-21T00:00:00Z
5128	2013	0.0	2017-04-21T00:00:00Z
5129	1028	0.0	2017-04-21T00:00:00Z
5130	799	0.0	2017-04-21T00:00:00Z

		GenBank FTP	\
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...		
5	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...		
14	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...		
18	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...		
19	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...		
...	...		
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...		
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...		
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...		
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...		
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...		

		RefSeq FTP	
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...		
5		NaN	
14	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...		
18	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...		
19	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...		
...	...		
5126		NaN	
5127		NaN	
5128		NaN	
5129		NaN	
5130		NaN	

[3350 rows x 16 columns]

```
[71]: data.loc[data["CDS"] > data["Scaffolds"]]
```

[71]:	#Organism Name	Organism Groups	\
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists	
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants	
2	Medicago truncatula	Eukaryota;Plants;Land Plants	
3	Glycine max	Eukaryota;Plants;Land Plants	
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants	
...	
4980	Saccharomyces cerevisiae RM11-1a	Eukaryota;Fungi;Ascomycetes	
5002	Saccharomyces cerevisiae AWRI1631	Eukaryota;Fungi;Ascomycetes	
5004	Saccharomyces cerevisiae Fosters0	Eukaryota;Fungi;Ascomycetes	
5020	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	
5022	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
...
4980	RM11-1a	SAMN02953602	PRJNA13674	GCA_000149365.1	Scaffold
5002	AWRI1631	SAMN02953734	PRJNA30553	GCA_000182175.1	Contig
5004	Fosters0	SAMN02981272	PRJNA48567	GCA_000326005.1	Contig
5020	131	SAMN06011376	PRJNA353176	GCA_001983315.1	Contig
5022	I-328	SAMN07811430	PRJNA414946	GCA_002804325.1	Contig

	Size(Mb)	GC%	Replicons \
0	167.6760	64.5000	NaN
1	119.6680	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...
2	412.9240	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...
3	978.9720	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...
4	823.7860	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...
...
4980	11.7370	38.3000	NaN
5002	11.1769	38.2000	NaN
5004	11.6849	38.2000	NaN
5020	12.0056	38.1000	NaN
5022	11.5053	38.3000	NaN

	WGS	Scaffolds	CDS	Release Date \
0	AHAL01	7795	38554.0	2013-04-19T00:00:00Z
1	NaN	7	48350.0	2001-08-13T00:00:00Z
2	APN001	2187	57661.0	2011-08-12T00:00:00Z
3	ACUP02	1191	71525.0	2010-01-05T00:00:00Z
4	AEKE02	3224	36010.0	2010-12-10T00:00:00Z
...
4980	AAEG01	17	5377.0	2005-03-16T00:00:00Z
5002	ABSV01	2484	5451.0	2008-09-30T00:00:00Z
5004	AEEZ01	222	3575.0	2011-02-11T00:00:00Z
5020	MQVN01	236	5445.0	2017-02-02T00:00:00Z
5022	PEJR01	531	5327.0	2017-12-04T00:00:00Z

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...

```

4980 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5002 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5004 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5020 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
5022 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...

```

```

                                RefSeq FTP
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
4980                                     NaN
5002                                     NaN
5004                                     NaN
5020                                     NaN
5022                                     NaN

```

[1661 rows x 16 columns]

```
[72]: data.loc[data["CDS"] > data["Scaffolds"], "Level"]
```

```

[72]: 0      Scaffold
      1      Chromosome
      2      Chromosome
      3      Chromosome
      4      Chromosome
      ...
      4980      Scaffold
      5002      Contig
      5004      Contig
      5020      Contig
      5022      Contig
      Name: Level, Length: 1661, dtype: object

```

```
[73]: filtered_data = data[(data["Organism Groups"] == "Arabidopsis thaliana") &
    ↪ (data["CDS"] == "Scaffold")]
```

```
[74]: print(filtered_data)
```

```

Empty DataFrame
Columns: [#Organism Name, Organism Groups, Strain, BioSample, BioProject,
Assembly, Level, Size(Mb), GC%, Replicons, WGS, Scaffolds, CDS, Release Date,
GenBank FTP, RefSeq FTP]
Index: []

```

```
[75]: # Both statement need to evaluate to true
filtered_data = data.loc[(data["Organism Groups"] == "Chromosome") &
↳ (data["Level"] == "Eukaryota;Plants;Land Plants")]
↳ .sort_values(by= "CDS",
↳ ascending=False)
```

```
[76]: print(filtered_data)
```

Empty DataFrame

Columns: [#Organism Name, Organism Groups, Strain, BioSample, BioProject, Assembly, Level, Size(Mb), GC%, Replicons, WGS, Scaffolds, CDS, Release Date, GenBank FTP, RefSeq FTP]

Index: []

```
[77]: rpg = data["Organism Groups"] == "Eukaryota;Plants;Land Plants"
```

```
[78]: filtered_data = data[rpg]
```

```
[79]: print(filtered_data)
```

	#Organism Name	Organism Groups	Strain	\
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants	NaN	
2	Medicago truncatula	Eukaryota;Plants;Land Plants	A17	
3	Glycine max	Eukaryota;Plants;Land Plants	NaN	
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants	NaN	
5	Hordeum vulgare	Eukaryota;Plants;Land Plants	NaN	
...	
4341	Oryza sativa Indica Group	Eukaryota;Plants;Land Plants	NaN	
4367	Oryza sativa Indica Group	Eukaryota;Plants;Land Plants	NaN	
4368	Oryza sativa Indica Group	Eukaryota;Plants;Land Plants	NaN	
4389	Oryza sativa	Eukaryota;Plants;Land Plants	NaN	
4390	Oryza sativa	Eukaryota;Plants;Land Plants	NaN	

	BioSample	BioProject	Assembly	Level	Size(Mb)	\
1	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome	119.668	
2	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome	412.924	
3	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome	978.972	
4	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome	823.786	
5	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold	9788.860	
...	
4341	SAMN04275185	PRJNA305605	GCA_001611195.1	Scaffold	351.225	
4367	SAMN04275173	PRJNA305604	GCA_001611235.1	Scaffold	352.227	
4368	SAMN04275184	PRJNA305606	GCA_001611255.1	Scaffold	331.819	
4389	SAMN04524138	PRJNA313502	GCA_001648735.1	Scaffold	307.225	
4390	SAMN04538187	PRJNA313502	GCA_001648745.1	Scaffold	295.390	

	GC%	Replicons	WGS	\
1	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN	
2	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001	

3	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02
4	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02
5	0.0000		NaN FJWB02
...
4341	43.1000		NaN LQHG01
4367	41.8000		NaN LQHE01
4368	42.0000		NaN LQHF01
4389	41.4000		NaN LVCG01
4390	41.0000		NaN LVCH01

	Scaffolds	CDS	Release Date \
1	7	48350.0	2001-08-13T00:00:00Z
2	2187	57661.0	2011-08-12T00:00:00Z
3	1191	71525.0	2010-01-05T00:00:00Z
4	3224	36010.0	2010-12-10T00:00:00Z
5	72295	0.0	2016-08-25T00:00:00Z
...
4341	39669	0.0	2016-04-05T00:00:00Z
4367	116212	0.0	2016-04-05T00:00:00Z
4368	75421	0.0	2016-04-05T00:00:00Z
4389	55637	0.0	2016-05-19T00:00:00Z
4390	64800	0.0	2016-05-19T00:00:00Z

	GenBank FTP \
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
...	...
4341	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
4367	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
4368	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
4389	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
4390	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...

	RefSeq FTP
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
5	NaN
...	...
4341	NaN
4367	NaN
4368	NaN
4389	NaN
4390	NaN

[430 rows x 16 columns]

```
[80]: # Filter with OR
# Or needs at least one contition to be true
scaffold = data["Level"] == "Scaffold"
chromosome = data["Level"] == "Chromosome"
data.loc[scaffold | chromosome]
```

```
[80]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
5	Hordeum vulgare	Eukaryota;Plants;Land Plants
14	Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes
18	Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts
19	Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
5	NaN	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold
14	FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2	Scaffold
18	CL Brener	SAMN02953627	PRJNA11755	GCA_000209065.1	Scaffold
19	WB C6	SAMN02952905	PRJNA1439	GCA_000002435.1	Scaffold
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%	Replicons	WGS \
0	167.6760	64.5000	NaN	AHAL01
5	9788.8600	0.0000	NaN	FJWB02
14	30.2760	50.2721	mitochondrion MT:NC_017896.1/JQ435097.1	AACD01
18	89.9375	51.7000	NaN	AAHK01
19	11.2136	49.2000	NaN	AACB02
...
5126	12.0392	38.1000	NaN	FXDE01
5127	11.7086	38.1000	NaN	FXCK01
5128	11.8194	38.0000	NaN	FXCO01
5129	11.7081	38.1000	NaN	FXCP01
5130	11.8189	38.1000	NaN	FXCG01

	Scaffolds	CDS	Release Date \
0	7795	38554.0	2013-04-19T00:00:00Z
5	72295	0.0	2016-08-25T00:00:00Z
14	91	9561.0	2003-04-07T00:00:00Z
18	29495	19607.0	2005-07-14T00:00:00Z
19	92	6502.0	2003-03-26T00:00:00Z
...
5126	1925	0.0	2017-04-21T00:00:00Z
5127	590	0.0	2017-04-21T00:00:00Z
5128	2013	0.0	2017-04-21T00:00:00Z
5129	1028	0.0	2017-04-21T00:00:00Z
5130	799	0.0	2017-04-21T00:00:00Z

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
14	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
18	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
19	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
5	NaN
14	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
18	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
19	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
5126	NaN
5127	NaN
5128	NaN
5129	NaN
5130	NaN

[3350 rows x 16 columns]

```
[81]: # Filter with bot OR and AND condition
```

```
data["Release Date"] = pd.to_datetime(data["Release Date"])
```

```
[82]: since_2017 = data["Release Date"] >= '2017-01-01'
scaffold_or_chromosome = (data["Level"] == "Scaffold") | (data["Level"] == "Chromosome")
filtered_data = data.loc[(since_2017) & (scaffold_or_chromosome)]
```

```
[83]: data.loc[since_2017 & scaffold | chromosome]
```

```
[83]:
```

	#Organism Name	Organism Groups	Strain \
122	Schmidtea mediterranea	Eukaryota;Animals;Flatworms	S2
157	Meloidogyne incognita	Eukaryota;Animals;Roundworms	NaN
159	Petromyzon marinus	Eukaryota;Animals;Fishes	W1
187	Lactuca sativa	Eukaryota;Plants;Land Plants	NaN
201	Ambystoma mexicanum	Eukaryota;Animals;Amphibians	DD151
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	NaN
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	ZP848
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	VKMY373
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	245
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	M2ON0800_1A

	BioSample	BioProject	Assembly	Level	Size(Mb) \
122	SAMN06660624	PRJNA379262	GCA_002600895.1	Scaffold	942.0420
157	SAMEA104032784	PRJEB8714	GCA_900182535.1	Scaffold	183.5320
159	SAMN06131802	PRJNA357048	GCA_002833325.1	Scaffold	1130.4200
187	SAMN01730979	PRJNA432228	GCA_002870075.1	Scaffold	2384.1900
201	SAMN06554622	PRJNA378970	GCA_002915635.1	Scaffold	32393.6000
...
5126	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold	12.0392
5127	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold	11.7086
5128	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold	11.8194
5129	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold	11.7081
5130	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold	11.8189

	GC%	Replicons	WGS	Scaffolds	CDS \
122	29.7000	NaN	NNSW01	1990	0.0
157	0.0000	NaN	FXSY01	12091	0.0
159	0.0000	NaN	PIZI01	12061	0.0
187	35.0002	chloroplast:NC_007578.1/AP007232.1	NBSK01	11453	38378.0
201	40.3000	NaN	PGSH01	125724	0.0
...
5126	38.1000	NaN	FXDE01	1925	0.0
5127	38.1000	NaN	FXCK01	590	0.0
5128	38.0000	NaN	FXC001	2013	0.0
5129	38.1000	NaN	FXCP01	1028	0.0
5130	38.1000	NaN	FXCG01	799	0.0

Release Date \

```

122 2017-10-18 00:00:00+00:00
157 2017-05-05 00:00:00+00:00
159 2017-12-11 00:00:00+00:00
187 2018-01-09 00:00:00+00:00
201 2018-02-05 00:00:00+00:00
...
5126 2017-04-21 00:00:00+00:00
5127 2017-04-21 00:00:00+00:00
5128 2017-04-21 00:00:00+00:00
5129 2017-04-21 00:00:00+00:00
5130 2017-04-21 00:00:00+00:00

```

```

                                     GenBank FTP \
122 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
157 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
159 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
187 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
201 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
...
5126 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

```

                                     RefSeq FTP
122                                     NaN
157                                     NaN
159                                     NaN
187 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002...
201                                     NaN
...
5126                                     NaN
5127                                     NaN
5128                                     NaN
5129                                     NaN
5130                                     NaN

```

[898 rows x 16 columns]

```
[84]: # isin()
top_3_Level = data.Level.isin(["Chromosome", "Scaffold", "Contig"])
```

```
[85]: data.loc[top_3_Level]
```

```

[85]:          #Organism Name          Organism Groups \
0      Emiliania huxleyi CCMP1516  Eukaryota;Protists;Other Protists

```

5	Hordeum vulgare	Eukaryota;Plants;Land Plants
7	Triticum aestivum	Eukaryota;Plants;Land Plants
9	Pneumocystis carinii B80	Eukaryota;Fungi;Ascomycetes
12	Phanerochaete chrysosporium	Eukaryota;Fungi;Basidiomycetes
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
5	NaN	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold
7	NaN	SAMN07284949	PRJNA392179	GCA_002220415.2	Contig
9	B80	SAMN02380717	PRJNA223511	GCA_001477545.1	Contig
12	ATCC 20696	SAMN05790961	PRJNA343563	GCA_001910725.1	Contig
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%	Replicons	WGS	Scaffolds	CDS \
0	167.67600	64.5	NaN	AHAL01	7795	38554.0
5	9788.86000	0.0	NaN	FJWB02	72295	0.0
7	15344.70000	46.1	NaN	NMPL02	279439	0.0
9	7.66146	27.8	NaN	LFVZ01	62	3646.0
12	39.20510	56.5	NaN	MJGA01	174	0.0
...
5126	12.03920	38.1	NaN	FXDE01	1925	0.0
5127	11.70860	38.1	NaN	FXCK01	590	0.0
5128	11.81940	38.0	NaN	FXC001	2013	0.0
5129	11.70810	38.1	NaN	FXCP01	1028	0.0
5130	11.81890	38.1	NaN	FXCG01	799	0.0

	Release Date \
0	2013-04-19 00:00:00+00:00
5	2016-08-25 00:00:00+00:00
7	2017-07-13 00:00:00+00:00
9	2015-12-22 00:00:00+00:00
12	2016-12-20 00:00:00+00:00
...	...
5126	2017-04-21 00:00:00+00:00
5127	2017-04-21 00:00:00+00:00
5128	2017-04-21 00:00:00+00:00

```
5129 2017-04-21 00:00:00+00:00
5130 2017-04-21 00:00:00+00:00
```

```

                                GenBank FTP  \
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
7      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
9      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
12     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
...
5126   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
```

```

                                RefSeq FTP
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
5                                           NaN
7                                           NaN
9      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...
12                                           NaN
...
5126                                           NaN
5127                                           NaN
5128                                           NaN
5129                                           NaN
5130                                           NaN
```

[4339 rows x 16 columns]

```
[86]: Level = data["Organism Groups"].isin(["Eukaryota;Protists;Other Protists",
      ↪ "Eukaryota;Plants;Land Plants", "Eukaryota;Fungi;Ascomycetes"])
      filtered_data = data[Level]
```

```
[87]: filtered_data
```

```

[87]:      #Organism Name      Organism Groups  \
0      Emiliana huxleyi CCMP1516  Eukaryota;Protists;Other Protists
1      Arabidopsis thaliana      Eukaryota;Plants;Land Plants
2      Medicago truncatula      Eukaryota;Plants;Land Plants
3      Glycine max              Eukaryota;Plants;Land Plants
4      Solanum lycopersicum      Eukaryota;Plants;Land Plants
...
5126   Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5127   Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5128   Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
```

5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%	Replicons \
0	167.6760	64.5000	NaN
1	119.6680	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...
2	412.9240	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...
3	978.9720	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...
4	823.7860	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...
...
5126	12.0392	38.1000	NaN
5127	11.7086	38.1000	NaN
5128	11.8194	38.0000	NaN
5129	11.7081	38.1000	NaN
5130	11.8189	38.1000	NaN

	WGS	Scaffolds	CDS	Release Date \
0	AHAL01	7795	38554.0	2013-04-19 00:00:00+00:00
1	NaN	7	48350.0	2001-08-13 00:00:00+00:00
2	APN001	2187	57661.0	2011-08-12 00:00:00+00:00
3	ACUP02	1191	71525.0	2010-01-05 00:00:00+00:00
4	AEKE02	3224	36010.0	2010-12-10 00:00:00+00:00
...
5126	FXDE01	1925	0.0	2017-04-21 00:00:00+00:00
5127	FXCK01	590	0.0	2017-04-21 00:00:00+00:00
5128	FXC001	2013	0.0	2017-04-21 00:00:00+00:00
5129	FXCP01	1028	0.0	2017-04-21 00:00:00+00:00
5130	FXCG01	799	0.0	2017-04-21 00:00:00+00:00

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...


```

4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...
5126  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

                                     RefSeq FTP
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
5126                                     NaN
5127                                     NaN
5128                                     NaN
5129                                     NaN
5130                                     NaN

```

[2881 rows x 16 columns]

```

[88]: import pandas as pd

# Assuming "Release Date" is your datetime column
data['Release Date'] = pd.to_datetime(data['Release Date']) # Convert to_
↳datetime if not already

# Filter based on the date range
date_range_90s = data['Release Date'].between('1990-01-01', '1999-12-31')
date_range_00s = data['Release Date'].between('2000-01-01', '2009-12-31')
date_range_10s = data['Release Date'].between('2010-01-01', '2019-12-31')

# Now you can use these boolean masks to filter your DataFrame
data_90s = data[date_range_90s]
data_00s = data[date_range_00s]
data_10s = data[date_range_10s]

```

```

[90]: data['Release Date'] = pd.to_datetime(data['Release Date'])

```

```

[91]: between_2M_10M = (data['Size(Mb)'] >= 2) & (data['Size(Mb)'] <= 10)
date_range_90s = (data['Release Date'].dt.year >= 1990) & (data['Release Date'].
↳dt.year < 2000)

```

```

[92]: filtered_organism_groups = data.loc[between_2M_10M & date_range_90s, 'Organism_
↳Groups']

```

```
[93]: print(filtered_organism_groups)
```

```
Series([], Name: Organism Groups, dtype: object)
```

```
[94]: data.loc[between_2M_10M & date_range_90s, "Organism Groups"].
      ↪value_counts(normalize=True)
```

```
[94]: Series([], Name: proportion, dtype: float64)
```

```
[95]: data.loc[-data["Level"].isin(["Scaffold", "Chromosome", "Contig"])]
```

```
[95]:
```

	#Organism Name	Organism Groups \
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
6	Oryza sativa Japonica Group	Eukaryota;Plants;Land Plants
...
4969	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
4970	Saccharomyces cerevisiae W303	Eukaryota;Fungi;Ascomycetes
4971	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
4972	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
4973	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
6	NaN	SAMD00000397	PRJNA12269	GCA_001433935.1	Chromosome
...
4969	beer002	SAMN05190363	PRJNA323691	GCA_001741815.1	Chromosome
4970	W303	SAMN05199423	PRJNA324291	GCA_002163515.1	Chromosome
4971	DBVPG6044	SAMN06218131	PRJNA360861	GCA_002271905.1	Chromosome
4972	Y12	SAMN06218130	PRJNA360861	GCA_002271945.1	Chromosome
4973	CEN.PK113-7D	SAMN07328112	PRJNA393501	GCA_002571405.1	Chromosome

	Size(Mb)	GC%	Replicons \
1	119.6680	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...
2	412.9240	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...
3	978.9720	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...
4	823.7860	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...
6	374.4230	43.5769	chromosome 1:NC_029256.1/AP014957.1; chromosom...
...
4969	12.1359	38.2781	chromosome I:CM007167.1; chromosome II:CM00716...
4970	12.4334	38.2030	chromosome I:CM007964.1; chromosome II:CM00796...
4971	11.7774	38.3553	chromosome I:CM008214.1; chromosome II:CM00821...
4972	11.6880	38.3125	chromosome I:CM008200.1; chromosome II:CM00820...

4973 12.1478 38.1989 chromosome I:CM008859.1; chromosome II:CM00886...

	WGS	Scaffolds	CDS	Release Date	\
1	NaN	7	48350.0	2001-08-13 00:00:00+00:00	
2	APN001	2187	57661.0	2011-08-12 00:00:00+00:00	
3	ACUP02	1191	71525.0	2010-01-05 00:00:00+00:00	
4	AEKE02	3224	36010.0	2010-12-10 00:00:00+00:00	
6	NaN	58	41070.0	2015-10-10 00:00:00+00:00	
...	
4969	MCAA01	195	0.0	2016-09-12 00:00:00+00:00	
4970	LYZE01	21	0.0	2017-06-01 00:00:00+00:00	
4971	MZZF01	23	0.0	2017-08-11 00:00:00+00:00	
4972	MZZG01	23	0.0	2017-08-11 00:00:00+00:00	
4973	PCYP01	20	0.0	2017-10-17 00:00:00+00:00	

	GenBank FTP	\
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
6	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...	
...	...	
4969	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...	
4970	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...	
4971	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...	
4972	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...	
4973	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...	

	RefSeq FTP
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
6	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...
...	...
4969	NaN
4970	NaN
4971	NaN
4972	NaN
4973	NaN

[792 rows x 16 columns]

```
[96]: data.loc[-data["Level"].isin(["Scaffold", "Chromosome", "Contig"])] &
↳ date_range_90s]
```

```
[96]:
```

	#Organism Name	Organism Groups \
11	Saccharomyces cerevisiae S288C	Eukaryota;Fungi;Ascomycetes
17	Leishmania major strain Friedlin	Eukaryota;Protists;Kinetoplasts
26	Plasmodium falciparum 3D7	Eukaryota;Protists;Apicomplexans
88	Equus caballus	Eukaryota;Animals;Mammals

	Strain	BioSample	BioProject	Assembly	Level \
11	S288C	NaN	PRJNA43747	GCA_000146045.2	Complete
17	Friedlin	SAMEA3138173	PRJNA10724	GCA_000002725.2	Complete
26	3D7	NaN	PRJNA430688	GCA_000002765.1	Chromosome
88	Twilight	SAMN02953672	PRJNA421018	GCA_002863925.1	Chromosome

	Size(Mb)	GC%	Replicons \
11	12.1571	38.1556	chromosome I:NC_001133.9/BK006935.2; chromosom...
17	32.8551	59.7114	chromosome 1:NC_001905.3/FR796397.1; chromosom...
26	23.2703	19.3630	chromosome 1:NC_004325.1/AL844501.1; chromosom...
88	2506.9700	42.0000	mitochondrion MT:NC_001640.1/X79547.1

	WGS	Scaffolds	CDS	Release Date \
11	NaN	17	6002.0	1999-11-08 00:00:00+00:00
17	NaN	36	8316.0	1998-06-29 00:00:00+00:00
26	NaN	15	5339.0	1998-12-17 00:00:00+00:00
88	PJAA01	4701	13.0	1994-11-22 00:00:00+00:00

	GenBank FTP \
11	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
17	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
26	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
88	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...

	RefSeq FTP
11	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
17	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
26	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
88	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002...

```
[97]: data.loc[-data["Level"].isin(["Scaffold", "Chromosome", "Contig"]) &
↳date_range_90s, "Organism Groups"]
```

```
[97]: 11      Eukaryota;Fungi;Ascomycetes
      17      Eukaryota;Protists;Kinetoplasts
      26      Eukaryota;Protists;Apicomplexans
      88      Eukaryota;Animals;Mammals
      Name: Organism Groups, dtype: object
```

2.10 Most Useful Methods for Data Analysis

[98]: `# query()`

```
data.query("CDS < Scaffolds")
```

```
[98]:
```

	#Organism Name	Organism Groups \
5	Hordeum vulgare	Eukaryota;Plants;Land Plants
7	Triticum aestivum	Eukaryota;Plants;Land Plants
12	Phanerochaete chrysosporium	Eukaryota;Fungi;Basidiomycetes
18	Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts
36	Brugia malayi	Eukaryota;Animals;Roundworms
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
5	NaN	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold
7	NaN	SAMN07284949	PRJNA392179	GCA_002220415.2	Contig
12	ATCC 20696	SAMN05790961	PRJNA343563	GCA_001910725.1	Contig
18	CL Brener	SAMN02953627	PRJNA11755	GCA_000209065.1	Scaffold
36	NaN	SAMN02953652	PRJNA10729	GCA_000002995.2	Scaffold
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%	Replicons	WGS \
5	9788.8600	0.000	NaN	FJWB02
7	15344.7000	46.100	NaN	NMPL02
12	39.2051	56.500	NaN	MJGA01
18	89.9375	51.700	NaN	AAHK01
36	93.6591	31.499	mitochondrion MT:NC_004298.1/AF538716.1	AAQA01
...
5126	12.0392	38.100	NaN	FXDE01
5127	11.7086	38.100	NaN	FXCK01
5128	11.8194	38.000	NaN	FXCO01
5129	11.7081	38.100	NaN	FXCP01
5130	11.8189	38.100	NaN	FXCG01

	Scaffolds	CDS	Release Date \
5	72295	0.0	2016-08-25 00:00:00+00:00

7	279439	0.0	2017-07-13	00:00:00+00:00
12	174	0.0	2016-12-20	00:00:00+00:00
18	29495	19607.0	2005-07-14	00:00:00+00:00
36	24286	11472.0	2002-09-25	00:00:00+00:00
...
5126	1925	0.0	2017-04-21	00:00:00+00:00
5127	590	0.0	2017-04-21	00:00:00+00:00
5128	2013	0.0	2017-04-21	00:00:00+00:00
5129	1028	0.0	2017-04-21	00:00:00+00:00
5130	799	0.0	2017-04-21	00:00:00+00:00

GenBank FTP \

5	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			
7	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...			
12	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...			
18	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...			
36	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...			
...	...			
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...			

RefSeq FTP

5	NaN
7	NaN
12	NaN
18	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
36	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
5126	NaN
5127	NaN
5128	NaN
5129	NaN
5130	NaN

[3470 rows x 16 columns]

```
[99]: Level = data.query("Level == 'Scaffold'")
```

```
[100]: Level
```

```
[100]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
5	Hordeum vulgare	Eukaryota;Plants;Land Plants
14	Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes

18	Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts
19	Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
5	NaN	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold
14	FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2	Scaffold
18	CL Brener	SAMN02953627	PRJNA11755	GCA_000209065.1	Scaffold
19	WB C6	SAMN02952905	PRJNA1439	GCA_000002435.1	Scaffold
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M20N0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%		Replicons	WGS \
0	167.6760	64.5000		NaN	AHAL01
5	9788.8600	0.0000		NaN	FJWB02
14	30.2760	50.2721	mitochondrion MT:NC_017896.1/JQ435097.1		AACD01
18	89.9375	51.7000		NaN	AAHK01
19	11.2136	49.2000		NaN	AACB02
...
5126	12.0392	38.1000		NaN	FXDE01
5127	11.7086	38.1000		NaN	FXCK01
5128	11.8194	38.0000		NaN	FXC001
5129	11.7081	38.1000		NaN	FXCP01
5130	11.8189	38.1000		NaN	FXCG01

	Scaffolds	CDS	Release Date \
0	7795	38554.0	2013-04-19 00:00:00+00:00
5	72295	0.0	2016-08-25 00:00:00+00:00
14	91	9561.0	2003-04-07 00:00:00+00:00
18	29495	19607.0	2005-07-14 00:00:00+00:00
19	92	6502.0	2003-03-26 00:00:00+00:00
...
5126	1925	0.0	2017-04-21 00:00:00+00:00
5127	590	0.0	2017-04-21 00:00:00+00:00
5128	2013	0.0	2017-04-21 00:00:00+00:00
5129	1028	0.0	2017-04-21 00:00:00+00:00
5130	799	0.0	2017-04-21 00:00:00+00:00

```

                                GenBank FTP \
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
5      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
14     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
18     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
19     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...
5126   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130   ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

                                RefSeq FTP
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
5
14     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
18     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
19     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
5126
5127
5128
5129
5130

```

[3350 rows x 16 columns]

```
[101]: # unique() and nunique()

data.Level.unique()
```

```
[101]: array(['Scaffold', ' Chromosome', 'Contig', 'Complete'], dtype=object)
```

```
[102]: data.Level.nunique()
```

```
[102]: 4
```

```
[103]: data.nunique()
```

```
[103]: #Organism Name      3458
      Organism Groups    19
      Strain             3962
      BioSample          5047
      BioProject         3859
      Assembly           5131
```


Level	4
Size(Mb)	5101
GC%	1352
Replicons	1068
WGS	4790
Scaffolds	3390
CDS	1764
Release Date	1580
GenBank FTP	5125
RefSeq FTP	859
dtype:	int64

```
[104]: data["Release Date"].nunique()
```

```
[104]: 1580
```

```
[105]: for i in data.select_dtypes("object"):
        print(data[i].unique())
```

```
['Emiliania huxleyi CCMP1516' 'Arabidopsis thaliana' 'Medicago truncatula'
... 'Saccharomyces cerevisiae M5' 'Saccharomyces cerevisiae Lalvin L2056'
'Saccharomyces cerevisiae NRRL Y-12632']
['Eukaryota;Protists;Other Protists' 'Eukaryota;Plants;Land Plants'
'Eukaryota;Fungi;Ascomycetes' 'Eukaryota;Fungi;Basidiomycetes'
'Eukaryota;Protists;Kinetoplasts' 'Eukaryota;Protists;Apicomplexans'
'Eukaryota;Fungi;Other Fungi' 'Eukaryota;Animals;Roundworms'
'Eukaryota;Animals;Insects' 'Eukaryota;Animals;Fishes'
'Eukaryota;Animals;Other Animals' 'Eukaryota;Animals;Mammals'
'Eukaryota;Other;Other' 'Eukaryota;Animals;Amphibians'
'Eukaryota;Animals;Birds' 'Eukaryota;Plants;Green Algae'
'Eukaryota;Animals;Flatworms' 'Eukaryota;Animals;Reptiles'
'Eukaryota;Plants;Other Plants']
['CCMP1516' nan 'A17' ... 'VKMY373' '245' 'M2ON0800_1A']
['SAMN02744062' 'SAMN03081427' 'SAMN02299339' ... 'SAMEA3110473'
'SAMEA3110476' 'SAMEA3110463']
['PRJNA77753' 'PRJNA10719' 'PRJNA10791' ... 'PRJEB6529' 'PRJEB8455'
'PRJEB7675']
['GCA_000372725.1' 'GCA_000001735.1' 'GCA_000219495.2' ...
'GCA_900178485.1' 'GCA_900178495.1' 'GCA_900178505.1']
['Scaffold' 'Chromosome' 'Contig' 'Complete']
[nan
'chromosome 1:NC_003070.9/CP002684.1; chromosome 2:NC_003071.7/CP002685.1;
chromosome 3:NC_003074.8/CP002686.1; chromosome 4:NC_003075.7/CP002687.1;
chromosome 5:NC_003076.8/CP002688.1; mitochondrion MT:NC_001284.2/Y08501.2;
chloroplast Pltd:NC_000932.1/AP000423.1'
'chromosome 1:NC_016407.2/CM001217.2; chromosome 2:NC_016408.2/CM001218.2;
chromosome 3:NC_016409.2/CM001219.2; chromosome 4:NC_016410.2/CM001220.2;
chromosome 5:NC_016411.2/CM001221.2; chromosome 6:NC_016412.2/CM001222.2;
```

chromosome 7:NC_016413.2/CM001223.2; chromosome 8:NC_016414.2/CM001224.2;
chloroplast Pltd:NC_003119.6/AC093544.8'

...

'chromosome I:CM008200.1; chromosome II:CM008201.1; chromosome III:CM008202.1;
chromosome IV:CM008203.1; chromosome V:CM008204.1; chromosome VI:CM008205.1;
chromosome VII:CM008206.1; chromosome VIII:CM008207.1; chromosome IX:CM008208.1;
chromosome X:CM008209.1; chromosome XI:CM008210.1; chromosome XIII:CM008211.1;
chromosome XV:CM008212.1; chromosome XVI:CM008213.1'

'chromosome I:CM008859.1; chromosome II:CM008860.1; chromosome III:CM008861.1;
chromosome IV:CM008862.1; chromosome V:CM008863.1; chromosome VI:CM008864.1;
chromosome VII:CM008865.1; chromosome VIII:CM008866.1; chromosome IX:CM008867.1;
chromosome X:CM008868.1; chromosome XI:CM008869.1; chromosome XII:CM008870.1;
chromosome XIII:CM008871.1; chromosome XIV:CM008872.1; chromosome XV:CM008873.1;
chromosome XVI:CM008874.1; plasmid p2uM:CM008875.1; mitochondrion MT:CM008876.1'

'plasmid p2um:CM007616.1']

['AHAL01' nan 'APN001' ... 'FXC001' 'FXCP01' 'FXCG01']

['ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/372/725/GCA_000372725.1_Emiliana_huxleyi_CCMP1516_main_genome_assembly_v1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/735/GCA_000001735.1_TAIR10'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/219/495/GCA_000219495.2_MedtrA17_4.0'

...

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/178/485/GCA_900178485.1_CF_SoucheMTF2386_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/178/495/GCA_900178495.1_DA_SoucheMTF2248_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/178/505/GCA_900178505.1_BC_Souche1152_V1']

['ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/372/725/GCF_000372725.1_Emiliana_huxleyi_CCMP1516_main_genome_assembly_v1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/735/GCF_000001735.3_TAIR10'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/219/495/GCF_000219495.3_MedtrA17_4.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/515/GCF_000004515.4_Glycine_max_v2.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/188/115/GCF_000188115.3_SL2.50'

nan

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/433/935/GCF_001433935.1_IRGSP-1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/005/GCF_000005005.2_B73_RefGen_v4'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/477/545/GCF_001477545.1_Pneu_cari_B80_V3'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/945/GCF_000002945.1_ASM294v2'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/182/965/GCF_000182965.3_ASM18296v3'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/205/GCF_000149205.2_ASM1492
 0v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/655/GCF_000002655.1_ASM265v
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/182/925/GCF_000182925.2_NC12'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/725/GCF_000002725.2_ASM272v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/209/065/GCF_000209065.1_ASM2090
 6v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/435/GCF_000002435.1_GL2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/208/925/GCF_000208925.1_JCVI_ES
 G2_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/210/295/GCF_000210295.1_ASM2102
 9v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/499/545/GCF_000499545.2_ETH001'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/165/345/GCF_000165345.1_ASM1653
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/565/GCF_000006565.2_TGA4'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/002/375/GCF_900002375.1_PBANKAO
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/765/GCF_000002765.3_ASM276v
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/002/335/GCF_900002335.2_PCHAS01'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/355/GCF_000006355.1_ASM635v
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/415/GCF_000002415.2_ASM241v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/165/395/GCF_000165395.1_ASM1653
 9v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/225/GCF_000003225.3_ASM322v
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/165/365/GCF_000165365.1_ASM1653
 6v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/091/225/GCF_000091225.1_ASM9122
 v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/985/GCF_000002985.6_WBcel23
 5'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/995/GCF_000002995.3_ASM299v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/204/515/GCF_002204515.2_AaegL5.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/876/365/GCF_001876365.2_canu_80
 X_arrow2.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/575/GCF_000005575.2_AgamP3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/215/GCF_000001215.4_Release
 _6_plus_IS01_MT'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/195/GCF_000002195.4_Amel_4.
 5'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/035/GCF_000002035.6_GRCz11'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/224/145/GCF_000224145.3_KH'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/405/GCF_000001405.37_GRCh38.p11'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/635/GCF_000001635.26_GRCm38.p6'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/615/GCF_000149615.1_ASM14961v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/405/GCF_000149405.2_ASM14940v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/695/GCF_000004695.1_dicty_2.7'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/315/625/GCF_000315625.1_Guith1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/240/135/GCF_000240135.3_ASM24013v3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/091/045/GCF_000091045.1_ASM9104v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/495/GCF_000002495.2_MG8'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/180/615/GCF_000180615.1_FUGU5'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/002/385/GCF_900002385.1_PY17X01'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/237/345/GCF_000237345.1_ASM23734v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/328/475/GCF_000328475.2_Umaydis521_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/182/895/GCF_000182895.1_CC3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/895/GCF_000001895.5_Rnor_6.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/091/025/GCF_000091025.4_ASM9102v4'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/815/GCF_000003815.1_Version_2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/625/GCF_000151625.1_ASM15162v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/181/335/GCF_000181335.3_Felis_catus_9.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/091/205/GCF_000091205.1_ASM9120v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/195/GCF_000004195.3_Xenopus_tropicalis_v9.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/663/975/GCF_001663975.1_Xenopus_laevis_v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/055/GCF_000003055.6_Bos_taurus_UMD_3.1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/298/735/GCF_000298735.2_Oar_v4.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/025/GCF_000003025.6_Sscrofa11.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/285/GCF_000002285.3_CanFam3

.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/235/GCF_000002235.4_Spur_4.
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/335/GCF_000151335.1_JCVI-
 cpa1-1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/775/GCF_000002775.3_Poptr2_
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/195/GCF_000003195.3_Sorghum
 _bicolor_NCBIV3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/275/GCF_000002275.2_Ornitho
 rhynchus_anatinus_5.0.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/315/GCF_000002315.4_Gallus_
 gallus-5.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/698/965/GCF_000698965.1_ASM6989
 6v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/337/955/GCF_000337955.1_F_pereg
 rinus_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/863/925/GCF_002863925.1_EquCab3
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/595/GCF_000002595.1_v3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/975/GCF_000005975.2_dyak_ca
 fl1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/605/GCF_000146605.2_Turkey_
 5.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/577/835/GCF_001577835.1_Coturni
 x_japonica_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/545/GCF_000002545.3_ASM254v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/525/GCF_000002525.2_ASM252v
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/515/GCF_000002515.2_ASM251v
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/445/GCF_000006445.2_ASM644v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/858/045/GCF_001858045.1_ASM1858
 04v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/660/625/GCF_001660625.1_IpCoco_
 1.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/163/495/GCF_002163495.1_Omyk_1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/754/195/GCF_000754195.2_ASM7541
 9v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/515/GCF_000001515.7_Pan_tro
 _3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/686/985/GCF_000686985.2_Bra_nap
 us_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/026/365/GCF_000026365.1_ASM2636
 v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/142/805/GCF_000142805.1_ASM1428
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/335/GCF_000006335.2_ASM633v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/335/GCF_000002335.3_Tcas5.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/765/GCF_000001765.3_Dpse_3.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/295/GCF_000002295.2_MonDom5'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/772/875/GCF_000772875.2_Mmul_8.
 0.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/755/GCF_000149755.1_P.sojae
 _V3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/905/GCF_000001905.1_Loxafr3
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/511/025/GCF_000511025.2_RefBeet
 -1.2.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/189/635/GCF_000189635.1_JCVI-
 TTA1-2.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/309/985/GCF_000309985.1_Brapa_1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/209/225/GCF_000209225.1_ASM2092
 2v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/735/GCF_000151735.1_Cavpor3
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/313/985/GCF_000313985.1_EchTel2
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/208/655/GCF_000208655.1_Dasnov3
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/237/925/GCF_000237925.1_ASM2379
 2v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/181/795/GCF_000181795.1_Trichin
 ella_spiralis-3.7.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/181/275/GCF_000181275.1_SorAra2
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/296/755/GCF_000296755.1_EriEur2
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/585/GCF_000149585.1_ASM1495
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/601/855/GCF_001601855.1_ASM1601
 85v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/208/865/GCF_000208865.1_ASM2088
 6v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/875/GCF_000002875.2_ASM287v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/135/GCF_000005135.1_dere_ca
 f1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/925/GCF_000005925.1_dwil_ca
 f1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/515/GCF_000149515.1_ASM14951v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/115/GCF_000005115.1_dana_caf1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/155/GCF_000005155.2_dgri_caf1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/825/GCF_000002825.2_ASM282v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/175/GCF_000005175.2_dmoj_caf1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/245/GCF_000005245.1_dvir_caf1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/985/GCF_000004985.1_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/195/GCF_000005195.2_dper_caf1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/215/GCF_000005215.3_dsec_caf1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/165/425/GCF_000165425.1_ASM16542v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/425/GCF_000149425.1_ASM14942v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/313/135/GCF_000313135.1_Acastellanii.strNEFF_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/405/GCF_000006405.1_JCVI_PMG_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/835/GCF_000003835.1_ASM383v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/142/945/GCF_000142945.1_ASM14294v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/665/GCF_000151665.1_ASM15166v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/143/365/GCF_000143365.1_ASM14336v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/625/GCF_000003625.3_OryCun2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/335/GCF_000149335.2_ASM14933v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/685/GCF_000149685.1_ASM14968v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/167/675/GCF_000167675.1_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/515/GCF_000003515.1_ASM351v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/545/GCF_000001545.4_P_pygmaeus_2.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/220/395/GCF_000220395.1_JCVI-IMG1-V.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/735/GCF_000150735.1_Paracocci_br_Pb18_V2'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/209/165/GCF_000209165.1_ASM20916v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/925/GCF_000149925.1_ASM14992v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/182/565/GCF_000182565.1_S_punctatus_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/143/185/GCF_000143185.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/127/325/GCF_002127325.1_HanXRQr1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/870/075/GCF_002870075.1_Lsat_Salinas_v7'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/275/GCF_000150275.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/457/365/GCF_000457365.1_ASM45736v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/148/765/GCF_000148765.1_MalDomGD1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/605/GCF_000003605.2_Skow_1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/275/GCF_000006275.2_JCVI-afl1-v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/805/GCF_000151805.1_Taeniopygia_guttata-3.2.4'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/233/375/GCF_000233375.1_ICSASG_v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/209/125/GCF_000209125.1_JCVI_EDISG_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/214/015/GCF_000214015.3_version_140606'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/330/505/GCF_000330505.1_EIA2_v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/499/845/GCF_000499845.1_PhaVulg1_0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/425/GCF_000002425.3_V1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/346/465/GCF_000346465.2_Prunus_persica_NCBIV2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/209/185/GCF_000209185.1_CulPip1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/264/685/GCF_000264685.3_Panu_3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/022/765/GCF_002022765.2_C_virginica-3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/745/GCF_000003745.3_12X'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/226/075/GCF_000226075.1_SolTub_3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/092/065/GCF_000092065.1_ASM9206v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/143/415/GCF_000143415.3_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/341/285/GCF_000341285.1_ASM34128v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/143/455/GCF_000143455.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/955/GCF_000150955.2_ASM15095v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/026/945/GCF_000026945.1_ASM2694v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/715/135/GCF_000715135.1_Ntab-TN90'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/186/865/GCF_000186865.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/659/605/GCF_001659605.1_Manihot_esculenta_v6'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/075/GCF_000002075.1_AplCal3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/505/GCF_000150505.1_S06'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/845/GCF_000149845.2_SJ5'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/203/795/GCF_000203795.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/142/985/GCF_000142985.2_Acyr2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/325/GCF_000002325.3_Nvit2.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/181/295/GCF_000181295.1_OtoGar3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/219/625/GCF_000219625.1_MYCGR_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/665/GCF_000004665.1_Callithrix_jacchus-3.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/915/GCF_000146915.1_ASM14691v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/236/235/GCF_000236235.1_SpeTri2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/795/GCF_000146795.2_Nleu3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/945/GCF_000146945.2_ASM14694v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/555/GCF_000149555.1_ASM14955v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/255/GCF_000004255.2_v.1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/143/535/GCF_000143535.1_BotFuc-Aug2005'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/504/015/GCF_000504015.1_Mimgu1_0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/375/325/GCF_000375325.1_Caprub1_0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/715/GCF_000002715.2_ASM271v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/645/GCF_000149645.1_ASM14964v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/265/GCF_000151265.2_Micromonas_pusilla_CCMP1545_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/654/055/GCF_001654055.1_ASM1654

05v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/535/GCF_000150535.2_Papaya1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/295/GCF_000006295.1_JCVI_LO
 USE_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/208/615/GCF_000208615.1_JCVI_IS
 G_i3_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/184/455/GCF_000184455.2_ASM1844
 5v3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/355/GCF_000151355.1_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/234/675/GCF_002234675.1_ASM2234
 67v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/208/745/GCF_000208745.1_Criollo
 _cocoa_genome_V2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/147/115/GCF_000147115.1_Myoluc2
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/685/GCF_000151685.1_JCVI_RC
 G_1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/143/565/GCF_000143565.1_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/638/985/GCF_001638985.1_Phybl2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/741/045/GCF_000741045.1_Vradiat
 a_ver6'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/854/935/GCF_001854935.1_ASM1854
 93v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/165/045/GCF_000165045.1_Callorh
 inchus_milii-6.1.3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/147/415/GCF_000147415.1_v_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/505/GCF_000005505.2_Brachyp
 odium_distachyon_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/181/695/GCF_000181695.1_ASM1816
 9v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/985/GCF_000149985.1_ASM1499
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/955/GCF_000149955.1_ASM1499
 5v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/090/745/GCF_000090745.1_AnoCar2
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/865/GCF_000002865.3_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/845/GCF_000002845.2_ASM284v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/035/GCF_000150035.1_ASM1500
 3v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/125/GCF_000003125.1_JCVI-
 TSTA1-3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/985/GCF_000001985.1_JCVI-
 PMFA1-2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/255/GCF_000006255.1_Postia_
 placenta_V1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/170/995/GCF_000170995.1_TRIVI_v
 2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/826/765/GCF_000826765.1_Fundulu
 s_heteroclitus-3.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/171/015/GCF_000171015.1_TRIAT_v
 2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/845/GCF_000151845.1_Pvam_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/315/GCF_000151315.2_C_owcza
 rzaki_V2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/922/835/GCF_001922835.1_NIST_Tu
 r_tru_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/507/365/GCF_000507365.1_N_ameri
 canus_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/292/845/GCF_000292845.1_OchPri3
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/885/GCF_000151885.1_Dord_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/364/345/GCF_000364345.1_Macaca_
 fascicularis_5.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/165/445/GCF_000165445.2_Mmur_3.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/517/525/GCF_000517525.1_Limulus
 _polyphemus-2.1.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/485/595/GCF_000485595.1_Priapul
 us_caudatus-5.0.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/471/725/GCF_000471725.1_UMD_CAS
 PUR_WB_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/696/425/GCF_000696425.1_G_varie
 gatus-3.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/145/GCF_000151145.1_ASM1511
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/425/GCF_000151425.1_ASM1514
 2v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/975/GCF_000150975.2_MS_CBS1
 18893'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/209/485/GCF_000209485.1_ASM2094
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/675/GCF_000150675.1_ASM1506
 7v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/515/GCF_000006515.1_JCVI_cm
 g_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/183/025/GCF_000183025.1_ASM1830
 2v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/625/215/GCF_001625215.1_ASM1625
 21v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/300/555/GCF_000300555.1_Agabi_v
 arbur_1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/475/195/GCF_000475195.1_Diaci_p
 syllid_genome_assembly_version_1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/125/GCF_000151125.1_ASM1511
 2v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/164/845/GCF_000164845.2_Vicugna
 _pacos-2.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/280/675/GCF_000280675.1_ASM2806
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/329/695/GCF_001329695.1_Rhoba1_
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/218/685/GCF_000218685.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/696/525/GCF_000696525.1_JatCur_
 1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/182/985/GCF_000182985.1_ASM1829
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/443/255/GCF_002443255.1_Vdes_3.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/223/465/GCF_000223465.1_Candida
 _tenuis_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/075/GCF_000004075.2_ASM407v
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/035/GCF_000149035.1_C_grami
 nicola_M1_001_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/905/GCF_000151905.2_gorGor4'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/612/305/GCF_000612305.1_Egrandi
 s1_0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/645/GCF_000151645.1_ASM1516
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/300/595/GCF_000300595.1_Phanero
 chaete_carnosa_HHB-10118-Sp_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/182/805/GCF_000182805.1_ASM1828
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/184/105/GCF_000184105.1_Geom_de
 structans_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/155/GCF_000004155.1_SCY4'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/505/GCF_000151505.1_ASM1515
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/143/105/GCF_000143105.1_Sanger-
 454-IlluminaPA_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/187/405/GCF_000187405.1_MetAcr_
 May2010'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/465/GCF_000146465.1_ASM1464
 6v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/203/815/GCF_000203815.1_DFas_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/354/255/GCF_000354255.1_CocheC4
 _1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/238/955/GCF_000238955.2_M_zebra

_UMD1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/465/895/GCF_001465895.1_Nfu_20140520'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/413/155/GCF_000413155.1_DPV01'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/442/705/GCF_000442705.1_EG5'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/258/705/GCF_000258705.1_Coccomyxa_subellipsoidea_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/090/795/GCF_000090795.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/335/GCF_000004335.2_AilMel_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/183/805/GCF_000183805.2_Loa_loa_V3.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/239/435/GCF_000239435.1_ASM23943v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/545/GCF_000151545.1_ASM15154v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/187/915/GCF_000187915.1_Pbar_UMD_V03'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/217/595/GCF_000217595.1_Lhum_UMD_V04'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/347/755/GCF_000347755.3_Ccap_2.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/214/255/GCF_000214255.1_Bter_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/184/785/GCF_000184785.2_Aflo_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/856/785/GCF_001856785.1_MPER_GO061.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/223/135/GCF_000223135.1_CriGri_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/355/885/GCF_000355885.1_BGI_duck_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/204/055/GCF_000204055.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/221/225/GCF_000221225.1_CTHT_3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/143/395/GCF_000143395.1_Attacep_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/340/665/GCF_000340665.1_C.cajan_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/188/075/GCF_000188075.1_Si_gnG'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/757/685/GCF_000757685.1_NilLug1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/230/625/GCF_000230625.1_Exop_derm_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/147/195/GCF_000147195.1_HarSal_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/147/175/GCF_000147175.1_CamFlo_1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/331/145/GCF_000331145.1_ASM33114v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/142/905/GCF_000142905.1_TheTra_May2010'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/320/565/GCF_000320565.1_Septoria_musiva_S02202_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/250/985/GCF_000250985.1_Nema_parisii_ERTm1_V3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/189/315/GCF_000189315.1_Devil_ref_v7.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/661/255/GCF_001661255.1_Wican1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/264/995/GCF_000264995.1_Punctularia_strigosoazonata_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/264/905/GCF_000264905.1_Stehi1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/708/205/GCF_000708205.1_Mixia_omundae_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/223/485/GCF_000223485.1_Spathaspora_passalidarum_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/338/995/GCF_000338995.1_Cocsa1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/327/365/GCF_000327365.1_Graimonddi2_0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/271/625/GCF_000271625.1_Conpu1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/271/585/GCF_000271585.1_Trametes_versicolor_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/271/605/GCF_000271605.1_Fomme1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/225/785/GCF_000225785.1_LatCha1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/190/715/GCF_000190715.1_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/223/845/GCF_000223845.1_GNI3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/215/625/GCF_000215625.1_MusPutFur1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/184/155/GCF_000184155.1_FraVesHawaii_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/239/415/GCF_000239415.1_AstBur1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/239/395/GCF_000239395.1_NeoBri1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/239/375/GCF_000239375.1_PunNye1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/188/095/GCF_000188095.1_BIMP_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/247/795/GCF_000247795.1_Bos_indicus_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/192/795/GCF_000192795.1_Vavr_culifloridensis_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/255/335/GCF_000255335.1_Mocc_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/236/285/GCF_000236285.1_Dbip_2.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/224/195/GCF_000224195.1_Dele_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/220/665/GCF_000220665.1_Dfic_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/224/215/GCF_000224215.1_Dkik_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/224/235/GCF_000224235.1_Dtak_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/233/415/GCF_000233415.1_Dbia_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/227/135/GCF_000227135.1_ASM2271
 3v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/204/515/GCF_000204515.1_Aech_3.
 9'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/102/435/GCF_002102435.1_Ovir.te
 _1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/344/685/GCF_000344685.1_Glotr1_
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/213/175/GCF_000213175.1_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/220/905/GCF_000220905.1_MROT_1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/236/305/GCF_000236305.1_Drho_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/236/325/GCF_000236325.1_Deug_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/235/385/GCF_000235385.1_SaiBol1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/247/695/GCF_000247695.1_HetGla_
 female_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/305/755/GCF_001305755.1_ASM1305
 75v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/007/265/GCF_002007265.1_ASM2007
 26v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/067/755/GCF_900067755.1_pvi1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/690/875/GCF_000690875.1_ASM6908
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/708/225/GCF_000708225.1_ASM7082
 2v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/769/185/GCF_000769185.1_Rrox_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/201/575/GCF_002201575.1_ASM2201
 57v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/776/465/GCF_002776465.1_AmpOce1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/288/905/GCF_002288905.1_ASM2288
 90v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/624/265/GCF_001624265.1_ASM1624
 26v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/466/805/GCF_001466805.2_Raegyp2

.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/754/665/GCF_000754665.1_Bison_U
MD1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/956/335/GCF_000956335.1_Plas_fr
ag_nilgiri_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/515/645/GCF_001515645.1_SAMN033
20097.WGS_v1.1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/984/765/GCF_001984765.1_C.can_g
enome_v1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/698/545/GCF_001698545.1_ASM1698
54v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/857/705/GCF_001857705.1_PanPar1
.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/949/145/GCF_001949145.1_OKI-
Apl_1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/682/695/GCF_001682695.1_Pygocen
trus_nattereri-1.0.2'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/242/695/GCF_000242695.1_LepOcu1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/276/665/GCF_000276665.1_ChiLan1
.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/349/705/GCF_000349705.1_LepWed1
.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/856/805/GCF_001856805.1_P_rapae
_3842_assembly_v2'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/222/465/GCF_000222465.1_Adig_1.
1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/225/545/GCF_000225545.1_AOL2492
7_1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/225/605/GCF_000225605.1_Cmilita
risCM01_v01'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/226/095/GCF_000226095.1_ASM2260
9v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/226/115/GCF_000226115.1_ASM2261
1v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/313/045/GCF_000313045.1_ASM3130
4v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/317/415/GCF_000317415.1_Csi_val
encia_1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/987/745/GCF_000987745.1_ASM9877
4v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/445/GCF_000699445.1_SchHae_
1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/742/605/GCF_002742605.1_O_europ
aea_v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/258/655/GCF_000258655.2_panpan1
.1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF_001704415.1_ARS1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/767/585/GCF_000767585.1_PRJNA23

4474_Ca_dromedarius_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/767/855/GCF_000767855.1_Ca_bactrianus_MBC_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/735/185/GCF_000735185.1_NC01'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/789/215/GCF_000789215.1_ASM78921v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/297/895/GCF_000297895.1_oyster_v9'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/775/205/GCF_002775205.1_X_maculatus-5.0-male'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/238/935/GCF_000238935.1_Melopsittacus_undulatus_6.3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/280/705/GCF_000280705.1_JacJac1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/493/695/GCF_000493695.1_BalAcu1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/400/835/GCF_000400835.1_PHO1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/464/555/GCF_000464555.1_PanTig1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/488/995/GCF_000488995.1_M_roreri_MCA_2997_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/226/545/GCF_000226545.1_ASM22654v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/951/615/GCF_000951615.1_common_carp_genome'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/970/005/GCF_001970005.1_Flounder_ref_guided_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/317/375/GCF_000317375.1_MicOch1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/260/355/GCF_000260355.1_ConCri1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/231/095/GCF_000231095.1_Oryza_brachyantha.v1.4b'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/661/265/GCF_001661265.1_Saico1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/710/875/GCF_000710875.1_Pepper_Zunla_1_Ref_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/417/885/GCF_001417885.1_Kmar_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/695/525/GCF_000695525.1_BOL'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/039/765/GCF_001039765.1_AptMant0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/269/505/GCF_000269505.1_DroMir_2.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/780/465/GCF_002780465.1_ASM278046v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/612/285/GCF_000612285.1_Gossypium_arboreum_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/239/965/GCF_900239965.1_Bicyclu

s_anynana_v1.2'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/957/025/GCF_001957025.1_Aet_MR_1.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/313/855/GCF_000313855.2_ASM31385v2'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/876/935/GCF_001876935.1_Aspof.V1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/263/155/GCF_000263155.2_Setaria_italica_v2.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/186/125/GCF_001186125.1_Spha_arctica_JP610_V1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/349/005/GCF_000349005.2_Pneumo_murina_B123_V4'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/477/535/GCF_001477535.1_Pneu_jiro_RU7_V2'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/271/645/GCF_000271645.1_Tremel'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/865/875/GCF_001865875.1_LupAngTanjil_v1.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/190/045/GCF_001190045.1_Vigan1.1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/342/415/GCF_000342415.1_JCVI-bewag-v1.1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/642/055/GCF_001642055.1_Altal1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/742/065/GCF_002742065.1_CB0940_V2'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/355/655/GCF_000355655.1_DendPond_male_1.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/534/875/GCF_000534875.1_SCA1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/015/335/GCF_001015335.1_Stomoxys_calcitrans-1.0.1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/648/675/GCF_000648675.1_Clec_1.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/672/515/GCF_001672515.1_ASM167251v1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/493/195/GCF_000493195.1_Citrus_clementina_v1.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/769/745/GCF_000769745.1_ASM76974v1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/500/345/GCF_000500345.1_Pman_1.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/263/595/GCF_001263595.1_ASM126359v1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/277/815/GCF_000277815.2_ASM27781v3'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/235/365/GCF_000235365.1_ASM23536v1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/230/375/GCF_000230375.1_ASM23037v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/512/975/GCF_000512975.1_S_indic
 um_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/330/985/GCF_000330985.1_DBM_FJ_
 V1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/653/235/GCF_001653235.2_ASM1653
 23v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/240/725/GCF_000240725.1_ASM2407
 2v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/257/125/GCF_000257125.1_ENU1_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/691/945/GCF_000691945.2_ASM6919
 4v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/308/155/GCF_000308155.1_EptFus1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/409/485/GCF_000409485.1_GLAREA'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/636/725/GCF_001636725.1_ISF_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/247/585/GCF_000247585.1_PP_INRA
 -310_V2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/207/925/GCF_002207925.1_PAV_r1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/523/025/GCF_000523025.1_Cse_v1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/806/345/GCF_000806345.1_ASM8063
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/298/275/GCF_000298275.1_OryAfe1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/296/735/GCF_000296735.1_ChraSi1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/299/155/GCF_000299155.1_EleEdw1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/260/255/GCF_000260255.1_OctDeg1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/247/815/GCF_000247815.1_FicAlb1
 .5'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/311/805/GCF_000311805.1_CB1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/788/275/GCF_000788275.1_BP.fa'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/349/665/GCF_000349665.1_MesAur1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/471/905/GCF_000471905.2_AMTR1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/571/385/GCF_002571385.1_Styloph
 ora_pistillata_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/442/555/GCF_001442555.1_ACSNU-
 2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/817/695/GCF_000817695.2_Aradu1.
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/325/575/GCF_000325575.1_ASM3255
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/227/115/GCF_000227115.2_ASM2271
 1v2'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/350/225/GCF_000350225.1_ASM3502
 2v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/241/765/GCF_000241765.3_Chrysem
 ys_picta_bellii-3.0.3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/910/315/GCF_002910315.2_ASM2910
 31v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/236/905/GCF_000236905.1_ASM2369
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/625/305/GCF_001625305.1_Haploid
 v18h27'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/113/885/GCF_002113885.1_ASM2113
 88v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/972/845/GCF_000972845.1_L_croce
 a_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/806/865/GCF_002806865.1_ASM2806
 86v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/390/325/GCF_000390325.2_Ntom_v0
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/243/375/GCF_000243375.1_ASM2433
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/478/725/GCF_000478725.1_Eutsalg
 1_0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/258/005/GCF_000258005.1_HHA1_v0
 2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/315/915/GCF_000315915.1_ASM3159
 1v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/304/475/GCF_000304475.1_Ka_CBS2
 517'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/348/985/GCF_000348985.1_ASM3489
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/220/235/GCF_002220235.1_ASM2220
 23v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/313/525/GCF_000313525.1_ASM3135
 2v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/315/875/GCF_000315875.1_ASM3158
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/359/705/GCF_000359705.1_Setospa
 eria_trucica_Et28A_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/890/905/GCF_001890905.1_Aspac1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/814/215/GCF_002814215.1_Sedor1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/260/705/GCF_002260705.1_Sdu_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/065/GCF_000699065.1_Lcup_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/648/655/GCF_000648655.2_Cflo_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/622/305/GCF_000622305.1_S.galil
 i_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/683/475/GCF_001683475.1_ASM1683

47v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/572/035/GCF_000572035.2_Mdem2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/499/605/GCF_000499605.1_EMW001'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/315/295/GCF_000315295.1_Pbr_v1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/648/695/GCF_000648695.1_Otau_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/500/325/GCF_000500325.1_Ldec_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/045/GCF_000699045.1_Apla_1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/095/GCF_000004095.1_Hydra_R
 P_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/995/035/GCF_001995035.1_ASM1995
 03v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/522/545/GCF_001522545.2_Parus_m
 ajor1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/499/425/GCF_000499425.1_EAH001'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/801/105/GCF_000801105.1_Rs1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/388/065/GCF_000388065.1_Font_al
 ba_ATCC_38817_V2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/334/495/GCF_000334495.1_TupChi_
 1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/472/045/GCF_000472045.1_Physete
 r_macrocephalus-2.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/485/575/GCF_000485575.1_Poecili
 a_formosa-5.1.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/372/685/GCF_000372685.2_Astyana
 x_mexicanus-2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/690/725/GCF_000690725.1_Stegast
 es_partitus-1.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/732/505/GCF_000732505.1_C_varie
 gatus-1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/292/725/GCF_000292725.1_SacArb1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/021/735/GCF_002021735.1_Okis_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/393/655/GCF_000393655.1_Nsyl'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/409/795/GCF_000409795.2_Chloroc
 ebus_sabeus_1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/266/775/GCF_001266775.1_Austrof
 undulus_limnaeus-1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/288/925/GCF_002288925.1_ASM2288
 92v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/042/975/GCF_002042975.1_ofav_do
 v_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/217/175/GCF_002217175.1_ASM2217
 17v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/079/805/GCF_900079805.1_Fusariu

m_fujikuroi_IMI58289_V2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/879/085/GCF_001879085.1_NIATTr2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/495/115/GCF_000495115.1_PopEup_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/956/065/GCF_000956065.1_Mnem_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/365/465/GCF_000365465.2_Ptep_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/320/785/GCF_000320785.1_RHOziaD V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/277/835/GCF_000277835.1_GeoFor_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/955/945/GCF_000955945.1_Caty_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/344/595/GCF_000344595.1_CheMyd_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/156/985/GCF_002156985.1_Harm_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/738/345/GCF_002738345.1_Cmax_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/281/105/GCF_000281105.1_Coni_ap ol_CBS100218_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/187/425/GCF_000187425.2_MAA_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/315/645/GCF_000315645.1_PdigPd1_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/281/125/GCF_000281125.3_ASM2811 2v4'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/523/455/GCF_000523455.1_Cochlio bolus_miyabeanus_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/523/435/GCF_000523435.1_Cochlio bolus_carbonum_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/527/765/GCF_000527765.1_Cochlio bolus_victoriae_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/280/035/GCF_000280035.1_ASM2800 3v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/723/895/GCF_001723895.1_CroPor_ compl'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/709/005/GCF_000709005.1_Plas_vinc vinckei_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/743/755/GCF_000743755.1_ASM7437 5v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/585/585/GCF_000585585.1_Capr_co ro_CBS_617_96_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/585/565/GCF_000585565.1_Capr_ep im_CBS_606_96_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/585/515/GCF_000585515.1_Clad_ye gr_CBS_114405_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/709/125/GCF_000709125.1_Exop_aq

ua_CBS_119918_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/346/735/GCF_000346735.1_P.mume_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/836/235/GCF_000836235.1_Pxut_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/540/865/GCF_001540865.1_ASM154086v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/755/165/GCF_000755165.1_ASM75516v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/661/405/GCF_001661405.1_Cybja1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/344/095/GCF_000344095.1_Aros_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/320/585/GCF_000320585.1_Heterobasidion_irregulare_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/321/225/GCF_000321225.1_Oros_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/390/285/GCF_000390285.2_Agla_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/331/955/GCF_000331955.2_Oorc_1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/275/845/GCF_000275845.1_Dichomitus_squalens_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/321/355/GCF_000321355.1_PcynB_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/633/955/GCF_000633955.1-Cs'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/078/875/GCF_002078875.1_NumMel1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/365/185/GCF_000365185.1_Chinese_Lotus_1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/696/795/GCF_000696795.1_Hhal_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/337/975/GCF_000337975.1_F_cherrug_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/671/375/GCF_000671375.1_Cexi_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/599/845/GCF_000599845.2_Tpre_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/341/935/GCF_000341935.1_Ccin1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/281/045/GCF_000281045.1_Sap_diclina_VS20_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/298/355/GCF_001298355.1_Pap_ma_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/640/805/GCF_001640805.1_ASM164080v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/706/865/GCF_002706865.1_ASM270686v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/313/485/GCF_000313485.1_ASM31348v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/303/195/GCF_000303195.2_FP7'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/664/045/GCF_001664045.1_Hanpo2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/187/245/GCF_000187245.1_Hansenu
 la_2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/952/055/GCF_000952055.2_Anan_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/371/365/GCF_000371365.1_Musca_d
 omestica-2.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/234/665/GCF_000234665.1_ASM2346
 6v4'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/293/215/GCF_000293215.1_Trichos
 poron_asahii_1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/661/235/GCF_001661235.1_Picme2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/231/115/GCF_000231115.1_Vitt_co
 rn_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/740/895/GCF_000740895.1_ASM7408
 9v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/983/325/GCF_001983325.1_ASM1983
 32v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/230/535/GCF_000230535.1_PelSin_
 1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/981/445/GCF_000981445.1_Bbig001'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/612/105/GCF_000612105.2_Oabi_2.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/443/585/GCF_001443585.1_aciJub1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/327/345/GCF_000327345.1_ASM3273
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/443/325/GCF_001443325.1_P_mexic
 ana-1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/723/915/GCF_001723915.1_GavGan_
 compl'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/298/775/GCF_000298775.1_ASM2987
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/080/235/GCF_900080235.1_BallGen_
 V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/365/145/GCF_000365145.1_Phia_eu
 ro_CBS_101466_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/442/215/GCF_000442215.1_Lipotes
 _vexillifer_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/516/985/GCF_000516985.1_PFICI'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/890/085/GCF_001890085.1_ASM1890
 08v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/469/605/GCF_000469605.1_Apis_do
 rsata_1.3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/326/865/GCF_000326865.1_Helobde
 lla_robusta_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/327/385/GCF_000327385.1_Helro1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/331/425/GCF_000331425.1_PseHum1

.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/524/495/GCF_000524495.1_Plas_in
ui_San_Antonio_1_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/705/GCF_000150705.2_Paracoc
ci_br_Pb01_V2'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/966/335/GCF_000966335.1_ASM9663
3v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/826/755/GCF_000826755.1_ZizJuj_
1.1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/708/625/GCF_002708625.1_Rhimi1_
1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/814/965/GCF_000814965.1_MBR_1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/365/165/GCF_000365165.1_Clad_ca
rr_CBS_160_54_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/747/765/GCF_000747765.1_ASM7477
6v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/585/535/GCF_000585535.1_Clad_ps
am_CBS_110553_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/520/075/GCF_000520075.1_Apha_as
ta_AP03_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/520/115/GCF_000520115.1_Apha_in
va_NJM9701_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/313/835/GCF_001313835.1_ASM1313
83v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/764/305/GCF_000764305.1_Hazt_2.
0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/891/065/GCF_001891065.1_H_comes
_QL1_v1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/349/385/GCF_000349385.1_UCREL1V
03'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/385/595/GCF_000385595.1_UCRNP2V
3'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/392/275/GCF_000392275.1_UCRPA7V
03'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/077/635/GCF_001077635.1_Thamnop
his_sirtalis-6.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/743/615/GCF_000743615.1_DMR_v1.
0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/507/465/GCF_000507465.1_Kwon_ma
ng_CBS8507_V2'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/512/605/GCF_000512605.1_Cryp_pi
nu_CBS10737_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/512/585/GCF_000512585.1_Cryp_be
st_CBS10118_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/512/565/GCF_000512565.1_Cryp_de
je_CBS10117_V1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/720/205/GCF_001720205.1_Cryp_am
yl_CBS6039_V3'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/349/305/GCF_000349305.1_ASM34930v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/890/805/GCF_001890805.1_Aspg11'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/687/225/GCF_000687225.1_UrsMar_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/093/755/GCF_002093755.1_Dser1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/709/895/GCF_000709895.1_ASM70989v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/188/975/GCF_001188975.1_gapfilled_joined_lt9474.gt500.covgt10'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/400/465/GCF_000400465.1_Wallemia_ichthyophaga_version_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/711/695/GCF_000711695.1_Tilletiaria_anomala_UBC_951_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/582/985/GCF_000582985.1_CFI001'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/441/935/GCF_000441935.1_MHA_v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/443/285/GCF_001443285.1_P_latipinna-1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/385/455/GCF_000385455.1_Zonotrichia_albicollis-1.0.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/105/GCF_000699105.1_ASM69910v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/720/155/GCF_001720155.1_Tsuc_wing_CBS7118_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/463/585/GCF_000463585.1_ASM46358v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/619/985/GCF_001619985.1_Xylona_heveae_TC161_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/411/555/GCF_001411555.1_wgs.5d'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/414/095/GCF_000414095.1_ASM41409v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/591/075/GCF_000591075.1_Eaff_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/319/635/GCF_000319635.1_Colletotrichum_gloeosporioides_Nara_gc-5'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/696/155/GCF_000696155.1_ZooNev1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/403/515/GCF_000403515.1_ASM40351v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/186/305/GCF_000186305.1_Python_molurus_bivittatus-5.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/499/385/GCF_000499385.1_ENH001'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/499/745/GCF_000499745.1_EMH001'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/527/695/GCF_001527695.2_P.Mucros_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/890/105/GCF_001890105.1_Aspzo1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/661/395/GCF_001661395.1_Hypbu1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/661/335/GCF_001661335.1_Babin1'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/661/415/GCF_001661415.1_Canta1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/661/345/GCF_001661345.1_Ascru1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/738/735/GCF_000738735.2_ASM73873v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/664/035/GCF_001664035.1_Metbi1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/412/655/GCF_000412655.1_ASM41265v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/472/105/GCF_000472105.1_Dsuzuki.i.v01'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/738/365/GCF_002738365.1_Cmos_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/150/825/GCF_000150825.1_ASM15082v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/455/745/GCF_000455745.1_ASM45574v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/961/545/GCF_000961545.1_S_schenckii_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/721/915/GCF_000721915.3_Eluc_V3'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/027/005/GCF_000027005.1_ASM2700v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/298/355/GCF_000298355.1_BosGru_v2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/649/575/GCF_001649575.1_ASM164957v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/503/995/GCF_000503995.1_CerSol_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/633/615/GCF_000633615.1_Guppy_female_1.0_MT'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/099/425/GCF_002099425.1_phaCin_unsw_v4.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/204/375/GCF_002204375.1_MunDraft-v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/952/655/GCF_001952655.1_M_albus_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/186/385/GCF_001186385.1_Dnoxia_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/406/875/GCF_001406875.1_SPENNV200'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/642/345/GCF_001642345.1_ASM164234v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/464/535/GCF_000464535.1_EPUS'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/956/105/GCF_000956105.1_Pcoq_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/188/695/GCF_000188695.1_Proterospongia_sp_ATCC50818'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/417/875/GCF_000417875.1_Pflocc_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/243/295/GCF_000243295.1_TriManL

at1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/283/155/GCF_000283155.1_CerSimS
 im1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/576/695/GCF_000576695.1_AUH_PRJ
 EB4427_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/602/025/GCF_001602025.1_ASM1602
 02v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/733/215/GCF_000733215.1_ASM7332
 1v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/497/045/GCF_000497045.1_PSEUBRA
 1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/971/095/GCF_000971095.1_AnsCyg_
 PRJNA183603_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/687/185/GCF_000687185.1_ASM6871
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/696/695/GCF_000696695.1_Burgud'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/605/985/GCF_001605985.1_ASM1605
 98v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/687/205/GCF_000687205.1_ASM6872
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/687/285/GCF_000687285.1_ASM6872
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/687/375/GCF_000687375.1_ASM6873
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/690/535/GCF_000690535.1_ASM6905
 3v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/690/715/GCF_000690715.1_ASM6907
 1v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/690/775/GCF_000690775.1_ASM6907
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/690/835/GCF_000690835.1_ASM6908
 3v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/691/245/GCF_000691245.1_Tgr_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/691/405/GCF_000691405.1_ASM6914
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/691/845/GCF_000691845.1_ASM6918
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/715/985/GCF_001715985.1_ASM1715
 98v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/692/075/GCF_000692075.1_ASM6920
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/691/785/GCF_000691785.1_ASM6917
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/695/815/GCF_000695815.1_ASM6958
 1v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/766/835/GCF_000766835.1_Aquila_
 chrysaetos-1.0.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/695/765/GCF_000695765.1_ASM6957

6v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/691/975/GCF_000691975.1_ASM6919
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/005/GCF_000699005.1_ASM6990
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/085/GCF_000699085.1_ASM6990
 8v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/245/GCF_000699245.1_ASM6992
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/700/745/GCF_000700745.1_ASM7007
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/145/GCF_000699145.1_ASM6991
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/708/025/GCF_000708025.1_ASM7080
 2v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/703/405/GCF_000703405.1_ASM7034
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/709/325/GCF_000709325.1_ASM7093
 2v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/708/925/GCF_000708925.1_ASM7089
 2v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/709/365/GCF_000709365.1_ASM7093
 6v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/705/375/GCF_000705375.1_ASM7053
 7v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/710/305/GCF_000710305.1_ASM7103
 0v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/721/765/GCF_000721765.1_Aureoba
 sidium_pullulans_var._namibiae_CBS_147.97_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/721/755/GCF_000721755.1_Aureoba
 sidium_pullulans_var._subglaciale_EXF-2481_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/715/545/GCF_000715545.1_OpiViv1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/680/005/GCF_001680005.1_ASM1680
 00v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/732/125/GCF_000732125.1_ScApio1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/737/465/GCF_000737465.1_Haliaee
 tus_leucocephalus-4.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/226/395/GCF_000226395.1_PenChr_
 Nov2007'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/696/875/GCF_000696875.1_ASM6968
 7v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/747/805/GCF_000747805.1_ChaPel_
 1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/836/215/GCF_000836215.1_Ppol_1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/760/515/GCF_000760515.1_UGP1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/769/155/GCF_000769155.1_ASM76915v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/094/665/GCF_900094665.1_CAROLIEIJ_v1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/695/195/GCF_000695195.1_ASM69519v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/836/275/GCF_000836275.1_Exop_mesoso_CBS40295_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/835/495/GCF_000835495.1_Cladimu_CBS83496_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/787/575/GCF_000787575.1_Asub_2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/835/505/GCF_000835505.1_Exop_xeno_CBS118157_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/803/265/GCF_000803265.1_ASM80326v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/806/365/GCF_000806365.1_ASM80636v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/816/755/GCF_000816755.2_Araip1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/835/455/GCF_000835455.1_Fons_pedr_CBS_271_37_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/836/115/GCF_000836115.1_Exop_spin_CBS89968_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/836/295/GCF_000836295.1_0_gall_CBS43764'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/835/475/GCF_000835475.1_Clad_bant_CBS_173_52_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/835/515/GCF_000835515.1_Exop_olig_CBS72588_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/835/555/GCF_000835555.1_Rhin_mack_CBS_650_93_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/836/435/GCF_000836435.1_Fons_mult_CBS_102226_V1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/611/645/GCF_000611645.1_mono_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/935/625/GCF_000935625.1_ASM93562v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/938/715/GCF_000938715.1_LALAO'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/458/135/GCF_001458135.1_marMar2.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/949/405/GCF_000949405.1_V.emery_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/951/045/GCF_000951045.1_Mleu.le_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/951/035/GCF_000951035.1_Cang.pa_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/956/235/GCF_000956235.1_wasmannia.A_1.0'

'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/968/595/GCF_000968595.1_ASM96859v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/685/135/GCF_001685135.1_ManJav1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/980/195/GCF_000980195.1_M.pharaonis_V2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/814/945/GCF_000814945.1_MAJ_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/277/935/GCF_001277935.1_ASM127793v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/515/605/GCF_001515605.1_SAMN03320099.WGS_v1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/515/625/GCF_001515625.1_SAMN03320098_v1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/027/345/GCF_001027345.1_Trio11'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/039/355/GCF_001039355.1_LinAna1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/189/475/GCF_001189475.1_ASM118947v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/204/775/GCF_001204775.1_ASM120477v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/186/105/GCF_001186105.1_ASM118610v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/263/575/GCF_001263575.1_Nlec1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/278/385/GCF_001278385.1_MalaPachy'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/272/555/GCF_001272555.1_ASM127255v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/263/375/GCF_000263375.1_Wallemia_sebi_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/263/275/GCF_001263275.1_ASM126327v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/293/395/GCF_001293395.1_ASM129339v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/338/955/GCF_000338955.1_Bauco1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/299/255/GCF_001299255.1_ASM129925v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/855/GCF_000003855.2_ASM385v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/447/785/GCF_001447785.1_Gekko_japonicus_V1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/313/825/GCF_001313825.1_ASM131382v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/412/225/GCF_001412225.1_Nicve_v1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/412/515/GCF_001412515.1_Dall1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/417/965/GCF_001417965.1_Aiptasia_genome_1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/265/015/GCF_000265015.1_Auricul

aria_subglabra_SS-5_V1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/298/625/GCF_001298625.1_SEUB3.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/431/845/GCF_001431845.1_ASM143184v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/194/135/GCF_001194135.1_Octopus_bimaculoides_v2_0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/447/265/GCF_001447265.1_Sturnus_vulgaris-1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/447/935/GCF_001447935.1_debFab1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/465/965/GCF_001465965.1_Pdom_r1.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/653/265/GCF_001653265.1_ASM165326v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/481/775/GCF_001481775.1_ASM148177v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/500/285/GCF_001500285.1_Phisc1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/074/925/GCF_900074925.1_version_1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/197/715/GCF_002197715.1_LonStrDom1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/853/355/GCF_001853355.1_ASM185335v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/595/765/GCF_001595765.1_Mnat.v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/594/045/GCF_001594045.1_Acol1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/594/065/GCF_001594065.1_Ccos11.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/594/075/GCF_001594075.1_Tcor1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/594/115/GCF_001594115.1_Tsep1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/604/755/GCF_001604755.1_Lepidotrix_coronata-1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/594/055/GCF_001594055.1_Tzet1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/604/975/GCF_001604975.1_Cebus_imitator-1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/090/985/GCF_000090985.2_ASM9098v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/642/045/GCF_001642045.1_Parsp1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/642/475/GCF_001642475.1_ASM164247v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/640/025/GCF_001640025.1_ASM164002v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/483/705/GCF_001483705.1_ASM148370v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/646/965/GCF_001646965.1_ASM164696v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/651/985/GCF_001651985.1_ASM165198v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/652/005/GCF_001652005.1_ASM1652

00v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/654/025/GCF_001654025.1_ASM1654
 02v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/654/015/GCF_001654015.1_ASM1654
 01v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/185/945/GCF_000185945.1_ASM1859
 4v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/662/655/GCF_001662655.1_ASM1662
 65v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/164/805/GCF_000164805.1_Tarsius
 _syrichtha-2.0.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/340/215/GCF_000340215.1_Mycfi2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/687/245/GCF_001687245.1_Rhagole
 tis_zephyria_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/792/695/GCF_001792695.1_ASM1792
 69v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/773/325/GCF_001773325.1_ASM1773
 32v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/879/475/GCF_001879475.1_Asagao_
 1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/831/195/GCF_001831195.1_CORC01'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/883/845/GCF_001883845.1_ASM1883
 84v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/888/835/GCF_001888835.1_ASM1888
 83v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/907/595/GCF_001907595.1_ASM1907
 59v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/937/115/GCF_001937115.1_Atum_1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/145/635/GCF_000145635.1_Gae_gra
 minis_V2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/006/095/GCF_002006095.1_ASM2006
 09v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/095/145/GCF_900095145.1_PAHARI_
 EIJ_v1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/102/565/GCF_002102565.1_Kocim1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/105/155/GCF_002105155.1_Lobtra1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/109/545/GCF_002109545.1_ASM2109
 54v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/168/275/GCF_002168275.1_ASM2168
 27v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/611/835/GCF_000611835.1_CerBir1
 .0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/217/835/GCF_002217835.1_Dobs_1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/303/985/GCF_002303985.1_Duzib1.
 0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/532/875/GCF_002532875.1_vjacob_

1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/776/525/GCF_002776525.1_ASM277652v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/891/405/GCF_002891405.1_Csec_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/872/115/GCF_002872115.1_PKINGS_0.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/825/GCF_000004825.1_PolPal_Dec2009'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/906/115/GCF_002906115.1_CorkOak1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/692/015/GCF_000692015.1_ASM69201v2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002/837/175/GCF_002837175.1_ASM283717v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/241/075/GCF_000241075.1_Xiphophorus_maculatus-4.4.2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/337/935/GCF_000337935.1_Cliv_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/230/445/GCF_000230445.1_HetGla_1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/239/315/GCF_000239315.1_MusPutFurMale1.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/224/045/GCF_000224045.1_CSB10A_v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/300/575/GCF_000300575.1_Agabi_varbish97_2'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/865/GCF_000151865.2_Ttru_1.4'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/403/535/GCF_000403535.1_Theobroma_cacao_20110822'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/006/425/GCF_000006425.1_ASM642v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/723/685/GCF_000723685.1_PREICH01'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/255/GCF_000002255.3_Mmul_051212'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/315/895/GCF_000315895.1_ASM31589v1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/188/235/GCF_000188235.2_Orenil1.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/305/GCF_000002305.2_EquCab2.0'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/205/GCF_000003205.7_Btau_5.0.1'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/265/GCF_000002265.2_Rn_Celebra'
 'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/245/GCF_000149245.1_CNA3'

```
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/015/GCF_000004015.4_AaegL3'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/445/GCF_000002445.1_ASM244v
1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/125/GCF_000002125.1_HuRef'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/259/055/GCF_000259055.1_dsima
f1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/419/365/GCF_000419365.1_C_grise
us_v1.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/166/005/GCF_000166005.1_PyrTer_
1.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/742/935/GCF_000742935.1_ASM7429
3v1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/313/675/GCF_000313675.1_ASM3136
7v1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/765/115/GCF_000765115.1_Oori1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/085/GCF_000003085.2_ASM308v
2'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/149/385/GCF_000149385.1_ASM1493
8v1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/075/GCF_000003075.1_ASM307v
1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/425/GCF_000005425.2_Build_4
.0'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/306/695/GCF_000306695.2_CHM1_1.
1'
'ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/395/GCF_000005395.1_ASM539v
1']
```

[106]: # rename and set index

Level

```
[106]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
5	Hordeum vulgare	Eukaryota;Plants;Land Plants
14	Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes
18	Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts
19	Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold

5		NaN	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold
14	FGSC A4		SAMN02953587	PRJNA130	GCA_000149205.2	Scaffold
18	CL Brener		SAMN02953627	PRJNA11755	GCA_000209065.1	Scaffold
19	WB C6		SAMN02952905	PRJNA1439	GCA_000002435.1	Scaffold
...
5126		NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848		SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373		SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245		SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M20N0800_1A		SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%		Replicons	WGS	\
0	167.6760	64.5000		NaN	AHAL01	
5	9788.8600	0.0000		NaN	FJWB02	
14	30.2760	50.2721	mitochondrion MT:NC_017896.1/JQ435097.1		AACD01	
18	89.9375	51.7000		NaN	AAHK01	
19	11.2136	49.2000		NaN	AACB02	
...
5126	12.0392	38.1000		NaN	FXDE01	
5127	11.7086	38.1000		NaN	FXCK01	
5128	11.8194	38.0000		NaN	FXCO01	
5129	11.7081	38.1000		NaN	FXCP01	
5130	11.8189	38.1000		NaN	FXCG01	

	Scaffolds	CDS	Release Date	\
0	7795	38554.0	2013-04-19 00:00:00+00:00	
5	72295	0.0	2016-08-25 00:00:00+00:00	
14	91	9561.0	2003-04-07 00:00:00+00:00	
18	29495	19607.0	2005-07-14 00:00:00+00:00	
19	92	6502.0	2003-03-26 00:00:00+00:00	
...
5126	1925	0.0	2017-04-21 00:00:00+00:00	
5127	590	0.0	2017-04-21 00:00:00+00:00	
5128	2013	0.0	2017-04-21 00:00:00+00:00	
5129	1028	0.0	2017-04-21 00:00:00+00:00	
5130	799	0.0	2017-04-21 00:00:00+00:00	

	GenBank FTP	\
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
5	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
14	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
18	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
19	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	

```

5129 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

                                RefSeq FTP
0      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
5                                NaN
14     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
18     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
19     ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
5126                                NaN
5127                                NaN
5128                                NaN
5129                                NaN
5130                                NaN

```

[3350 rows x 16 columns]

```
[107]: # rename and set index
```

```
Level.reset_index()
```

```
[107]:
```

	index	#Organism Name	Organism Groups \
0	0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	5	Hordeum vulgare	Eukaryota;Plants;Land Plants
2	14	Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes
3	18	Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts
4	19	Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists
...
3345	5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
3346	5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
3347	5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
3348	5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
3349	5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold
2	FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2	Scaffold
3	CL Brener	SAMN02953627	PRJNA11755	GCA_000209065.1	Scaffold
4	WB C6	SAMN02952905	PRJNA1439	GCA_000002435.1	Scaffold
...
3345	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
3346	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
3347	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
3348	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
3349	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%		Replicons	WGS \
0	167.6760	64.5000		NaN	AHAL01
1	9788.8600	0.0000		NaN	FJWB02
2	30.2760	50.2721	mitochondrion MT:NC_017896.1/JQ435097.1		AACD01
3	89.9375	51.7000		NaN	AAHK01
4	11.2136	49.2000		NaN	AACB02
...
3345	12.0392	38.1000		NaN	FXDE01
3346	11.7086	38.1000		NaN	FXCK01
3347	11.8194	38.0000		NaN	FXC001
3348	11.7081	38.1000		NaN	FXCP01
3349	11.8189	38.1000		NaN	FXCG01

	Scaffolds	CDS	Release Date \
0	7795	38554.0	2013-04-19 00:00:00+00:00
1	72295	0.0	2016-08-25 00:00:00+00:00
2	91	9561.0	2003-04-07 00:00:00+00:00
3	29495	19607.0	2005-07-14 00:00:00+00:00
4	92	6502.0	2003-03-26 00:00:00+00:00
...
3345	1925	0.0	2017-04-21 00:00:00+00:00
3346	590	0.0	2017-04-21 00:00:00+00:00
3347	2013	0.0	2017-04-21 00:00:00+00:00
3348	1028	0.0	2017-04-21 00:00:00+00:00
3349	799	0.0	2017-04-21 00:00:00+00:00

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
3345	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
3346	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
3347	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
3348	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
3349	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	NaN
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...

```

3345
3346
3347
3348
3349

```

[3350 rows x 17 columns]

```
[108]: #reset_index and rename and set index
```

```
Level.reset_index(drop=True, inplace=True)
```

```
[109]: Level
```

```
[109]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Hordeum vulgare	Eukaryota;Plants;Land Plants
2	Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes
3	Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts
4	Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists
...
3345	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
3346	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
3347	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
3348	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
3349	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMEA3903302	PRJEB13020	GCA_900075435.2	Scaffold
2	FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2	Scaffold
3	CL Brener	SAMN02953627	PRJNA11755	GCA_000209065.1	Scaffold
4	WB C6	SAMN02952905	PRJNA1439	GCA_000002435.1	Scaffold
...
3345	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
3346	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
3347	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
3348	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
3349	M20N0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%		Replicons	WGS \
0	167.6760	64.5000		NaN	AHAL01
1	9788.8600	0.0000		NaN	FJWB02
2	30.2760	50.2721	mitochondrion MT:NC_017896.1/JQ435097.1	AACD01	
3	89.9375	51.7000		NaN	AAHK01
4	11.2136	49.2000		NaN	AACB02
...

3345	12.0392	38.1000	NaN	FXDE01
3346	11.7086	38.1000	NaN	FXCK01
3347	11.8194	38.0000	NaN	FXC001
3348	11.7081	38.1000	NaN	FXCP01
3349	11.8189	38.1000	NaN	FXCG01

	Scaffolds	CDS	Release Date	\
0	7795	38554.0	2013-04-19 00:00:00+00:00	
1	72295	0.0	2016-08-25 00:00:00+00:00	
2	91	9561.0	2003-04-07 00:00:00+00:00	
3	29495	19607.0	2005-07-14 00:00:00+00:00	
4	92	6502.0	2003-03-26 00:00:00+00:00	
...	
3345	1925	0.0	2017-04-21 00:00:00+00:00	
3346	590	0.0	2017-04-21 00:00:00+00:00	
3347	2013	0.0	2017-04-21 00:00:00+00:00	
3348	1028	0.0	2017-04-21 00:00:00+00:00	
3349	799	0.0	2017-04-21 00:00:00+00:00	

	GenBank FTP	\
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
...	...	
3345	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
3346	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
3347	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
3348	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
3349	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	NaN
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
3345	NaN
3346	NaN
3347	NaN
3348	NaN
3349	NaN

[3350 rows x 16 columns]


```
[110]: Level.set_index("#Organism Name")
```

```
[110]:
```

	Organism Groups	Strain \
#Organism Name		
Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists	CCMP1516
Hordeum vulgare	Eukaryota;Plants;Land Plants	NaN
Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes	FGSC A4
Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts	CL Brener
Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists	WB C6
...
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	NaN
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	ZP848
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	VKMY373
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	245
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	M20N0800_1A

	BioSample	BioProject	Assembly \
#Organism Name			
Emiliana huxleyi CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1
Hordeum vulgare	SAMEA3903302	PRJEB13020	GCA_900075435.2
Aspergillus nidulans FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2
Trypanosoma cruzi	SAMN02953627	PRJNA11755	GCA_000209065.1
Giardia lamblia ATCC 50803	SAMN02952905	PRJNA1439	GCA_000002435.1
...
Saccharomyces cerevisiae	SAMEA3110475	PRJEB7675	GCA_900178465.1
Saccharomyces cerevisiae	SAMEA3110493	PRJEB7675	GCA_900178475.1
Saccharomyces cerevisiae	SAMEA3110473	PRJEB7675	GCA_900178485.1
Saccharomyces cerevisiae	SAMEA3110476	PRJEB7675	GCA_900178495.1
Saccharomyces cerevisiae	SAMEA3110463	PRJEB7675	GCA_900178505.1

	Level	Size(Mb)	GC% \
#Organism Name			
Emiliana huxleyi CCMP1516	Scaffold	167.6760	64.5000
Hordeum vulgare	Scaffold	9788.8600	0.0000
Aspergillus nidulans FGSC A4	Scaffold	30.2760	50.2721
Trypanosoma cruzi	Scaffold	89.9375	51.7000
Giardia lamblia ATCC 50803	Scaffold	11.2136	49.2000
...
Saccharomyces cerevisiae	Scaffold	12.0392	38.1000
Saccharomyces cerevisiae	Scaffold	11.7086	38.1000
Saccharomyces cerevisiae	Scaffold	11.8194	38.0000
Saccharomyces cerevisiae	Scaffold	11.7081	38.1000
Saccharomyces cerevisiae	Scaffold	11.8189	38.1000

	Replicons	WGS \
#Organism Name		
Emiliana huxleyi CCMP1516	NaN	AHAL01

Hordeum vulgare		NaN	FJWB02
Aspergillus nidulans FGSC A4	mitochondrion MT:NC_017896.1/JQ435097.1	AACD01	
Trypanosoma cruzi		NaN	AAHK01
Giardia lamblia ATCC 50803		NaN	AACB02
...	
Saccharomyces cerevisiae		NaN	FXDE01
Saccharomyces cerevisiae		NaN	FXCK01
Saccharomyces cerevisiae		NaN	FXCD01
Saccharomyces cerevisiae		NaN	FXCP01
Saccharomyces cerevisiae		NaN	FXCG01

#Organism Name	Scaffolds	CDS	Release Date \
Emiliana huxleyi CCMP1516	7795	38554.0	2013-04-19 00:00:00+00:00
Hordeum vulgare	72295	0.0	2016-08-25 00:00:00+00:00
Aspergillus nidulans FGSC A4	91	9561.0	2003-04-07 00:00:00+00:00
Trypanosoma cruzi	29495	19607.0	2005-07-14 00:00:00+00:00
Giardia lamblia ATCC 50803	92	6502.0	2003-03-26 00:00:00+00:00
...
Saccharomyces cerevisiae	1925	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	590	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	2013	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	1028	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	799	0.0	2017-04-21 00:00:00+00:00

GenBank FTP

#Organism Name	
Emiliana huxleyi CCMP1516	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Hordeum vulgare	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Aspergillus nidulans FGSC A4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Trypanosoma cruzi	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Giardia lamblia ATCC 50803	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

RefSeq FTP

#Organism Name	
Emiliana huxleyi CCMP1516	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Hordeum vulgare	NaN
Aspergillus nidulans FGSC A4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Trypanosoma cruzi	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Giardia lamblia ATCC 50803	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...

```

...
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN

```

[3350 rows x 15 columns]

[111]: Level

```

[111]:      #Organism Name      Organism Groups \
0      Emiliana huxleyi CCMP1516 Eukaryota;Protists;Other Protists
1      Hordeum vulgare      Eukaryota;Plants;Land Plants
2      Aspergillus nidulans FGSC A4      Eukaryota;Fungi;Ascomycetes
3      Trypanosoma cruzi      Eukaryota;Protists;Kinetoplasts
4      Giardia lamblia ATCC 50803 Eukaryota;Protists;Other Protists
...
3345      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
3346      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
3347      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
3348      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
3349      Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes

```

```

      Strain      BioSample      BioProject      Assembly      Level \
0      CCMP1516      SAMN02744062      PRJNA77753      GCA_000372725.1      Scaffold
1      NaN      SAMEA3903302      PRJEB13020      GCA_900075435.2      Scaffold
2      FGSC A4      SAMN02953587      PRJNA130      GCA_000149205.2      Scaffold
3      CL Brener      SAMN02953627      PRJNA11755      GCA_000209065.1      Scaffold
4      WB C6      SAMN02952905      PRJNA1439      GCA_000002435.1      Scaffold
...
3345      NaN      SAMEA3110475      PRJEB7675      GCA_900178465.1      Scaffold
3346      ZP848      SAMEA3110493      PRJEB7675      GCA_900178475.1      Scaffold
3347      VKMY373      SAMEA3110473      PRJEB7675      GCA_900178485.1      Scaffold
3348      245      SAMEA3110476      PRJEB7675      GCA_900178495.1      Scaffold
3349      M2ON0800_1A      SAMEA3110463      PRJEB7675      GCA_900178505.1      Scaffold

```

```

      Size(Mb)      GC%      Replicons      WGS \
0      167.6760      64.5000      NaN      AHAL01
1      9788.8600      0.0000      NaN      FJWB02
2      30.2760      50.2721      mitochondrion MT:NC_017896.1/JQ435097.1      AACD01
3      89.9375      51.7000      NaN      AAHK01
4      11.2136      49.2000      NaN      AACB02
...
3345      12.0392      38.1000      NaN      FXDE01
3346      11.7086      38.1000      NaN      FXCK01
3347      11.8194      38.0000      NaN      FXCO01

```

3348	11.7081	38.1000	NaN	FXCP01
3349	11.8189	38.1000	NaN	FXCG01

	Scaffolds	CDS	Release Date	\
0	7795	38554.0	2013-04-19 00:00:00+00:00	
1	72295	0.0	2016-08-25 00:00:00+00:00	
2	91	9561.0	2003-04-07 00:00:00+00:00	
3	29495	19607.0	2005-07-14 00:00:00+00:00	
4	92	6502.0	2003-03-26 00:00:00+00:00	
...	
3345	1925	0.0	2017-04-21 00:00:00+00:00	
3346	590	0.0	2017-04-21 00:00:00+00:00	
3347	2013	0.0	2017-04-21 00:00:00+00:00	
3348	1028	0.0	2017-04-21 00:00:00+00:00	
3349	799	0.0	2017-04-21 00:00:00+00:00	

	GenBank FTP	\
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...	
...	...	
3345	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
3346	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
3347	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
3348	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	
3349	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...	

	RefSeq FTP	
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...	
1		NaN
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...	
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...	
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...	
...	...	
3345		NaN
3346		NaN
3347		NaN
3348		NaN
3349		NaN

[3350 rows x 16 columns]

```
[112]: Level.set_index("#Organism Name",inplace=True)
```

```
[113]: Level
```

[113]:

	Organism Groups	Strain \
#Organism Name		
Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists	CCMP1516
Hordeum vulgare	Eukaryota;Plants;Land Plants	NaN
Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes	FGSC A4
Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts	CL Brener
Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists	WB C6
...
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	NaN
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	ZP848
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	VKMY373
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	245
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	M20N0800_1A

	BioSample	BioProject	Assembly \
#Organism Name			
Emiliana huxleyi CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1
Hordeum vulgare	SAMEA3903302	PRJEB13020	GCA_900075435.2
Aspergillus nidulans FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2
Trypanosoma cruzi	SAMN02953627	PRJNA11755	GCA_000209065.1
Giardia lamblia ATCC 50803	SAMN02952905	PRJNA1439	GCA_000002435.1
...
Saccharomyces cerevisiae	SAMEA3110475	PRJEB7675	GCA_900178465.1
Saccharomyces cerevisiae	SAMEA3110493	PRJEB7675	GCA_900178475.1
Saccharomyces cerevisiae	SAMEA3110473	PRJEB7675	GCA_900178485.1
Saccharomyces cerevisiae	SAMEA3110476	PRJEB7675	GCA_900178495.1
Saccharomyces cerevisiae	SAMEA3110463	PRJEB7675	GCA_900178505.1

	Level	Size(Mb)	GC% \
#Organism Name			
Emiliana huxleyi CCMP1516	Scaffold	167.6760	64.5000
Hordeum vulgare	Scaffold	9788.8600	0.0000
Aspergillus nidulans FGSC A4	Scaffold	30.2760	50.2721
Trypanosoma cruzi	Scaffold	89.9375	51.7000
Giardia lamblia ATCC 50803	Scaffold	11.2136	49.2000
...
Saccharomyces cerevisiae	Scaffold	12.0392	38.1000
Saccharomyces cerevisiae	Scaffold	11.7086	38.1000
Saccharomyces cerevisiae	Scaffold	11.8194	38.0000
Saccharomyces cerevisiae	Scaffold	11.7081	38.1000
Saccharomyces cerevisiae	Scaffold	11.8189	38.1000

	Replicons	WGS \
#Organism Name		
Emiliana huxleyi CCMP1516	NaN	AHAL01
Hordeum vulgare	NaN	FJWB02
Aspergillus nidulans FGSC A4	mitochondrion MT:NC_017896.1/JQ435097.1	AACD01

Trypanosoma cruzi	NaN	AAHK01
Giardia lamblia ATCC 50803	NaN	AACB02
...
Saccharomyces cerevisiae	NaN	FXDE01
Saccharomyces cerevisiae	NaN	FXCK01
Saccharomyces cerevisiae	NaN	FXC001
Saccharomyces cerevisiae	NaN	FXCP01
Saccharomyces cerevisiae	NaN	FXCG01

#Organism Name	Scaffolds	CDS	Release Date \
Emiliana huxleyi CCMP1516	7795	38554.0	2013-04-19 00:00:00+00:00
Hordeum vulgare	72295	0.0	2016-08-25 00:00:00+00:00
Aspergillus nidulans FGSC A4	91	9561.0	2003-04-07 00:00:00+00:00
Trypanosoma cruzi	29495	19607.0	2005-07-14 00:00:00+00:00
Giardia lamblia ATCC 50803	92	6502.0	2003-03-26 00:00:00+00:00
...
Saccharomyces cerevisiae	1925	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	590	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	2013	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	1028	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	799	0.0	2017-04-21 00:00:00+00:00

GenBank FTP

#Organism Name	
Emiliana huxleyi CCMP1516	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Hordeum vulgare	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Aspergillus nidulans FGSC A4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Trypanosoma cruzi	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Giardia lamblia ATCC 50803	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

RefSeq FTP

#Organism Name	
Emiliana huxleyi CCMP1516	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Hordeum vulgare	NaN
Aspergillus nidulans FGSC A4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Trypanosoma cruzi	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Giardia lamblia ATCC 50803	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
Saccharomyces cerevisiae	NaN

Saccharomyces cerevisiae	NaN
Saccharomyces cerevisiae	NaN
Saccharomyces cerevisiae	NaN
Saccharomyces cerevisiae	NaN

[3350 rows x 15 columns]

```
[114]: Level.rename({"Scaffold":"Chromosome","Contig": "Complete"})
```

```
[114]:
```

	Organism Groups	Strain \
#Organism Name		
Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists	CCMP1516
Hordeum vulgare	Eukaryota;Plants;Land Plants	NaN
Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes	FGSC A4
Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts	CL Brener
Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists	WB C6
...
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	NaN
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	ZP848
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	VKMY373
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	245
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	M20N0800_1A

	BioSample	BioProject	Assembly \
#Organism Name			
Emiliana huxleyi CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1
Hordeum vulgare	SAMEA3903302	PRJEB13020	GCA_900075435.2
Aspergillus nidulans FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2
Trypanosoma cruzi	SAMN02953627	PRJNA11755	GCA_000209065.1
Giardia lamblia ATCC 50803	SAMN02952905	PRJNA1439	GCA_000002435.1
...
Saccharomyces cerevisiae	SAMEA3110475	PRJEB7675	GCA_900178465.1
Saccharomyces cerevisiae	SAMEA3110493	PRJEB7675	GCA_900178475.1
Saccharomyces cerevisiae	SAMEA3110473	PRJEB7675	GCA_900178485.1
Saccharomyces cerevisiae	SAMEA3110476	PRJEB7675	GCA_900178495.1
Saccharomyces cerevisiae	SAMEA3110463	PRJEB7675	GCA_900178505.1

	Level	Size(Mb)	GC% \
#Organism Name			
Emiliana huxleyi CCMP1516	Scaffold	167.6760	64.5000
Hordeum vulgare	Scaffold	9788.8600	0.0000
Aspergillus nidulans FGSC A4	Scaffold	30.2760	50.2721
Trypanosoma cruzi	Scaffold	89.9375	51.7000
Giardia lamblia ATCC 50803	Scaffold	11.2136	49.2000
...
Saccharomyces cerevisiae	Scaffold	12.0392	38.1000
Saccharomyces cerevisiae	Scaffold	11.7086	38.1000

Saccharomyces cerevisiae	Scaffold	11.8194	38.0000
Saccharomyces cerevisiae	Scaffold	11.7081	38.1000
Saccharomyces cerevisiae	Scaffold	11.8189	38.1000

	Replicons	WGS \
#Organism Name		
Emiliana huxleyi CCMP1516	NaN	AHAL01
Hordeum vulgare	NaN	FJWB02
Aspergillus nidulans FGSC A4 mitochondrion MT:NC_017896.1/JQ435097.1		AACD01
Trypanosoma cruzi	NaN	AAHK01
Giardia lamblia ATCC 50803	NaN	AACB02
...
Saccharomyces cerevisiae	NaN	FXDE01
Saccharomyces cerevisiae	NaN	FXCK01
Saccharomyces cerevisiae	NaN	FXCO01
Saccharomyces cerevisiae	NaN	FXCP01
Saccharomyces cerevisiae	NaN	FXCG01

	Scaffolds	CDS	Release Date \
#Organism Name			
Emiliana huxleyi CCMP1516	7795	38554.0	2013-04-19 00:00:00+00:00
Hordeum vulgare	72295	0.0	2016-08-25 00:00:00+00:00
Aspergillus nidulans FGSC A4	91	9561.0	2003-04-07 00:00:00+00:00
Trypanosoma cruzi	29495	19607.0	2005-07-14 00:00:00+00:00
Giardia lamblia ATCC 50803	92	6502.0	2003-03-26 00:00:00+00:00
...
Saccharomyces cerevisiae	1925	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	590	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	2013	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	1028	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	799	0.0	2017-04-21 00:00:00+00:00

	GenBank FTP
\	
#Organism Name	
Emiliana huxleyi CCMP1516	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Hordeum vulgare	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Aspergillus nidulans FGSC A4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Trypanosoma cruzi	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Giardia lamblia ATCC 50803	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...


```

#Organism Name
Emiliana huxleyi CCMP1516 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Hordeum vulgare NaN
Aspergillus nidulans FGSC A4 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Trypanosoma cruzi ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Giardia lamblia ATCC 50803 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN

```

[3350 rows x 15 columns]

```
[115]: Level.rename({"Scaffold":"Chromosome","Contig": "Complete"},axis=1)
```

```

[115]:
#Organism Name
Emiliana huxleyi CCMP1516 Eukaryota;Protists;Other Protists CCMP1516
Hordeum vulgare Eukaryota;Plants;Land Plants NaN
Aspergillus nidulans FGSC A4 Eukaryota;Fungi;Ascomycetes FGSC A4
Trypanosoma cruzi Eukaryota;Protists;Kinetoplasts CL Brener
Giardia lamblia ATCC 50803 Eukaryota;Protists;Other Protists WB C6
...
Saccharomyces cerevisiae Eukaryota;Fungi;Ascomycetes NaN
Saccharomyces cerevisiae Eukaryota;Fungi;Ascomycetes ZP848
Saccharomyces cerevisiae Eukaryota;Fungi;Ascomycetes VKMY373
Saccharomyces cerevisiae Eukaryota;Fungi;Ascomycetes 245
Saccharomyces cerevisiae Eukaryota;Fungi;Ascomycetes M2ON0800_1A

```

```

BioSample BioProject Assembly \
#Organism Name
Emiliana huxleyi CCMP1516 SAMN02744062 PRJNA77753 GCA_000372725.1
Hordeum vulgare SAMEA3903302 PRJEB13020 GCA_900075435.2
Aspergillus nidulans FGSC A4 SAMN02953587 PRJNA130 GCA_000149205.2
Trypanosoma cruzi SAMN02953627 PRJNA11755 GCA_000209065.1
Giardia lamblia ATCC 50803 SAMN02952905 PRJNA1439 GCA_000002435.1
...
Saccharomyces cerevisiae SAMEA3110475 PRJEB7675 GCA_900178465.1
Saccharomyces cerevisiae SAMEA3110493 PRJEB7675 GCA_900178475.1
Saccharomyces cerevisiae SAMEA3110473 PRJEB7675 GCA_900178485.1
Saccharomyces cerevisiae SAMEA3110476 PRJEB7675 GCA_900178495.1
Saccharomyces cerevisiae SAMEA3110463 PRJEB7675 GCA_900178505.1

```

```
Level Size(Mb) GC% \
```

#Organism Name				
Emiliana huxleyi CCMP1516	Scaffold	167.6760	64.5000	
Hordeum vulgare	Scaffold	9788.8600	0.0000	
Aspergillus nidulans FGSC A4	Scaffold	30.2760	50.2721	
Trypanosoma cruzi	Scaffold	89.9375	51.7000	
Giardia lamblia ATCC 50803	Scaffold	11.2136	49.2000	
...	
Saccharomyces cerevisiae	Scaffold	12.0392	38.1000	
Saccharomyces cerevisiae	Scaffold	11.7086	38.1000	
Saccharomyces cerevisiae	Scaffold	11.8194	38.0000	
Saccharomyces cerevisiae	Scaffold	11.7081	38.1000	
Saccharomyces cerevisiae	Scaffold	11.8189	38.1000	

		Replicons	WGS \
#Organism Name			
Emiliana huxleyi CCMP1516		NaN	AHAL01
Hordeum vulgare		NaN	FJWB02
Aspergillus nidulans FGSC A4	mitochondrion MT:NC_017896.1/JQ435097.1		AACD01
Trypanosoma cruzi		NaN	AAHK01
Giardia lamblia ATCC 50803		NaN	AACB02
...	
Saccharomyces cerevisiae		NaN	FXDE01
Saccharomyces cerevisiae		NaN	FXCK01
Saccharomyces cerevisiae		NaN	FXC001
Saccharomyces cerevisiae		NaN	FXCP01
Saccharomyces cerevisiae		NaN	FXCG01

	Scaffolds	CDS	Release Date \
#Organism Name			
Emiliana huxleyi CCMP1516	7795	38554.0	2013-04-19 00:00:00+00:00
Hordeum vulgare	72295	0.0	2016-08-25 00:00:00+00:00
Aspergillus nidulans FGSC A4	91	9561.0	2003-04-07 00:00:00+00:00
Trypanosoma cruzi	29495	19607.0	2005-07-14 00:00:00+00:00
Giardia lamblia ATCC 50803	92	6502.0	2003-03-26 00:00:00+00:00
...
Saccharomyces cerevisiae	1925	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	590	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	2013	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	1028	0.0	2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	799	0.0	2017-04-21 00:00:00+00:00

GenBank FTP

#Organism Name	
Emiliana huxleyi CCMP1516	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Hordeum vulgare	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Aspergillus nidulans FGSC A4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...

Trypanosoma cruzi	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Giardia lamblia ATCC 50803	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

RefSeq FTP

#Organism Name	
Emiliana huxleyi CCMP1516	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Hordeum vulgare	NaN
Aspergillus nidulans FGSC A4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Trypanosoma cruzi	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Giardia lamblia ATCC 50803	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
Saccharomyces cerevisiae	NaN
Saccharomyces cerevisiae	NaN
Saccharomyces cerevisiae	NaN
Saccharomyces cerevisiae	NaN
Saccharomyces cerevisiae	NaN

[3350 rows x 15 columns]

[116]: Level

	Organism Groups	Strain \
#Organism Name		
Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists	CCMP1516
Hordeum vulgare	Eukaryota;Plants;Land Plants	NaN
Aspergillus nidulans FGSC A4	Eukaryota;Fungi;Ascomycetes	FGSC A4
Trypanosoma cruzi	Eukaryota;Protists;Kinetoplasts	CL Brener
Giardia lamblia ATCC 50803	Eukaryota;Protists;Other Protists	WB C6
...
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	NaN
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	ZP848
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	VKMY373
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	245
Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes	M2ON0800_1A

	BioSample	BioProject	Assembly \
#Organism Name			
Emiliana huxleyi CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1
Hordeum vulgare	SAMEA3903302	PRJEB13020	GCA_900075435.2
Aspergillus nidulans FGSC A4	SAMN02953587	PRJNA130	GCA_000149205.2
Trypanosoma cruzi	SAMN02953627	PRJNA11755	GCA_000209065.1

Giardia lamblia ATCC 50803	SAMN02952905	PRJNA1439	GCA_000002435.1
...
Saccharomyces cerevisiae	SAMEA3110475	PRJEB7675	GCA_900178465.1
Saccharomyces cerevisiae	SAMEA3110493	PRJEB7675	GCA_900178475.1
Saccharomyces cerevisiae	SAMEA3110473	PRJEB7675	GCA_900178485.1
Saccharomyces cerevisiae	SAMEA3110476	PRJEB7675	GCA_900178495.1
Saccharomyces cerevisiae	SAMEA3110463	PRJEB7675	GCA_900178505.1

	Level	Size(Mb)	GC%	\
#Organism Name				
Emiliana huxleyi CCMP1516	Scaffold	167.6760	64.5000	
Hordeum vulgare	Scaffold	9788.8600	0.0000	
Aspergillus nidulans FGSC A4	Scaffold	30.2760	50.2721	
Trypanosoma cruzi	Scaffold	89.9375	51.7000	
Giardia lamblia ATCC 50803	Scaffold	11.2136	49.2000	
...	
Saccharomyces cerevisiae	Scaffold	12.0392	38.1000	
Saccharomyces cerevisiae	Scaffold	11.7086	38.1000	
Saccharomyces cerevisiae	Scaffold	11.8194	38.0000	
Saccharomyces cerevisiae	Scaffold	11.7081	38.1000	
Saccharomyces cerevisiae	Scaffold	11.8189	38.1000	

	Replicons	WGS	\
#Organism Name			
Emiliana huxleyi CCMP1516	NaN	AHAL01	
Hordeum vulgare	NaN	FJWB02	
Aspergillus nidulans FGSC A4	mitochondrion MT:NC_017896.1/JQ435097.1	AACD01	
Trypanosoma cruzi	NaN	AAHK01	
Giardia lamblia ATCC 50803	NaN	AACB02	
...	
Saccharomyces cerevisiae	NaN	FXDE01	
Saccharomyces cerevisiae	NaN	FXCK01	
Saccharomyces cerevisiae	NaN	FXCO01	
Saccharomyces cerevisiae	NaN	FXCP01	
Saccharomyces cerevisiae	NaN	FXCG01	

	Scaffolds	CDS	Release Date	\
#Organism Name				
Emiliana huxleyi CCMP1516	7795	38554.0	2013-04-19 00:00:00+00:00	
Hordeum vulgare	72295	0.0	2016-08-25 00:00:00+00:00	
Aspergillus nidulans FGSC A4	91	9561.0	2003-04-07 00:00:00+00:00	
Trypanosoma cruzi	29495	19607.0	2005-07-14 00:00:00+00:00	
Giardia lamblia ATCC 50803	92	6502.0	2003-03-26 00:00:00+00:00	
...	
Saccharomyces cerevisiae	1925	0.0	2017-04-21 00:00:00+00:00	
Saccharomyces cerevisiae	590	0.0	2017-04-21 00:00:00+00:00	
Saccharomyces cerevisiae	2013	0.0	2017-04-21 00:00:00+00:00	

Saccharomyces cerevisiae	1028	0.0 2017-04-21 00:00:00+00:00
Saccharomyces cerevisiae	799	0.0 2017-04-21 00:00:00+00:00

GenBank FTP

```

\
#Organism Name
Emiliana huxleyi CCMP1516 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Hordeum vulgare ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Aspergillus nidulans FGSC A4 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Trypanosoma cruzi ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Giardia lamblia ATCC 50803 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...
Saccharomyces cerevisiae ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
Saccharomyces cerevisiae ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

RefSeq FTP

```

#Organism Name
Emiliana huxleyi CCMP1516 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Hordeum vulgare NaN
Aspergillus nidulans FGSC A4 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Trypanosoma cruzi ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Giardia lamblia ATCC 50803 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN
Saccharomyces cerevisiae NaN

```

[3350 rows x 15 columns]

[117]: *#nlargest and nsmallest*

[118]: Level.nlargest(n=10, columns="CDS")

[118]:	Organism Groups	Strain \
#Organism Name		
Nicotiana tabacum	Eukaryota;Plants;Land Plants	NaN
Quercus suber	Eukaryota;Plants;Land Plants	HL8
Sinocyclocheilus rhinoceros	Eukaryota;Animals;Fishes	Xijiao
Sinocyclocheilus anshuiensis	Eukaryota;Animals;Fishes	Anshui
Sinocyclocheilus grahami	Eukaryota;Animals;Fishes	Dianchi
Cercocebus atys	Eukaryota;Animals;Mammals	FAK
Ancylostoma ceylanicum	Eukaryota;Animals;Roundworms	HY135

Chenopodium quinoa	Eukaryota;Plants;Land Plants	NaN
Durio zibethinus	Eukaryota;Plants;Land Plants	D1
Macaca nemestrina	Eukaryota;Animals;Mammals	M95218

	BioSample	BioProject	Assembly \
#Organism Name			
Nicotiana tabacum	SAMN02316627	PRJNA319578	GCA_000715135.1
Quercus suber	SAMN07312269	PRJNA433227	GCA_002906115.1
Sinocyclocheilus rhinoceros	SAMN03320098	PRJNA316320	GCA_001515625.1
Sinocyclocheilus anshuiensis	SAMN03320099	PRJNA316320	GCA_001515605.1
Sinocyclocheilus grahami	SAMN03320097	PRJNA316320	GCA_001515645.1
Cercocebus atys	SAMN03379531	PRJNA279144	GCA_000955945.1
Ancylostoma ceylanicum	SAMN02585415	PRJNA259754	GCA_000688135.1
Chenopodium quinoa	SAMN04338310	PRJNA394242	GCA_001683475.1
Durio zibethinus	SAMN07563003	PRJNA407962	GCA_002303985.1
Macaca nemestrina	SAMN03121842	PRJNA279145	GCA_000956065.1

	Level	Size(Mb)	GC% \
#Organism Name			
Nicotiana tabacum	Scaffold	3643.470	39.2006
Quercus suber	Scaffold	953.299	35.5000
Sinocyclocheilus rhinoceros	Scaffold	1655.790	38.1000
Sinocyclocheilus anshuiensis	Scaffold	1632.720	38.0001
Sinocyclocheilus grahami	Scaffold	1750.290	38.7000
Cercocebus atys	Scaffold	2848.250	41.1000
Ancylostoma ceylanicum	Scaffold	313.093	43.5000
Chenopodium quinoa	Scaffold	1333.550	37.0000
Durio zibethinus	Scaffold	715.230	32.7000
Macaca nemestrina	Scaffold	2948.700	41.3000

	Replicons
\	
#Organism Name	
Nicotiana tabacum	plastid Pltd:NC_001879.2/Z00044.2; mitochondri...
Quercus suber	NaN
Sinocyclocheilus rhinoceros	mitochondrion MT:NC_027168.1/KR069119.1
Sinocyclocheilus anshuiensis	mitochondrion MT:NC_027169.1/KR069120.1
Sinocyclocheilus grahami	mitochondrion MT:NC_013189.1/GQ148557.1
Cercocebus atys	NaN
Ancylostoma ceylanicum	NaN
Chenopodium quinoa	chloroplast Pltd:NC_034949.1/KY419706.1
Durio zibethinus	NaN
Macaca nemestrina	NaN

	WGS	Scaffolds	CDS \
#Organism Name			
Nicotiana tabacum	AYMY01	168247	84255.0

Quercus suber	PKMF01	23344	83154.0
Sinocyclocheilus rhinoceros	LAVF01	164173	68562.0
Sinocyclocheilus anshuiensis	LAVE01	85682	68474.0
Sinocyclocheilus grahami	LCYQ01	31277	67410.0
Cercocebus atys	JZLG01	11433	65920.0
Ancylostoma ceylanicum	JARK01	1736	65583.0
Chenopodium quinoa	LPWI01	3487	63173.0
Durio zibethinus	NSDW01	677	63007.0
Macaca nemestrina	JZLF01	9733	62876.0

Release Date \

#Organism Name	
Nicotiana tabacum	2014-05-29 00:00:00+00:00
Quercus suber	2018-01-29 00:00:00+00:00
Sinocyclocheilus rhinoceros	2015-12-14 00:00:00+00:00
Sinocyclocheilus anshuiensis	2015-12-14 00:00:00+00:00
Sinocyclocheilus grahami	2015-12-16 00:00:00+00:00
Cercocebus atys	2015-03-16 00:00:00+00:00
Ancylostoma ceylanicum	2014-03-25 00:00:00+00:00
Chenopodium quinoa	2016-07-11 00:00:00+00:00
Durio zibethinus	2017-09-15 00:00:00+00:00
Macaca nemestrina	2015-03-16 00:00:00+00:00

GenBank FTP

\

#Organism Name	
Nicotiana tabacum	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Quercus suber	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
Sinocyclocheilus rhinoceros	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
Sinocyclocheilus anshuiensis	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
Sinocyclocheilus grahami	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
Cercocebus atys	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Ancylostoma ceylanicum	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Chenopodium quinoa	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
Durio zibethinus	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
Macaca nemestrina	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...

RefSeq FTP

#Organism Name	
Nicotiana tabacum	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Quercus suber	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002...
Sinocyclocheilus rhinoceros	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...
Sinocyclocheilus anshuiensis	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...
Sinocyclocheilus grahami	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...
Cercocebus atys	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
Ancylostoma ceylanicum	NaN
Chenopodium quinoa	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...

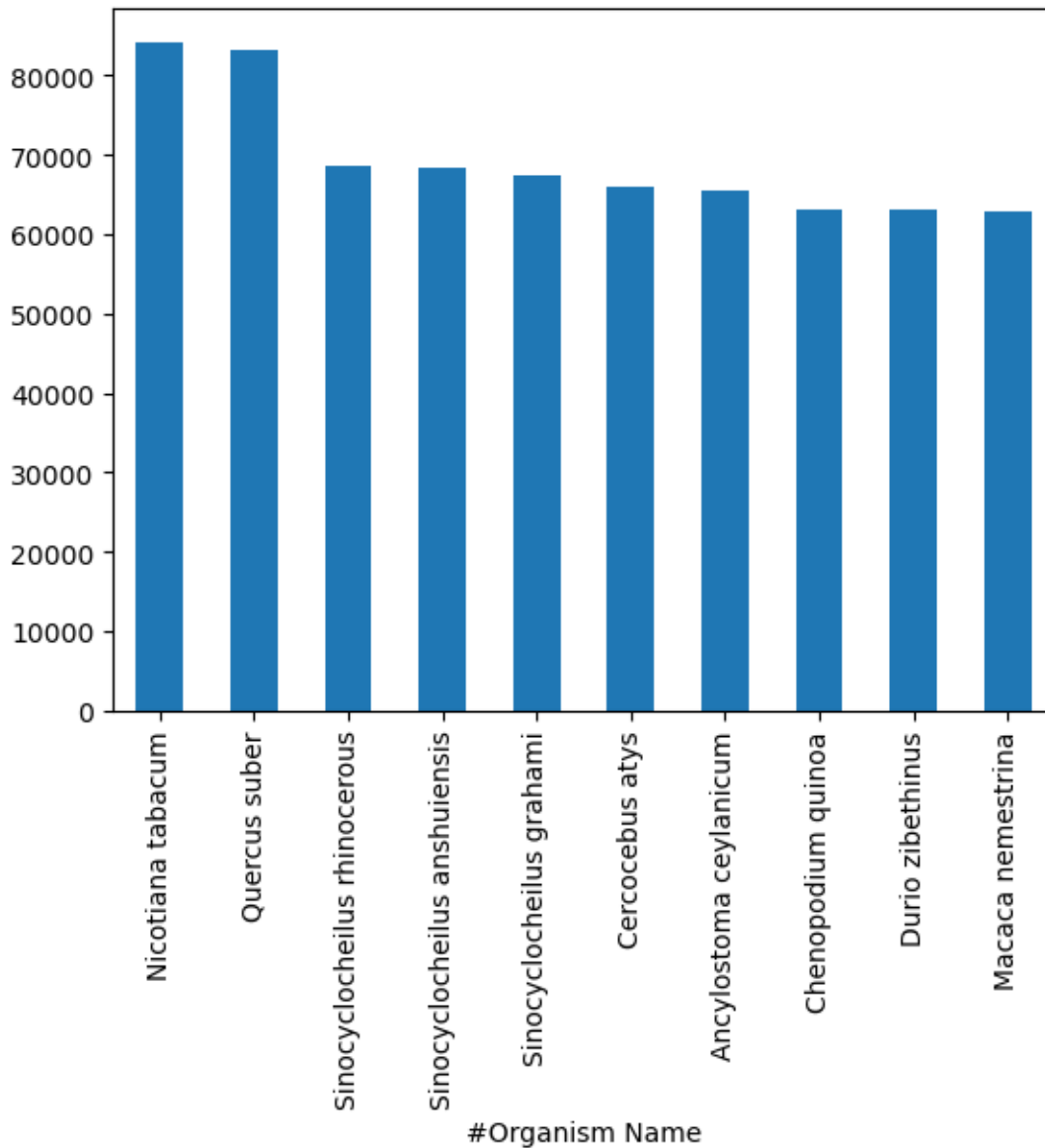
Durio zibethinus	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002...
Macaca nemestrina	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...

```
[119]: Level.nlargest(n=10, columns="CDS").loc[:, "CDS"]
```

```
[119]: #Organism Name
Nicotiana tabacum      84255.0
Quercus suber          83154.0
Sinocyclocheilus rhinoceros 68562.0
Sinocyclocheilus anshuiensis 68474.0
Sinocyclocheilus grahami 67410.0
Cercopithecus atys     65920.0
Ancylostoma ceylanicum 65583.0
Chenopodium quinoa     63173.0
Durio zibethinus       63007.0
Macaca nemestrina      62876.0
Name: CDS, dtype: float64
```

```
[120]: Level.nlargest(n=10, columns="CDS").loc[:, "CDS"].plot.bar()
```

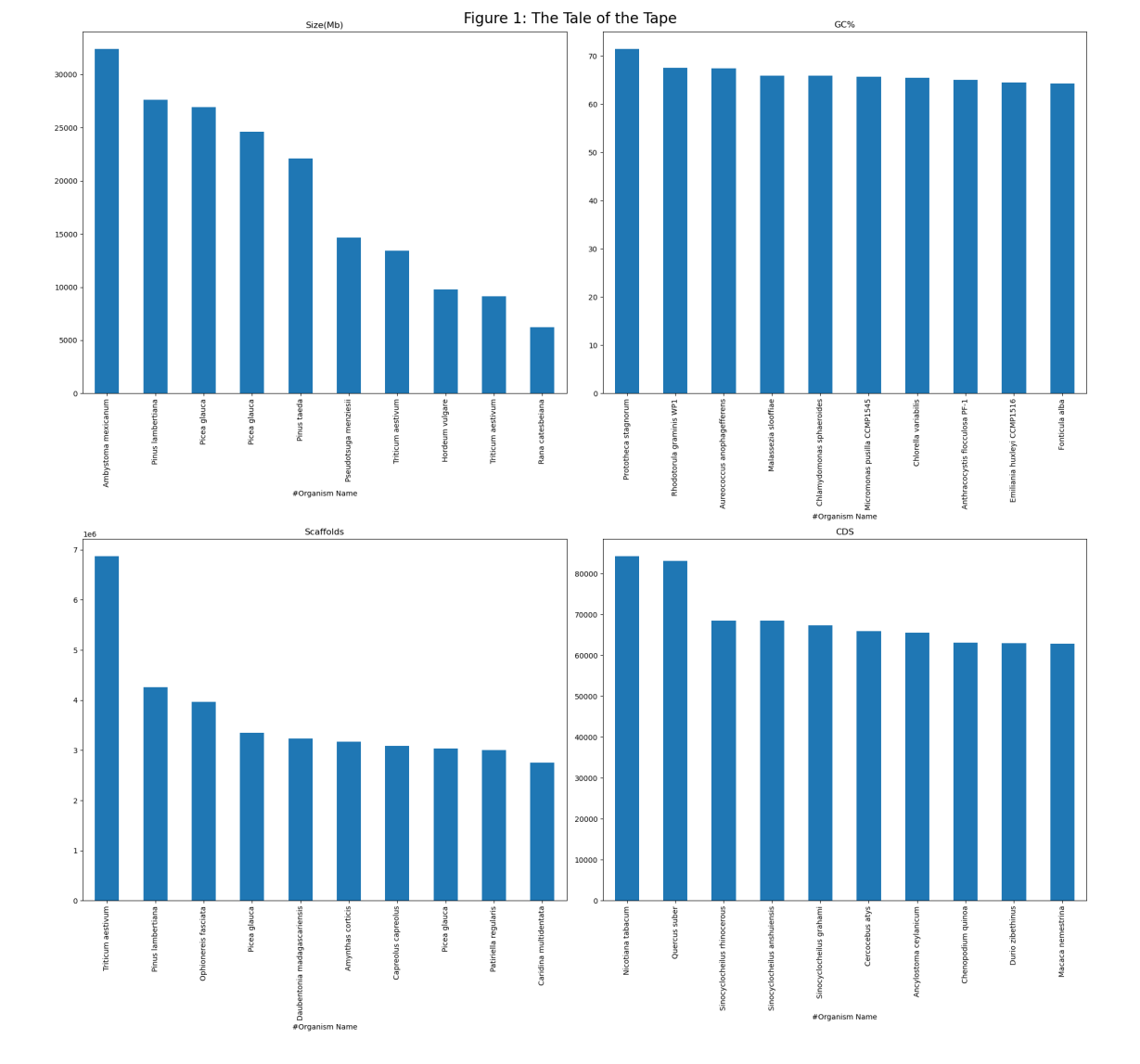
```
[120]: <Axes: xlabel='#Organism Name'>
```

```
[121]: fig, ax = plt.subplots(2,2, figsize=(20,20))

Level.nlargest(n=10, columns="Size(Mb)").loc[:, "Size(Mb)"].plot.bar(ax=ax[0][0])
ax[0][0].set_title("Size(Mb)")
Level.nlargest(n=10, columns="GC%").loc[:, "GC%"].plot.bar(ax=ax[0][1])
ax[0][1].set_title("GC%")
Level.nlargest(n=10, columns="Scaffolds").loc[:, "Scaffolds"].plot.
    ↪ bar(ax=ax[1][0])
ax[1][0].set_title("Scaffolds")
Level.nlargest(n=10, columns="CDS").loc[:, "CDS"].plot.bar(ax=ax[1][1])
ax[1][1].set_title("CDS")
```

```
fig.suptitle("Figure 1: The Tale of the Tape", fontsize=20)
fig.tight_layout()
```



```
[122]: Level.nsmallest(n=5, columns="Size(Mb)")
```

[122]:		Organism Groups	Strain \
#Organism Name			
Rotaria magnacalcarata	Eukaryota;Animals;Other Animals	TS_RM01	
Diospyros lotus	Eukaryota;Plants;Land Plants	NaN	
Astrammina rara	Eukaryota;Protists;Other Protists	NaN	
Primula vulgaris	Eukaryota;Plants;Land Plants	NaN	
Nematocida displodere	Eukaryota;Fungi;Other Fungi	JUm2807	

	BioSample	BioProject	Assembly	Level \
#Organism Name				
Rotaria magnacalcarata	SAMN04076332	PRJNA295611	GCA_001317505.1	Scaffold
Diospyros lotus	SAMN03019951	PRJNA256104	GCA_000774125.1	Scaffold
Astrammina rara	SAMN02953772	PRJNA47149	GCA_000211355.2	Scaffold
Primula vulgaris	SAMEA2802728	PRJEB7311	GCA_001077355.1	Scaffold
Nematocida displodere	SAMN04499096	PRJNA312555	GCA_001642395.1	Scaffold

	Size(Mb)	GC%	Replicons	WGS	Scaffolds	CDS \
#Organism Name						
Rotaria magnacalcarata	0.702958	34.9	NaN	LJPC01	75	0.0
Diospyros lotus	1.104190	39.5	NaN	JRBH01	796	0.0
Astrammina rara	1.450090	42.0	NaN	ADNL01	3231	0.0
Primula vulgaris	1.504780	35.6	NaN	CDJJ02	229	0.0
Nematocida displodere	3.066400	49.1	NaN	LTDL01	42	2278.0

	Release Date \
#Organism Name	
Rotaria magnacalcarata	2015-10-16 00:00:00+00:00
Diospyros lotus	2014-11-10 00:00:00+00:00
Astrammina rara	2011-04-27 00:00:00+00:00
Primula vulgaris	2015-01-28 00:00:00+00:00
Nematocida displodere	2016-05-11 00:00:00+00:00

	GenBank FTP \
#Organism Name	
Rotaria magnacalcarata	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
Diospyros lotus	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Astrammina rara	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
Primula vulgaris	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
Nematocida displodere	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...

	RefSeq FTP
#Organism Name	
Rotaria magnacalcarata	NaN
Diospyros lotus	NaN
Astrammina rara	NaN
Primula vulgaris	NaN
Nematocida displodere	NaN

```
[123]: # value_counts()

# Assuming 'data' is your DataFrame containing the relevant data
Strain = pd.DataFrame(data)
top_100 = Strain.nlargest(n=100, columns="Scaffolds")
```

```
[124]: top_100["Strain"].value_counts()
```

```
[124]: Strain
USFS_5038          1
Rimini            1
isofemale 4       1
motley            1
NA19240           1
WOAK              1
Rspec5            1
CLI525            1
Kalimpong         1
mixed             1
4540-9            1
ML1               1
Etri01            1
WM_1_Wirral       1
Blida - Haddad    1
MA1               1
2881A             1
Ipuh              1
2006-1            1
0317_1            1
A0245             1
Dm6514m           1
Shenzhen          1
Guangxi           1
Petshop1          1
PSP_Ac18          1
A                 1
Bruno             1
AM-GW-M-EMB-4M-20130531 1
SBHNorthAmerica   1
2KALR             1
A0248             1
MSM/Ms            1
Weyco1            1
Teifi             1
LN10              1
Name: count, dtype: int64
```

```
[125]: top_100["Strain"].value_counts(normalize=True)
```

```
[125]: Strain
USFS_5038          0.027778
Rimini            0.027778
isofemale 4       0.027778
```

motley	0.027778
NA19240	0.027778
WOAK	0.027778
Rspec5	0.027778
CLI525	0.027778
Kalimpong	0.027778
mixed	0.027778
4540-9	0.027778
ML1	0.027778
Etri01	0.027778
WM_1_Wirral	0.027778
Blida - Haddad	0.027778
MA1	0.027778
2881A	0.027778
Ipuh	0.027778
2006-1	0.027778
0317_1	0.027778
A0245	0.027778
Dm6514m	0.027778
Shenzhen	0.027778
Guangxi	0.027778
Petshop1	0.027778
PSP_Ac18	0.027778
A	0.027778
Bruno	0.027778
AM-GW-M-EMB-4M-20130531	0.027778
SBHNorthAmerica	0.027778
2KALR	0.027778
A0248	0.027778
MSM/Ms	0.027778
Weyco1	0.027778
Teifi	0.027778
LN10	0.027778

Name: proportion, dtype: float64

```
[126]: top_100["Strain"].value_counts(normalize=True,ascending=False)
```

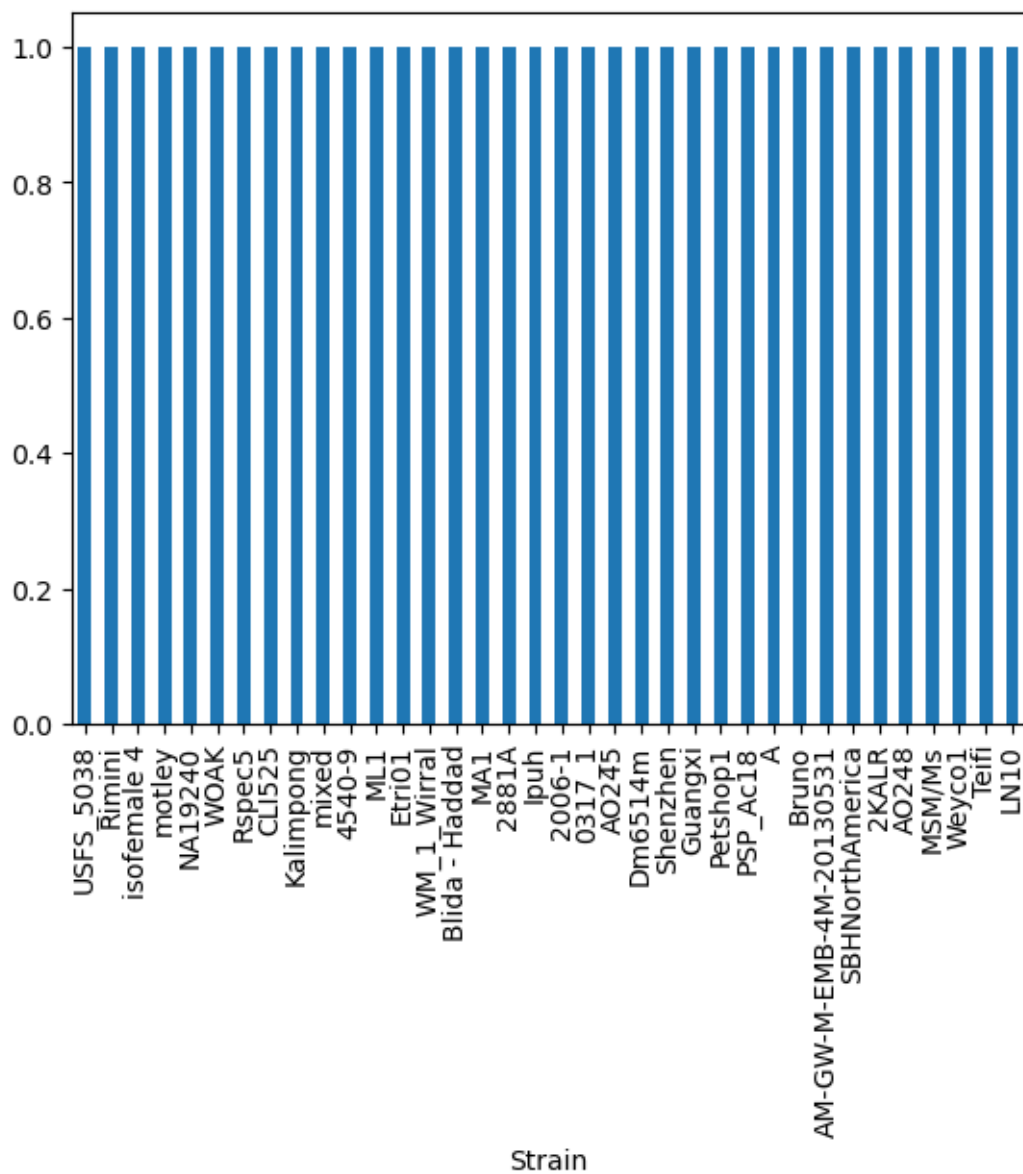
```
[126]: Strain
USFS_5038      0.027778
Rimini         0.027778
isofemale 4    0.027778
motley         0.027778
NA19240        0.027778
WOAK           0.027778
Rspec5         0.027778
CLI525         0.027778
Kalimpong      0.027778
```

mixed	0.027778
4540-9	0.027778
ML1	0.027778
Etri01	0.027778
WM_1_Wirral	0.027778
Blida - Haddad	0.027778
MA1	0.027778
2881A	0.027778
Ipuh	0.027778
2006-1	0.027778
0317_1	0.027778
A0245	0.027778
Dm6514m	0.027778
Shenzhen	0.027778
Guangxi	0.027778
Petshop1	0.027778
PSP_Ac18	0.027778
A	0.027778
Bruno	0.027778
AM-GW-M-EMB-4M-20130531	0.027778
SBHNorthAmerica	0.027778
2KALR	0.027778
A0248	0.027778
MSM/Ms	0.027778
Weyco1	0.027778
Teifi	0.027778
LN10	0.027778

Name: proportion, dtype: float64

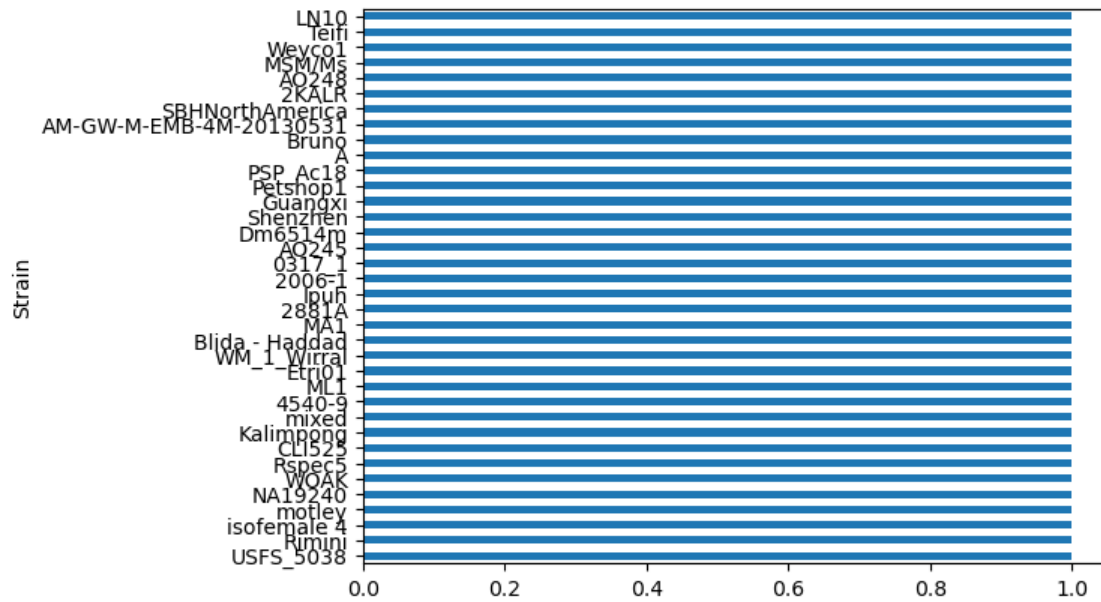
```
[127]: top_100["Strain"].value_counts().plot.bar()
```

```
[127]: <Axes: xlabel='Strain'>
```



```
[128]: top_100["Strain"].value_counts().plot.barh()
```

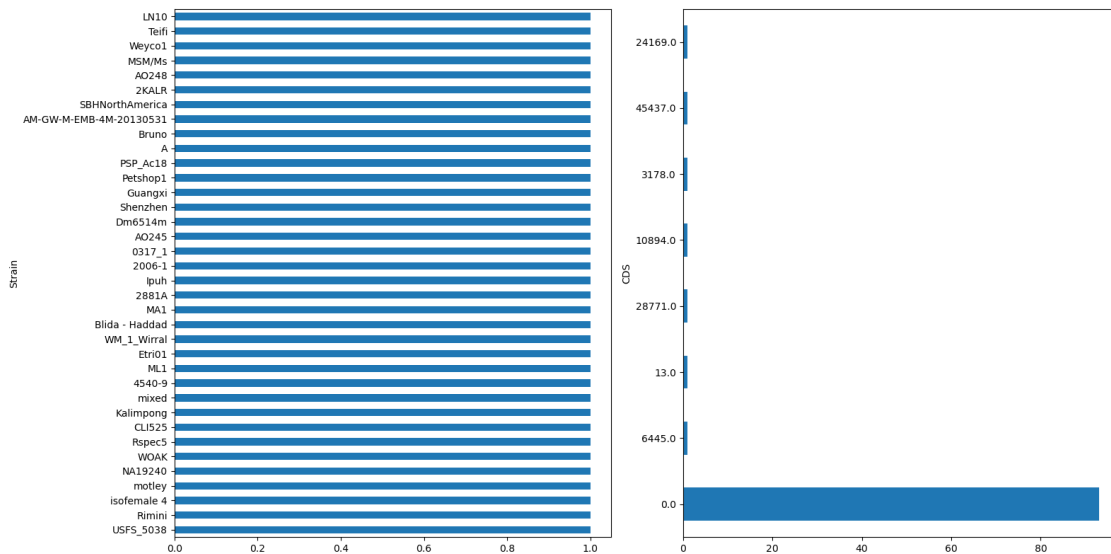
```
[128]: <Axes: ylabel='Strain'>
```



```
[129]: fig, ax =plt.subplots(1,2,figsize=(16,8))

top_100["Strain"].value_counts().plot.barh(ax=ax[0])
top_100["CDS"].value_counts().plot.barh(ax=ax[1])

plt.tight_layout()
```



[130]: # sort_values

Strain

[130]:

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold
5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%	Replicons \
0	167.6760	64.5000	NaN
1	119.6680	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...
2	412.9240	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...
3	978.9720	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...
4	823.7860	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...
...
5126	12.0392	38.1000	NaN
5127	11.7086	38.1000	NaN
5128	11.8194	38.0000	NaN
5129	11.7081	38.1000	NaN
5130	11.8189	38.1000	NaN

	WGS	Scaffolds	CDS	Release Date \
0	AHAL01	7795	38554.0	2013-04-19 00:00:00+00:00
1	NaN	7	48350.0	2001-08-13 00:00:00+00:00
2	APN001	2187	57661.0	2011-08-12 00:00:00+00:00

3	ACUP02	1191	71525.0	2010-01-05	00:00:00+00:00
4	AEKE02	3224	36010.0	2010-12-10	00:00:00+00:00
...
5126	FXDE01	1925	0.0	2017-04-21	00:00:00+00:00
5127	FXCK01	590	0.0	2017-04-21	00:00:00+00:00
5128	FXC001	2013	0.0	2017-04-21	00:00:00+00:00
5129	FXCP01	1028	0.0	2017-04-21	00:00:00+00:00
5130	FXCG01	799	0.0	2017-04-21	00:00:00+00:00

GenBank FTP \

0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

RefSeq FTP

0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
5126	NaN
5127	NaN
5128	NaN
5129	NaN
5130	NaN

[5131 rows x 16 columns]

```
[131]: data.sort_values(by="Strain", ascending=False)
```

	#Organism Name	Organism Groups \
2439	Phytophthora megakarya	Eukaryota;Protists;Other Protists
3096	Paralichthys olivaceus	Eukaryota;Animals;Fishes
1325	Bos mutus	Eukaryota;Animals;Mammals
2981	Papilio xuthus	Eukaryota;Animals;Insects
1003	Papilio machaon	Eukaryota;Animals;Insects
...
5087	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

5088	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5089	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5090	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	\
2439	zdh0120	SAMN04632436	PRJNA318028	GCA_002215365.1	
3096	ysfri_paol	SAMN05981185	PRJNA73673	GCA_001904815.2	
1325	yakQH1	SAMN00744358	PRJNA74739	GCA_000298355.1	
2981	ya'a_city_454_Px	SAMN03290671	PRJNA270384	GCA_001298345.1	
1003	ya'a_city_454_Pm	SAMN03290678	PRJNA300299	GCA_001298355.1	
...	
5087	NaN	SAMEA3249804	PRJEB8455	GCA_900178035.1	
5088	NaN	SAMEA3249826	PRJEB8455	GCA_900178045.1	
5089	NaN	SAMEA3249812	PRJEB8455	GCA_900178065.1	
5090	NaN	SAMEA3249800	PRJEB8455	GCA_900178075.1	
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	

	Level	Size(Mb)	GC%	\
2439	Scaffold	101.5050	48.7000	
3096	Chromosome	545.7750	41.6195	
1325	Scaffold	2645.1600	42.0000	
2981	Scaffold	243.2320	34.9000	
1003	Scaffold	278.4360	34.3992	
...	
5087	Scaffold	13.8535	41.9000	
5088	Scaffold	11.5139	38.2000	
5089	Scaffold	11.5691	38.1000	
5090	Scaffold	12.5459	40.0000	
5126	Scaffold	12.0392	38.1000	

	Replicons	WGS	Scaffolds	\
2439	NaN	NBNE01	24039	
3096	chromosome 1:CM007757.1; chromosome 2:CM007758...	AGQT02	7202	
1325	mitochondrion MT:NC_006380.3/EF494179.1	AGSK01	41192	
2981	NaN	LADI01	15362	
1003	mitochondrion MT:NC_018047.1/HM243594.1	LADJ01	63187	
...	
5087	NaN	FXDY01	8158	
5088	NaN	FXDN01	1441	
5089	NaN	FXLH01	1144	
5090	NaN	FXDS01	6246	
5126	NaN	FXDE01	1925	

	CDS	Release Date	\
2439	34804.0	2017-07-06 00:00:00+00:00	
3096	0.0	2016-12-06 00:00:00+00:00	

```

1325  28881.0  2012-05-21  00:00:00+00:00
2981  15322.0  2015-09-24  00:00:00+00:00
1003  17758.0  2015-09-24  00:00:00+00:00
...
5087      0.0  2017-04-25  00:00:00+00:00
5088      0.0  2017-04-25  00:00:00+00:00
5089      0.0  2017-04-26  00:00:00+00:00
5090      0.0  2017-04-25  00:00:00+00:00
5126      0.0  2017-04-21  00:00:00+00:00

```

GenBank FTP \

```

2439  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
3096  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
1325  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2981  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
1003  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...
...
5087  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5088  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5089  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5090  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5126  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

```

RefSeq FTP

```

2439      NaN
3096      NaN
1325  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2981      NaN
1003  ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...
...
5087      NaN
5088      NaN
5089      NaN
5090      NaN
5126      NaN

```

[5131 rows x 16 columns]

```
[132]: data.sort_values(by="Release Date",ascending=False)
```

```

[132]:      #Organism Name      Organism Groups \
2685      Monilinia laxa      Eukaryota;Fungi;Ascomycetes
2667      Psilocybe cyanescens      Eukaryota;Fungi;Basidiomycetes
1072      Boehmeria nivea      Eukaryota;Plants;Land Plants
1904      Eutrema heterophyllum      Eukaryota;Plants;Land Plants
1903      Eutrema yunnanense      Eukaryota;Plants;Land Plants
...

```

3119	Taenia asiatica	Eukaryota;Animals;Flatworms
3156	Wuchereria bancrofti	Eukaryota;Animals;Roundworms
3209	Botrytis cinerea T4	Eukaryota;Fungi;Ascomycetes
3357	Caenorhabditis elegans	Eukaryota;Animals;Roundworms
4976	Saccharomyces cerevisiae EC1118	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level	\
2685	8L	SAMN08475542	PRJNA433296	GCA_002938945.1	Scaffold	
2667	2631	SAMN07169033	PRJNA387735	GCA_002938375.1	Scaffold	
1072	NaN	SAMN07982878	PRJNA417329	GCA_002937015.1	Scaffold	
1904	NaN	SAMN08200758	PRJNA419026	GCA_002933915.1	Scaffold	
1903	NaN	SAMN08200759	PRJNA419026	GCA_002933935.1	Scaffold	
...	
3119	NaN	SAMEA1463525	PRJEB532	GCA_000951535.1	Scaffold	
3156	NaN	SAMEA1686771	PRJEB536	GCA_000951595.1	Scaffold	
3209	T4	NaN	PRJNA64593	GCA_000227075.1	Scaffold	
3357	N2	SAMEA1920834	PRJEB4373	GCA_000939815.1	Scaffold	
4976	Lalvin EC1118	SAMEA2272624	PRJEA37863	GCA_000218975.1	Scaffold	

	Size(Mb)	GC%	Replicons	WGS	Scaffolds	CDS	\
2685	40.7994	40.5	NaN	PTPU01	618	0.0	
2667	47.4654	46.1	NaN	NHYD01	3976	15936.0	
1072	344.6170	32.7	NaN	PHNS01	12775	0.0	
1904	348.9710	37.0	NaN	PKMM01	57686	0.0	
1903	415.3640	35.4	NaN	PKML01	78020	0.0	
...	
3119	135.9780	43.0	NaN	NaN	24137	0.0	
3156	76.9915	28.8	NaN	NaN	13520	0.0	
3209	39.5114	43.5	NaN	NaN	118	16353.0	
3357	101.1690	35.8	NaN	NaN	3806	0.0	
4976	11.6595	38.3	NaN	NaN	31	6017.0	

	Release Date	\
2685	2018-02-22 00:00:00+00:00	
2667	2018-02-22 00:00:00+00:00	
1072	2018-02-21 00:00:00+00:00	
1904	2018-02-20 00:00:00+00:00	
1903	2018-02-20 00:00:00+00:00	
...	...	
3119	NaN	
3156	NaN	
3209	NaN	
3357	NaN	
4976	NaN	

	GenBank	FTP	RefSeq	FTP
2685	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...			NaN

```

2667 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002... NaN
1072 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002... NaN
1904 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002... NaN
1903 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002... NaN
...
3119 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN
3156 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN
3209 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN
3357 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN
4976 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN

```

[5131 rows x 16 columns]

```
[133]: data.sort_values(by="Release Date",ascending=True)
```

```

[133]:
#Organism Name      Organism Groups \
88      Equus caballus      Eukaryota;Animals;Mammals
17  Leishmania major strain Friedlin  Eukaryota;Protists;Kinetoplasts
26      Plasmodium falciparum 3D7  Eukaryota;Protists;Apicomplexans
11      Saccharomyces cerevisiae S288C      Eukaryota;Fungi;Ascomycetes
33      Encephalitozoon cuniculi GB-M1      Eukaryota;Fungi;Other Fungi
...
3119      Taenia asiatica      Eukaryota;Animals;Flatworms
3156      Wuchereria bancrofti      Eukaryota;Animals;Roundworms
3209      Botrytis cinerea T4      Eukaryota;Fungi;Ascomycetes
3357      Caenorhabditis elegans      Eukaryota;Animals;Roundworms
4976      Saccharomyces cerevisiae EC1118      Eukaryota;Fungi;Ascomycetes

Strain      BioSample      BioProject      Assembly      Level \
88      Twilight      SAMN02953672      PRJNA421018      GCA_002863925.1      Chromosome
17      Friedlin      SAMEA3138173      PRJNA10724      GCA_000002725.2      Complete
26      3D7      NaN      PRJNA430688      GCA_000002765.1      Chromosome
11      S288C      NaN      PRJNA43747      GCA_000146045.2      Complete
33      GB-M1      SAMEA3138295      PRJNA13833      GCA_000091225.1      Chromosome
...
3119      NaN      SAMEA1463525      PRJEB532      GCA_000951535.1      Scaffold
3156      NaN      SAMEA1686771      PRJEB536      GCA_000951595.1      Scaffold
3209      T4      NaN      PRJNA64593      GCA_000227075.1      Scaffold
3357      N2      SAMEA1920834      PRJEB4373      GCA_000939815.1      Scaffold
4976      Lalvin EC1118      SAMEA2272624      PRJEA37863      GCA_000218975.1      Scaffold

Size(Mb)      GC%      Replicons \
88      2506.97000      42.0000      mitochondrion MT:NC_001640.1/X79547.1
17      32.85510      59.7114      chromosome 1:NC_001905.3/FR796397.1; chromosom...
26      23.27030      19.3630      chromosome 1:NC_004325.1/AL844501.1; chromosom...
11      12.15710      38.1556      chromosome I:NC_001133.9/BK006935.2; chromosom...
33      2.49752      47.3005      chromosome I:NC_003242.2/AL391737.2; chromosom...

```

...
3119	135.97800	43.0000	NaN
3156	76.99150	28.8000	NaN
3209	39.51140	43.5000	NaN
3357	101.16900	35.8000	NaN
4976	11.65950	38.3000	NaN

	WGS	Scaffolds	CDS	Release Date \
88	PJAA01	4701	13.0	1994-11-22 00:00:00+00:00
17	NaN	36	8316.0	1998-06-29 00:00:00+00:00
26	NaN	15	5339.0	1998-12-17 00:00:00+00:00
11	NaN	17	6002.0	1999-11-08 00:00:00+00:00
33	NaN	11	1996.0	2001-02-05 00:00:00+00:00

...
3119	NaN	24137	0.0	NaT
3156	NaN	13520	0.0	NaT
3209	NaN	118	16353.0	NaT
3357	NaN	3806	0.0	NaT
4976	NaN	31	6017.0	NaT

	GenBank FTP \
88	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
17	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
26	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
11	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
33	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
3119	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3156	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3209	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3357	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4976	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...

	RefSeq FTP
88	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002...
17	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
26	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
11	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
33	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
3119	NaN
3156	NaN
3209	NaN
3357	NaN
4976	NaN

[5131 rows x 16 columns]

```
[134]: data.sort_values(by=["Release Date","CDS"])
```

```
[134]:
```

	#Organism Name	Organism Groups \
88	Equus caballus	Eukaryota;Animals;Mammals
17	Leishmania major strain Friedlin	Eukaryota;Protists;Kinetoplasts
26	Plasmodium falciparum 3D7	Eukaryota;Protists;Apicomplexans
11	Saccharomyces cerevisiae S288C	Eukaryota;Fungi;Ascomycetes
33	Encephalitozoon cuniculi GB-M1	Eukaryota;Fungi;Other Fungi
...
2863	Lichtheimia ramosa	Eukaryota;Fungi;Other Fungi
718	Leptosphaeria maculans JN3	Eukaryota;Fungi;Ascomycetes
1502	Penicillium rubens Wisconsin 54-1255	Eukaryota;Fungi;Ascomycetes
3209	Botrytis cinerea T4	Eukaryota;Fungi;Ascomycetes
638	Apteryx australis mantelli	Eukaryota;Animals;Birds

	Strain	BioSample	BioProject	Assembly \
88	Twilight	SAMN02953672	PRJNA421018	GCA_002863925.1
17	Friedlin	SAMEA3138173	PRJNA10724	GCA_000002725.2
26	3D7	NaN	PRJNA430688	GCA_000002765.1
11	S288C	NaN	PRJNA43747	GCA_000146045.2
33	GB-M1	SAMEA3138295	PRJNA13833	GCA_000091225.1
...
2863	JMRC FSU:6197	SAMEA2519541	PRJEB6289	GCA_000945115.1
718	v23.1.3	NaN	PRJNA63129	GCA_000230375.1
1502	Wisconsin 54-1255	SAMEA2272345	PRJEA27927	GCA_000226395.1
3209	T4	NaN	PRJNA64593	GCA_000227075.1
638	NaN	SAMEA2560515	PRJNA294519	GCA_001039765.1

	Level	Size(Mb)	GC% \
88	Chromosome	2506.97000	42.0000
17	Complete	32.85510	59.7114
26	Chromosome	23.27030	19.3630
11	Complete	12.15710	38.1556
33	Chromosome	2.49752	47.3005
...
2863	Scaffold	30.70900	41.2000
718	Scaffold	45.12460	45.3000
1502	Scaffold	32.22370	49.0000
3209	Scaffold	39.51140	43.5000
638	Scaffold	1523.97000	42.5000

	Replicons	WGS	Scaffolds \
88	mitochondrion MT:NC_001640.1/X79547.1	PJAA01	4701
17	chromosome 1:NC_001905.3/FR796397.1; chromosom...	NaN	36
26	chromosome 1:NC_004325.1/AL844501.1; chromosom...	NaN	15
11	chromosome I:NC_001133.9/BK006935.2; chromosom...	NaN	17
33	chromosome I:NC_003242.2/AL391737.2; chromosom...	NaN	11

...
2863	NaN	NaN	74
718	NaN	NaN	76
1502	NaN	NaN	49
3209	NaN	NaN	118
638	NaN	NaN	24720

	CDS	Release Date \
88	13.0	1994-11-22 00:00:00+00:00
17	8316.0	1998-06-29 00:00:00+00:00
26	5339.0	1998-12-17 00:00:00+00:00
11	6002.0	1999-11-08 00:00:00+00:00
33	1996.0	2001-02-05 00:00:00+00:00
...
2863	11542.0	NaT
718	12469.0	NaT
1502	12791.0	NaT
3209	16353.0	NaT
638	22793.0	NaT

	GenBank FTP \
88	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
17	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
26	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
11	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
33	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
2863	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
718	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1502	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3209	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
638	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001...

	RefSeq FTP
88	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002...
17	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
26	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
11	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
33	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
2863	NaN
718	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1502	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3209	NaN
638	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001...

[5131 rows x 16 columns]

```
[135]: data.sort_values(by=["Release Date","CDS"],ascending=False)
```

```
[135]:
```

	#Organism Name	Organism Groups	Strain \
2667	Psilocybe cyanescens	Eukaryota;Fungi;Basidiomycetes	2631
2685	Monilinia laxa	Eukaryota;Fungi;Ascomycetes	8L
1072	Boehmeria nivea	Eukaryota;Plants;Land Plants	NaN
2666	Cercospora berteroae	Eukaryota;Fungi;Ascomycetes	CBS538.71
1903	Eutrema yunnanense	Eukaryota;Plants;Land Plants	NaN
...
3081	Nippostrongylus brasiliensis	Eukaryota;Animals;Roundworms	NaN
3111	Toxocara canis	Eukaryota;Animals;Roundworms	NaN
3119	Taenia asiatica	Eukaryota;Animals;Flatworms	NaN
3156	Wuchereria bancrofti	Eukaryota;Animals;Roundworms	NaN
3357	Caenorhabditis elegans	Eukaryota;Animals;Roundworms	N2

	BioSample	BioProject	Assembly	Level	Size(Mb)	GC% \
2667	SAMN07169033	PRJNA387735	GCA_002938375.1	Scaffold	47.4654	46.1
2685	SAMN08475542	PRJNA433296	GCA_002938945.1	Scaffold	40.7994	40.5
1072	SAMN07982878	PRJNA417329	GCA_002937015.1	Scaffold	344.6170	32.7
2666	SAMN08288628	PRJNA270309	GCA_002933655.1	Scaffold	33.8928	52.2
1903	SAMN08200759	PRJNA419026	GCA_002933935.1	Scaffold	415.3640	35.4
...
3081	SAMEA1018098	PRJEB511	GCA_000951335.1	Scaffold	294.4000	43.0
3111	SAMEA1034752	PRJEB533	GCA_000951555.1	Scaffold	299.7880	40.5
3119	SAMEA1463525	PRJEB532	GCA_000951535.1	Scaffold	135.9780	43.0
3156	SAMEA1686771	PRJEB536	GCA_000951595.1	Scaffold	76.9915	28.8
3357	SAMEA1920834	PRJEB4373	GCA_000939815.1	Scaffold	101.1690	35.8

	Replicons	WGS	Scaffolds	CDS	Release Date \
2667	NaN	NHYD01	3976	15936.0	2018-02-22 00:00:00+00:00
2685	NaN	PTPU01	618	0.0	2018-02-22 00:00:00+00:00
1072	NaN	PHNS01	12775	0.0	2018-02-21 00:00:00+00:00
2666	NaN	PNEN01	1801	11903.0	2018-02-20 00:00:00+00:00
1903	NaN	PKML01	78020	0.0	2018-02-20 00:00:00+00:00
...
3081	NaN	NaN	29375	0.0	NaN
3111	NaN	NaN	29185	0.0	NaN
3119	NaN	NaN	24137	0.0	NaN
3156	NaN	NaN	13520	0.0	NaN
3357	NaN	NaN	3806	0.0	NaN

	GenBank	FTP	RefSeq	FTP
2667	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...		NaN	
2685	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...		NaN	
1072	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...		NaN	
2666	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...		NaN	
1903	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...		NaN	

```

...
3081 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN
3111 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN
3119 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN
3156 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN
3357 ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000... NaN

```

[5131 rows x 16 columns]

```
[136]: data.sort_values(by=["Release Date","CDS"],ascending= [True, False])
```

```

[136]:
#Organism Name      Organism Groups \
88      Equus caballus      Eukaryota;Animals;Mammals
17  Leishmania major strain Friedlin  Eukaryota;Protists;Kinetoplasts
26      Plasmodium falciparum 3D7  Eukaryota;Protists;Apicomplexans
11      Saccharomyces cerevisiae S288C      Eukaryota;Fungi;Ascomycetes
33      Encephalitozoon cuniculi GB-M1      Eukaryota;Fungi;Other Fungi
...
3081      Nippostrongylus brasiliensis      Eukaryota;Animals;Roundworms
3111      Toxocara canis      Eukaryota;Animals;Roundworms
3119      Taenia asiatica      Eukaryota;Animals;Flatworms
3156      Wuchereria bancrofti      Eukaryota;Animals;Roundworms
3357      Caenorhabditis elegans      Eukaryota;Animals;Roundworms

Strain      BioSample      BioProject      Assembly      Level \
88  Twilight      SAMN02953672      PRJNA421018      GCA_002863925.1      Chromosome
17  Friedlin      SAMEA3138173      PRJNA10724      GCA_000002725.2      Complete
26      3D7      NaN      PRJNA430688      GCA_000002765.1      Chromosome
11      S288C      NaN      PRJNA43747      GCA_000146045.2      Complete
33      GB-M1      SAMEA3138295      PRJNA13833      GCA_000091225.1      Chromosome
...
3081      NaN      SAMEA1018098      PRJEB511      GCA_000951335.1      Scaffold
3111      NaN      SAMEA1034752      PRJEB533      GCA_000951555.1      Scaffold
3119      NaN      SAMEA1463525      PRJEB532      GCA_000951535.1      Scaffold
3156      NaN      SAMEA1686771      PRJEB536      GCA_000951595.1      Scaffold
3357      N2      SAMEA1920834      PRJEB4373      GCA_000939815.1      Scaffold

Size(Mb)      GC%      Replicons \
88  2506.97000      42.0000      mitochondrion MT:NC_001640.1/X79547.1
17  32.85510      59.7114      chromosome 1:NC_001905.3/FR796397.1; chromosom...
26  23.27030      19.3630      chromosome 1:NC_004325.1/AL844501.1; chromosom...
11  12.15710      38.1556      chromosome I:NC_001133.9/BK006935.2; chromosom...
33  2.49752      47.3005      chromosome I:NC_003242.2/AL391737.2; chromosom...
...
3081      294.40000      43.0000      NaN
3111      299.78800      40.5000      NaN
3119      135.97800      43.0000      NaN

```

3156	76.99150	28.8000	NaN
3357	101.16900	35.8000	NaN

	WGS	Scaffolds	CDS	Release Date \
88	PJAA01	4701	13.0	1994-11-22 00:00:00+00:00
17	NaN	36	8316.0	1998-06-29 00:00:00+00:00
26	NaN	15	5339.0	1998-12-17 00:00:00+00:00
11	NaN	17	6002.0	1999-11-08 00:00:00+00:00
33	NaN	11	1996.0	2001-02-05 00:00:00+00:00
...
3081	NaN	29375	0.0	NaT
3111	NaN	29185	0.0	NaT
3119	NaN	24137	0.0	NaT
3156	NaN	13520	0.0	NaT
3357	NaN	3806	0.0	NaT

	GenBank FTP \
88	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002...
17	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
26	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
11	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
33	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
3081	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3111	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3119	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3156	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3357	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...

	RefSeq FTP
88	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/002...
17	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
26	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
11	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
33	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
3081	NaN
3111	NaN
3119	NaN
3156	NaN
3357	NaN

[5131 rows x 16 columns]

```
[137]: data["Release Date"] = pd.to_datetime(data["Release Date"], errors='coerce')
```

```
[138]: # aggfuncs + agg() & cumsum()

# Assuming 'data' is your DataFrame containing the relevant data
agg_data = data.agg({
    'Release Date': ['min', 'max', 'mean', 'median', 'std', 'count'],
    'CDS': 'sum',
    'BioSample': lambda x: x.mode()[0]
})
```

```
[139]: # Assuming 'data' is your DataFrame containing the relevant data
data['Release Date'] = data['Release Date'].astype('int64')
```

```
[140]: cumsum_data = data[['Release Date', 'CDS']].cumsum()
```

```
[141]: cumsum_data
```

```
[141]:
```

	Release Date	CDS
0	1366329600000000000	38554.00
1	2363990400000000000	86904.00
2	3677097600000000000	144565.00
3	4939747200000000000	216090.00
4	6231686400000000000	252100.00
...
5126	1103665579565318144	31790367.83
5127	2596398379565318144	31790367.83
5128	4089131179565318144	31790367.83
5129	5581863979565318144	31790367.83
5130	7074596779565318144	31790367.83

[5131 rows x 2 columns]

```
[142]: # aggfuncs + agg() & cumsum()data.Name.mode()[0]

print("Min Release Date:", data['Release Date'].min())
print("Max Release Date:", data['Release Date'].max())
print("Mean Release Date:", data['Release Date'].mean())
print("Median Release Date:", data['Release Date'].median())
print("std Release Date:", data['Release Date'].std())
print("Count number of rows in Release Date:", data['Release Date'].count())
print("Sum of CDS Release Date:", data['CDS'].sum())
print("Mode BioSample", data['BioSample'].mode()[0])
```

```
Min Release Date: -9223372036854775808
Max Release Date: 1519257600000000000
Mean Release Date: 1.3100154822860733e+18
Median Release Date: 1.4487552e+18
std Release Date: 1.0901304214295648e+18
Count number of rows in Release Date: 5131
```

```
Sum of CDS Release Date: 31790367.83
Mode BioSample SAMN02743421
```

```
[143]: data.BioSample.mode()[0]
```

```
[143]: 'SAMN02743421'
```

```
[144]: data["CDS"].quantile(0.8)
```

```
[144]: 10994.0
```

```
[145]: data["CDS"].quantile(0.99)
```

```
[145]: 53452.199999999924
```

```
[146]: data.CDS.agg(["min", "max", "mean", "median", "std", "sum", "count"])
```

```
[146]: min          0.000000e+00
      max          1.234670e+05
      mean         6.195745e+03
      median       0.000000e+00
      std          1.191627e+04
      sum          3.179037e+07
      count        5.131000e+03
      Name: CDS, dtype: float64
```

```
[147]: data.CDS.cumsum()
```

```
[147]: 0          38554.00
      1          86904.00
      2         144565.00
      3         216090.00
      4         252100.00
      ...
      5126      31790367.83
      5127      31790367.83
      5128      31790367.83
      5129      31790367.83
      5130      31790367.83
      Name: CDS, Length: 5131, dtype: float64
```

```
[148]: # qut and qcut
      # cut
      data["Release Date"] = pd.to_datetime(data["Release Date"], errors='coerce')
```

```
[149]: Release_Date_bins = pd.cut(data["Release Date"], bins=5)
```

```
[150]: Release_Date_counts = data["Release Date"].value_counts()
```

```
[151]: cumulative_counts = Release_Date_counts.cumsum()
```

```
[152]: Release_Date_bins = pd.cut(data["Release Date"], bins=5)
```

```
[153]: Release_Date_bins
```

```
[153]: 0      (2008-11-03 19:12:00, 2013-06-29 09:36:00]
      1      (1999-07-17 14:24:00, 2004-03-11 04:48:00]
      2      (2008-11-03 19:12:00, 2013-06-29 09:36:00]
      3      (2008-11-03 19:12:00, 2013-06-29 09:36:00]
      4      (2008-11-03 19:12:00, 2013-06-29 09:36:00]
      ...
      5126     (2013-06-29 09:36:00, 2018-02-22]
      5127     (2013-06-29 09:36:00, 2018-02-22]
      5128     (2013-06-29 09:36:00, 2018-02-22]
      5129     (2013-06-29 09:36:00, 2018-02-22]
      5130     (2013-06-29 09:36:00, 2018-02-22]
      Name: Release Date, Length: 5131, dtype: category
      Categories (5, interval[datetime64[ns], right]): [(1994-11-13 12:10:04.800000,
      1999-07-17 14:24:00] < (1999-07-17 14:24:00, 2004-03-11 04:48:00] < (2004-03-11
      04:48:00, 2008-11-03 19:12:00] < (2008-11-03 19:12:00, 2013-06-29 09:36:00] <
      (2013-06-29 09:36:00, 2018-02-22]]
```

```
[154]: Release_Date_bins = pd.cut(data["Release Date"], bins=pd.
      ↪to_datetime(['1999-01-01', '2004-01-01', '2008-01-01', '2013-01-01',
      ↪'2018-01-01']))
```

```
[155]: Release_Date_bins = pd.cut(
      data["Release Date"],
      bins=pd.to_datetime(['1999-01-01', '2004-01-01', '2008-01-01',
      ↪'2013-01-01', '2018-01-01'])
      )
```

```
[156]: Release_Date_bins
```

```
[156]: 0      (2013-01-01, 2018-01-01]
      1      (1999-01-01, 2004-01-01]
      2      (2008-01-01, 2013-01-01]
      3      (2008-01-01, 2013-01-01]
      4      (2008-01-01, 2013-01-01]
      ...
      5126     (2013-01-01, 2018-01-01]
      5127     (2013-01-01, 2018-01-01]
      5128     (2013-01-01, 2018-01-01]
      5129     (2013-01-01, 2018-01-01]
      5130     (2013-01-01, 2018-01-01]
      Name: Release Date, Length: 5131, dtype: category
```

```
Categories (4, interval[datetime64[ns], right]): [(1999-01-01, 2004-01-01] <
(2004-01-01, 2008-01-01] < (2008-01-01, 2013-01-01] < (2013-01-01, 2018-01-01]]
```

```
[157]: data["Release Date"] = pd.to_datetime(data["Release Date"], errors='coerce')
```

```
[158]: # qcut
```

```
Release_Date_bins = pd.qcut(data["Release Date"], q=4)
```

```
[159]: Release_Date_bins
```

```
[159]: 0      (1994-11-21 23:59:59.999999999, 2014-02-19]
      1      (1994-11-21 23:59:59.999999999, 2014-02-19]
      2      (1994-11-21 23:59:59.999999999, 2014-02-19]
      3      (1994-11-21 23:59:59.999999999, 2014-02-19]
      4      (1994-11-21 23:59:59.999999999, 2014-02-19]
      ...
      5126      (2016-12-23, 2018-02-22]
      5127      (2016-12-23, 2018-02-22]
      5128      (2016-12-23, 2018-02-22]
      5129      (2016-12-23, 2018-02-22]
      5130      (2016-12-23, 2018-02-22]
      Name: Release Date, Length: 5131, dtype: category
      Categories (4, interval[datetime64[ns], right]): [(1994-11-21
      23:59:59.999999999, 2014-02-19] < (2014-02-19, 2015-11-29] < (2015-11-29,
      2016-12-23] < (2016-12-23, 2018-02-22]]
```

```
[160]: Release_Date_bins = pd.qcut(data["Release Date"], q=4).value_counts()
```

```
[161]: Release_Date_bins
```

```
[161]: Release Date
      (2014-02-19, 2015-11-29]      1303
      (1994-11-21 23:59:59.999999999, 2014-02-19]      1272
      (2015-11-29, 2016-12-23]      1251
      (2016-12-23, 2018-02-22]      1251
      Name: count, dtype: int64
```

```
[162]: Scaffolds_bins = pd.qcut(data["Scaffolds"], q=4, duplicates='drop')
```

```
[163]: Scaffolds_bins
```

```
[163]: 0      (1591.0, 10342.5]
      1      (0.999, 196.0]
      2      (1591.0, 10342.5]
      3      (196.0, 1591.0]
      4      (1591.0, 10342.5]
      ...
```



```

5126      (1591.0, 10342.5]
5127      (196.0, 1591.0]
5128      (1591.0, 10342.5]
5129      (196.0, 1591.0]
5130      (196.0, 1591.0]
Name: Scaffolds, Length: 5131, dtype: category
Categories (4, interval[float64, right]): [(0.999, 196.0] < (196.0, 1591.0] <
(1591.0, 10342.5] < (10342.5, 6870110.0]]

```

```
[164]: Scaffolds_bins = pd.qcut(data["Scaffolds"], q=4,).value_counts()
```

```
[165]: Scaffolds_bins
```

```

[165]: Scaffolds
(0.999, 196.0]      1284
(196.0, 1591.0]     1283
(10342.5, 6870110.0] 1283
(1591.0, 10342.5]   1281
Name: count, dtype: int64

```

```
[166]: Scaffolds_bins = pd.qcut(data["Scaffolds"], q=4,precision=50).value_counts()
```

```
[167]: Scaffolds_bins
```

```

[167]: Scaffolds
(1.0, 196.0]      1284
(196.0, 1591.0]   1283
(10342.5, 6870110.0] 1283
(1591.0, 10342.5]  1281
Name: count, dtype: int64

```

```
[168]: Scaffolds_bins = pd.qcut(data["Scaffolds"], q=[0,0.5,0.75,0.95,1],precision=50)
```

```
[169]: Scaffolds_bins
```

```

[169]: 0      (1591.0, 10342.5]
1      (1.0, 1591.0]
2      (1591.0, 10342.5]
3      (1.0, 1591.0]
4      (1591.0, 10342.5]
...
5126   (1591.0, 10342.5]
5127   (1.0, 1591.0]
5128   (1591.0, 10342.5]
5129   (1.0, 1591.0]
5130   (1.0, 1591.0]
Name: Scaffolds, Length: 5131, dtype: category
Categories (4, interval[float64, right]): [(1.0, 1591.0] < (1591.0, 10342.5] <

```

```
(10342.5, 183821.0] < (183821.0, 6870110.0]]
```

```
[170]: scaffolds_bins = pd.qcut(data["Scaffolds"], q=[0, 0.5, 0.75, 0.95, 1])
```

```
[171]: scaffolds_bins
```

```
[171]: 0      (1591.0, 10342.5]
      1      (0.999, 1591.0]
      2      (1591.0, 10342.5]
      3      (0.999, 1591.0]
      4      (1591.0, 10342.5]
      ...
     5126    (1591.0, 10342.5]
     5127    (0.999, 1591.0]
     5128    (1591.0, 10342.5]
     5129    (0.999, 1591.0]
     5130    (0.999, 1591.0]
     Name: Scaffolds, Length: 5131, dtype: category
     Categories (4, interval[float64, right]): [(0.999, 1591.0] < (1591.0, 10342.5] <
     (10342.5, 183821.0] < (183821.0, 6870110.0]]
```

```
[172]: data_binned = pd.DataFrame(scaffolds_bins, columns=["Scaffolds_bin"])
```

```
[173]: data_combined = pd.concat([data, data_binned], axis=1)
```

```
[174]: print(data_combined)
```

	#Organism Name	Organism Groups \
0	Emiliania huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level \
0	CCMP1516	SAMN02744062	PRJNA77753	GCA_000372725.1	Scaffold
1	NaN	SAMN03081427	PRJNA10719	GCA_000001735.1	Chromosome
2	A17	SAMN02299339	PRJNA10791	GCA_000219495.2	Chromosome
3	NaN	SAMN00002965	PRJNA19861	GCA_000004515.3	Chromosome
4	NaN	SAMN02981290	PRJNA119	GCA_000188115.2	Chromosome
...
5126	NaN	SAMEA3110475	PRJEB7675	GCA_900178465.1	Scaffold

5127	ZP848	SAMEA3110493	PRJEB7675	GCA_900178475.1	Scaffold
5128	VKMY373	SAMEA3110473	PRJEB7675	GCA_900178485.1	Scaffold
5129	245	SAMEA3110476	PRJEB7675	GCA_900178495.1	Scaffold
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	GCA_900178505.1	Scaffold

	Size(Mb)	GC%	Replicons \
0	167.6760	64.5000	NaN
1	119.6680	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...
2	412.9240	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...
3	978.9720	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...
4	823.7860	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...
...
5126	12.0392	38.1000	NaN
5127	11.7086	38.1000	NaN
5128	11.8194	38.0000	NaN
5129	11.7081	38.1000	NaN
5130	11.8189	38.1000	NaN

	WGS	Scaffolds	CDS	Release Date \
0	AHAL01	7795	38554.0	2013-04-19
1	NaN	7	48350.0	2001-08-13
2	APNO01	2187	57661.0	2011-08-12
3	ACUP02	1191	71525.0	2010-01-05
4	AEKE02	3224	36010.0	2010-12-10
...
5126	FXDE01	1925	0.0	2017-04-21
5127	FXCK01	590	0.0	2017-04-21
5128	FXC001	2013	0.0	2017-04-21
5129	FXCP01	1028	0.0	2017-04-21
5130	FXCG01	799	0.0	2017-04-21

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

	RefSeq FTP Scaffolds_bin
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000... NaN
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000... NaN
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000... NaN

```

3      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...      NaN
4      ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...      NaN
...
5126      NaN      NaN
5127      NaN      NaN
5128      NaN      NaN
5129      NaN      NaN
5130      NaN      NaN

```

[5131 rows x 17 columns]

```
[176]: organismGroups = ["Emiliana huxleyi CCMP1516", "Arabidopsis thaliana",
    ↪ "Medicago truncatula", "Glycine max", "Saccharomyces cerevisiae"]
```

```
[177]: def categories(x):
    if x in organismGroups:
        return "organismGroups"
```

```
[178]: data["Scaffolds"]
```

```
[178]: 0      7795
1         7
2      2187
3      1191
4      3224
...
5126    1925
5127     590
5128    2013
5129    1028
5130     799
Name: Scaffolds, Length: 5131, dtype: int64
```

```
[179]: data["Scaffolds"].apply(categories)
```

```
[179]: 0      None
1      None
2      None
3      None
4      None
...
5126    None
5127    None
5128    None
5129    None
5130    None
Name: Scaffolds, Length: 5131, dtype: object
```

```
[180]: data["Scaffolds"].apply(categories).value_counts()
```

```
[180]: Series([], Name: count, dtype: int64)
```

```
[181]: import scipy.stats as stats
```

```
[182]: z_score = data.select_dtypes("float").apply(lambda x: stats.zscore(x))
```

```
[183]: z_score
```

```
[183]:
```

	Size(Mb)	GC%	CDS
0	-0.186909	2.122414	2.715734
1	-0.226714	-0.565356	3.537883
2	0.016434	-0.754870	4.319328
3	0.485763	-0.653414	5.482893
4	0.357093	-0.597773	2.502223
...
5126	-0.315953	-0.371931	-0.519991
5127	-0.316227	-0.371931	-0.519991
5128	-0.316135	-0.381379	-0.519991
5129	-0.316227	-0.371931	-0.519991
5130	-0.316136	-0.371931	-0.519991

```
[5131 rows x 3 columns]
```

```
[184]: z_score.mean()
```

```
[184]:
```

Size(Mb)	1.661764e-17
GC%	-1.107843e-17
CDS	2.215686e-17
dtype:	float64

```
[185]: z_score.std()
```

```
[185]:
```

Size(Mb)	1.000097
GC%	1.000097
CDS	1.000097
dtype:	float64

```
[186]: z_score
```

```
[186]:
```

	Size(Mb)	GC%	CDS
0	-0.186909	2.122414	2.715734
1	-0.226714	-0.565356	3.537883
2	0.016434	-0.754870	4.319328
3	0.485763	-0.653414	5.482893
4	0.357093	-0.597773	2.502223
...

```

5126 -0.315953 -0.371931 -0.519991
5127 -0.316227 -0.371931 -0.519991
5128 -0.316135 -0.381379 -0.519991
5129 -0.316227 -0.371931 -0.519991
5130 -0.316136 -0.371931 -0.519991

```

[5131 rows x 3 columns]

```

[187]: # corr()

corr_matrix = data.drop("Release Date", axis=1)

```

```

[188]: corr_matrix

```

```

[188]:
      #Organism Name      Organism Groups \
0      Emiliania huxleyi CCMP1516  Eukaryota;Protists;Other Protists
1      Arabidopsis thaliana      Eukaryota;Plants;Land Plants
2      Medicago truncatula      Eukaryota;Plants;Land Plants
3      Glycine max      Eukaryota;Plants;Land Plants
4      Solanum lycopersicum      Eukaryota;Plants;Land Plants
...
5126  Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5127  Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5128  Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5129  Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes
5130  Saccharomyces cerevisiae      Eukaryota;Fungi;Ascomycetes

      Strain      BioSample      BioProject      Assembly      Level \
0      CCMP1516      SAMN02744062      PRJNA77753      GCA_000372725.1      Scaffold
1      NaN      SAMN03081427      PRJNA10719      GCA_000001735.1      Chromosome
2      A17      SAMN02299339      PRJNA10791      GCA_000219495.2      Chromosome
3      NaN      SAMN00002965      PRJNA19861      GCA_000004515.3      Chromosome
4      NaN      SAMN02981290      PRJNA119      GCA_000188115.2      Chromosome
...
5126      NaN      SAMEA3110475      PRJEB7675      GCA_900178465.1      Scaffold
5127      ZP848      SAMEA3110493      PRJEB7675      GCA_900178475.1      Scaffold
5128      VKMY373      SAMEA3110473      PRJEB7675      GCA_900178485.1      Scaffold
5129      245      SAMEA3110476      PRJEB7675      GCA_900178495.1      Scaffold
5130      M2ON0800_1A      SAMEA3110463      PRJEB7675      GCA_900178505.1      Scaffold

      Size(Mb)      GC%      Replicons \
0      167.6760      64.5000      NaN
1      119.6680      36.0528      chromosome 1:NC_003070.9/CP002684.1; chromosom...
2      412.9240      34.0470      chromosome 1:NC_016407.2/CM001217.2; chromosom...
3      978.9720      35.1208      chromosome 1:NC_016088.2/CM000834.2; chromosom...
4      823.7860      35.7097      chromosome 1:NC_015438.2/CM001064.2; chromosom...
...

```

5126	12.0392	38.1000	NaN
5127	11.7086	38.1000	NaN
5128	11.8194	38.0000	NaN
5129	11.7081	38.1000	NaN
5130	11.8189	38.1000	NaN

	WGS	Scaffolds	CDS \
0	AHAL01	7795	38554.0
1	NaN	7	48350.0
2	APN001	2187	57661.0
3	ACUP02	1191	71525.0
4	AEKE02	3224	36010.0
...
5126	FXDE01	1925	0.0
5127	FXCK01	590	0.0
5128	FXC001	2013	0.0
5129	FXCP01	1028	0.0
5130	FXCG01	799	0.0

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

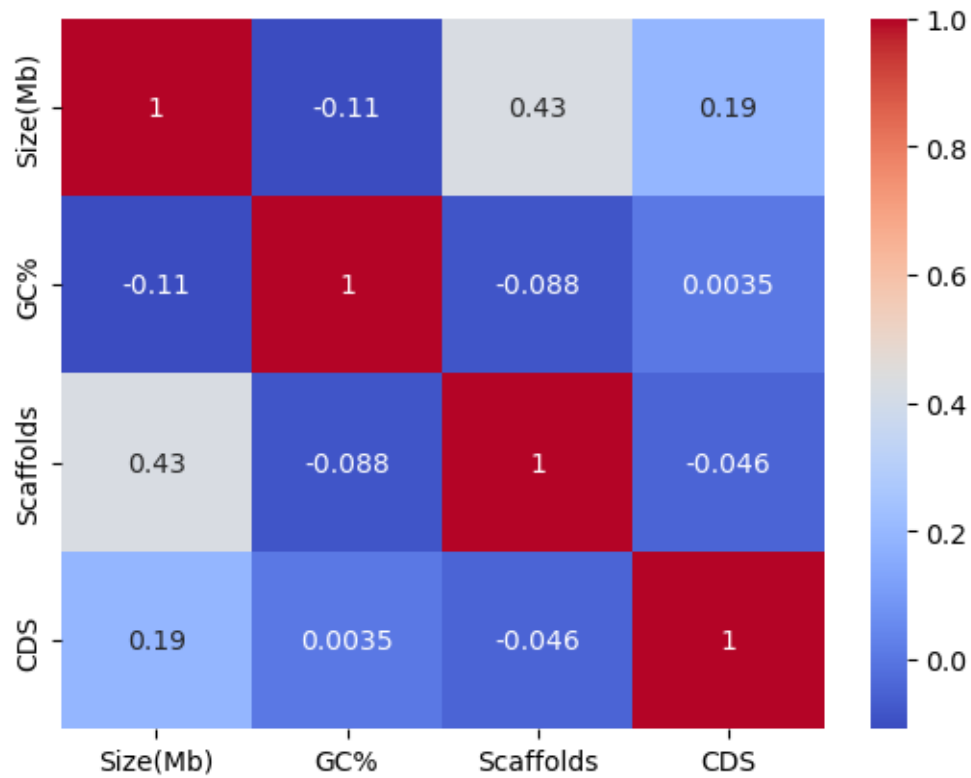
	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
5126	NaN
5127	NaN
5128	NaN
5129	NaN
5130	NaN

[5131 rows x 15 columns]

```
[189]: numeric_data = data.select_dtypes(include=['number'])
```

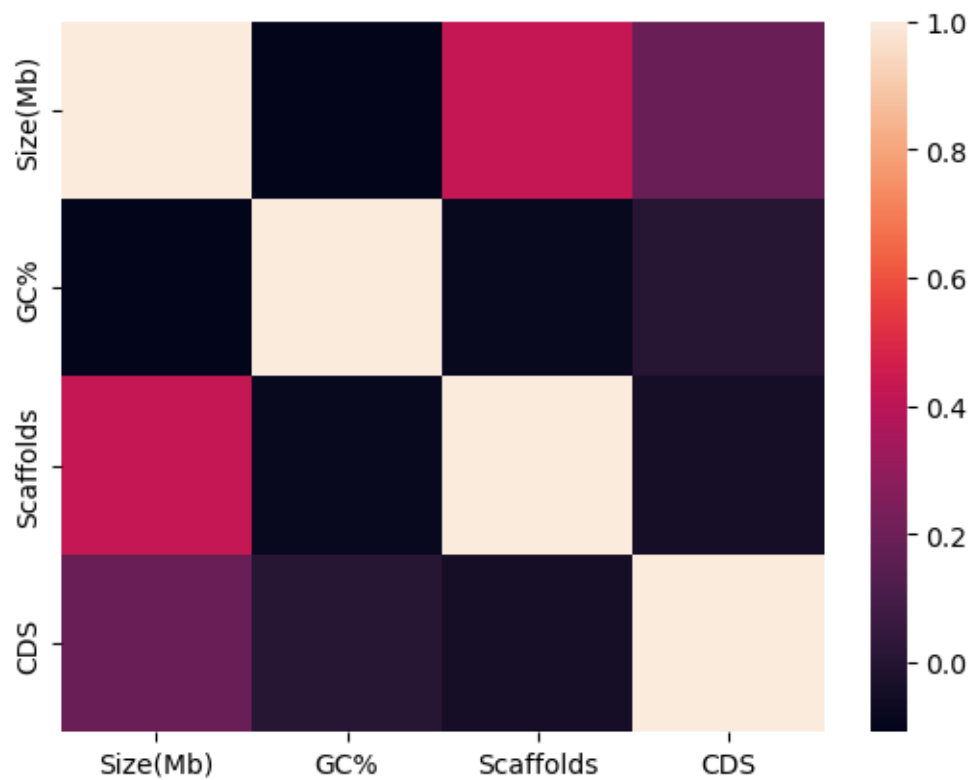
```
[190]: corr_matrix = numeric_data.corr()
```

```
[191]: sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')  
plt.show()
```



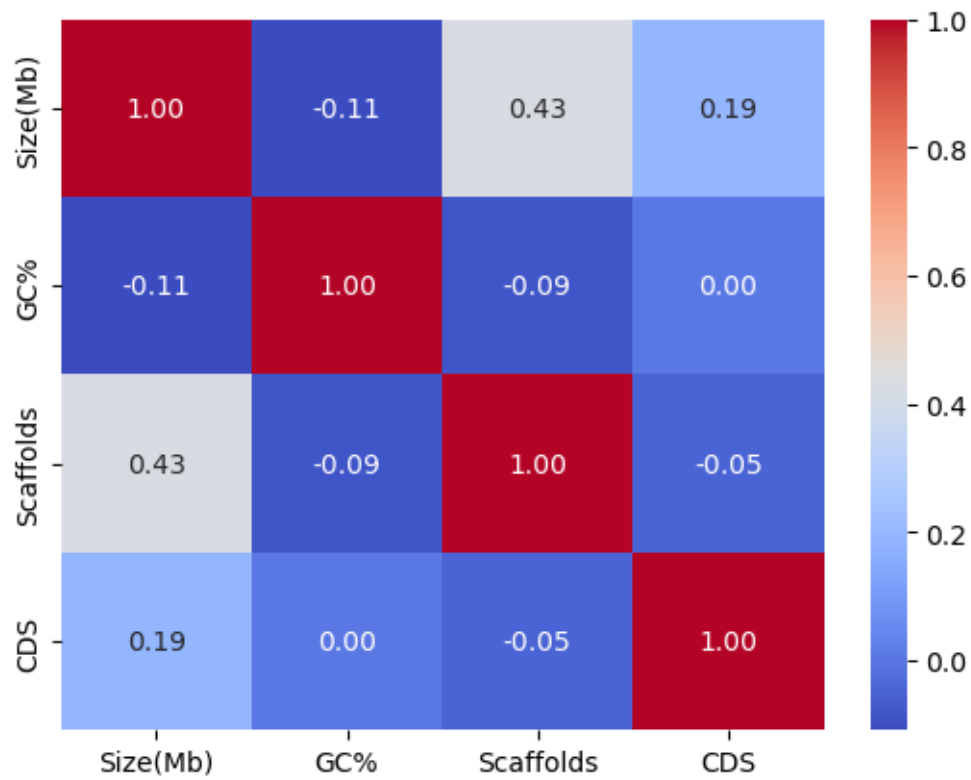
```
[192]: sns.heatmap(corr_matrix)
```

```
[192]: <Axes: >
```

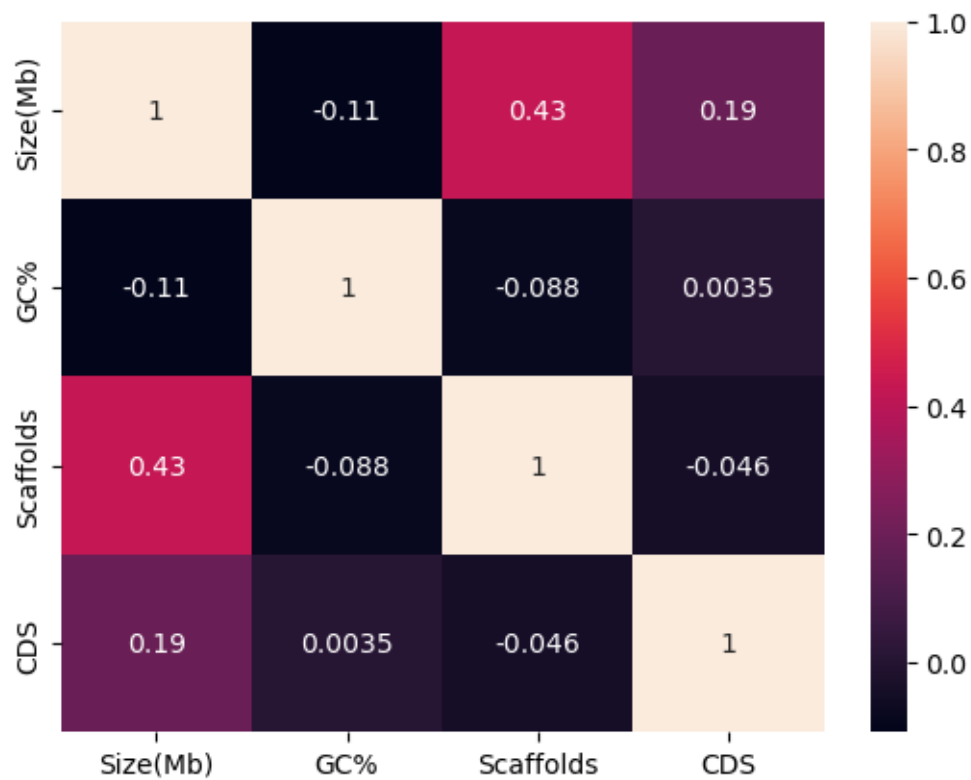
```
[193]: sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt='.2f')
```

```
[193]: <Axes: >
```



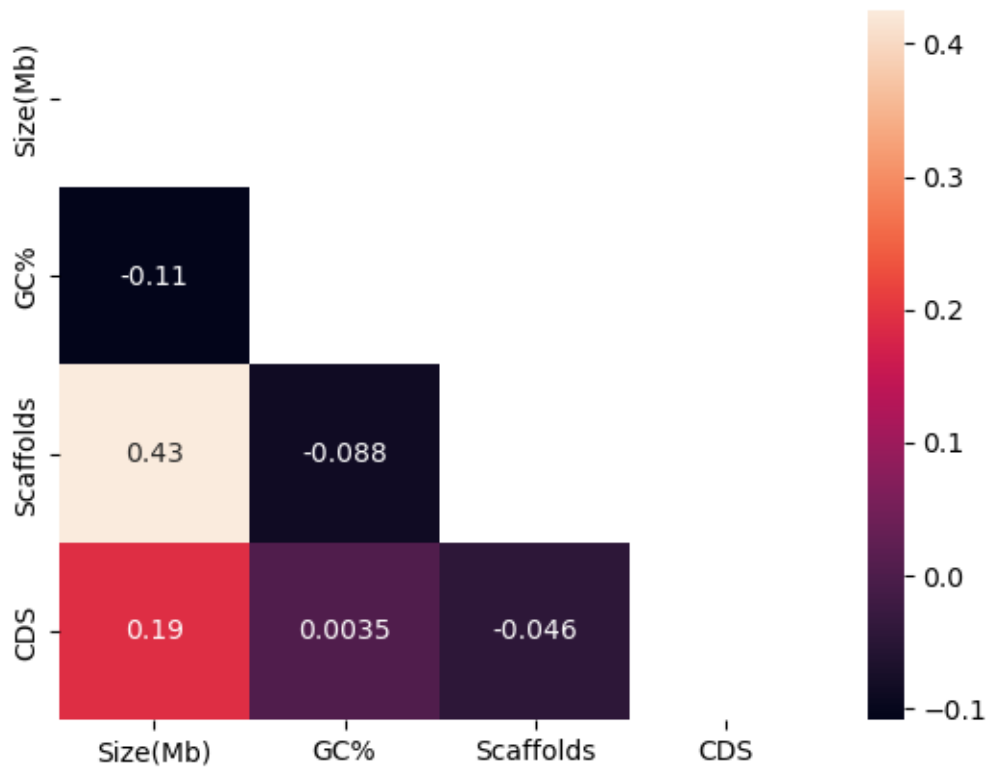
```
[194]: sns.heatmap(corr_matrix, annot=True)
```

```
[194]: <Axes: >
```



```
[195]: metrix = np.triu(corr_matrix)
sns.heatmap(corr_matrix, annot=True, mask= metrix)
```

```
[195]: <Axes: >
```



2.11 GroupBy , most important thing

```
[196]: data.groupby("Level")["CDS"].mean()
```

```
[196]: Level
      Chromosome    10028.494484
      Complete      4525.944444
      Contig        2200.587462
      Scaffold      6528.212239
      Name: CDS, dtype: float64
```

```
[197]: data.Level.unique()
```

```
[197]: array(['Scaffold', 'Chromosome', 'Contig', 'Complete'], dtype=object)
```

```
[198]: print(data["CDS"].mean())
```

```
6195.7450458000385
```

```
[199]: data["CDS"] = pd.to_numeric(data["CDS"])
      print(data["CDS"].mean())
```

```
6195.7450458000385
```

```
[200]: data["CDS"] = pd.to_numeric(data["CDS"], downcast="float")
print(data["CDS"].mean())
```

6195.745

```
[201]: print(data["CDS"].mean())
```

```
data.loc[data["Level"] == "Chromosome", "CDS"].mean()
```

6195.745

```
[201]: nan
```

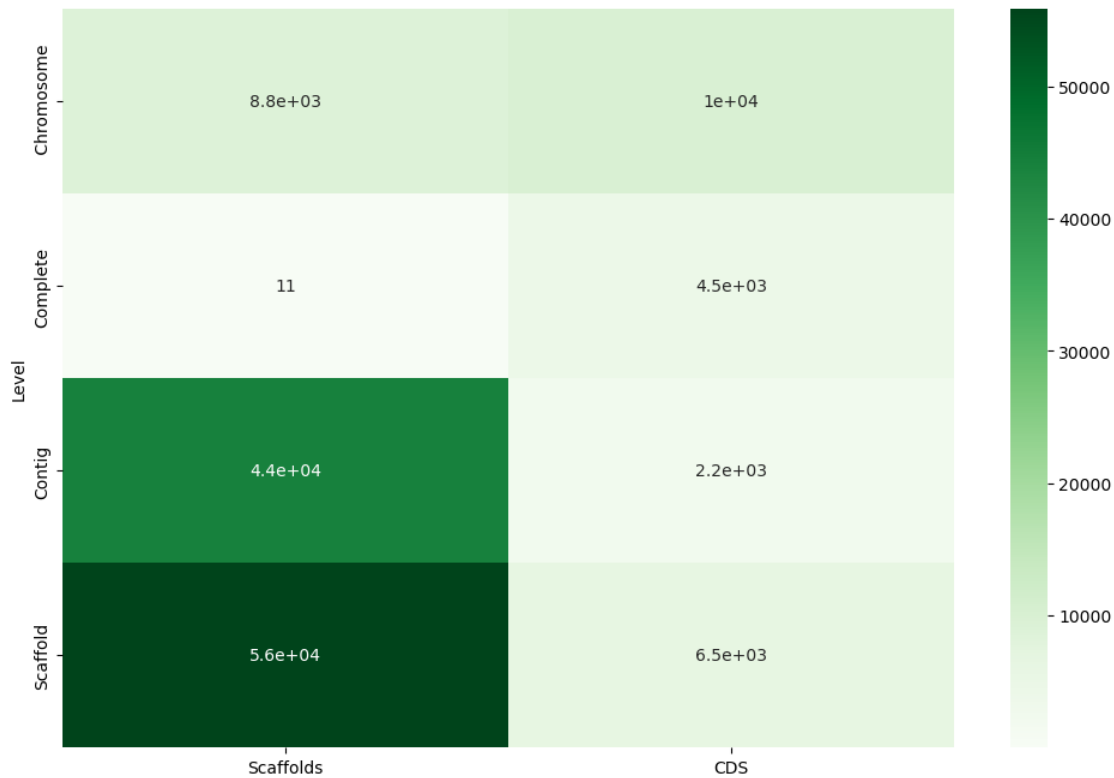
```
[202]: Level = data.groupby("Level")[["Scaffolds", "CDS"]].mean()
Level
```

```
[202]:
```

	Scaffolds	CDS
Level		
Chromosome	8847.941799	10028.495117
Complete	11.472222	4525.944336
Contig	44287.107179	2200.587402
Scaffold	55849.401194	6528.212402

```
[203]: plt.figure(figsize=(12,8))
sns.heatmap(Level, annot=True, cmap="Greens")
```

```
[203]: <Axes: ylabel='Level'>
```



```
[204]: Level = ["Scaffold", "Chromosome", "Contig", "Complete"]
#Organism Name = ["Eukaryota;Protists;Other Protists", "Eukaryota;Plants;Land
↳Plants", "Eukaryota;Fungi;Ascomycetes"]
```

```
[205]: data["Level"]
```

```
[205]: 0      Scaffold
1      Chromosome
2      Chromosome
3      Chromosome
4      Chromosome
...
5126   Scaffold
5127   Scaffold
5128   Scaffold
5129   Scaffold
5130   Scaffold
Name: Level, Length: 5131, dtype: object
```

```
[206]: data["Level"].apply(categories)
```

```
[206]: 0      None
        1      None
        2      None
        3      None
        4      None
        ...
        5126   None
        5127   None
        5128   None
        5129   None
        5130   None
        Name: Level, Length: 5131, dtype: object
```

```
[207]: data["Level"].apply(categories).value_counts()
```

```
[207]: Series([], Name: count, dtype: int64)
```

```
[208]: data["Assembly"] = data["Level"].apply(categories)
```

```
[209]: data
```

```
[209]:
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Assembly	Level	Size(Mb) \
0	CCMP1516	SAMN02744062	PRJNA77753	None	Scaffold	167.6760
1	NaN	SAMN03081427	PRJNA10719	None	Chromosome	119.6680
2	A17	SAMN02299339	PRJNA10791	None	Chromosome	412.9240
3	NaN	SAMN00002965	PRJNA19861	None	Chromosome	978.9720
4	NaN	SAMN02981290	PRJNA119	None	Chromosome	823.7860
...
5126	NaN	SAMEA3110475	PRJEB7675	None	Scaffold	12.0392
5127	ZP848	SAMEA3110493	PRJEB7675	None	Scaffold	11.7086
5128	VKMY373	SAMEA3110473	PRJEB7675	None	Scaffold	11.8194
5129	245	SAMEA3110476	PRJEB7675	None	Scaffold	11.7081
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	None	Scaffold	11.8189

	GC%		Replicons	WGS \
0	64.5000		NaN	AHAL01
1	36.0528	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN	
2	34.0470	chromosome 1:NC_016407.2/CM001217.2; chromosom...		APN001
3	35.1208	chromosome 1:NC_016088.2/CM000834.2; chromosom...		ACUP02
4	35.7097	chromosome 1:NC_015438.2/CM001064.2; chromosom...		AEKE02
...
5126	38.1000		NaN	FXDE01
5127	38.1000		NaN	FXCK01
5128	38.0000		NaN	FXCO01
5129	38.1000		NaN	FXCP01
5130	38.1000		NaN	FXCG01

	Scaffolds	CDS	Release Date \
0	7795	38554.0	2013-04-19
1	7	48350.0	2001-08-13
2	2187	57661.0	2011-08-12
3	1191	71525.0	2010-01-05
4	3224	36010.0	2010-12-10
...
5126	1925	0.0	2017-04-21
5127	590	0.0	2017-04-21
5128	2013	0.0	2017-04-21
5129	1028	0.0	2017-04-21
5130	799	0.0	2017-04-21

	GenBank FTP \
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...	...
5126	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
5126	NaN

5127	NaN
5128	NaN
5129	NaN
5130	NaN

[5131 rows x 16 columns]

```
[210]: data = data.drop(columns=["Level", "Assembly"])
```

```
[211]: print(data)
```

	#Organism Name	Organism Groups \
0	Emiliana huxleyi CCMP1516	Eukaryota;Protists;Other Protists
1	Arabidopsis thaliana	Eukaryota;Plants;Land Plants
2	Medicago truncatula	Eukaryota;Plants;Land Plants
3	Glycine max	Eukaryota;Plants;Land Plants
4	Solanum lycopersicum	Eukaryota;Plants;Land Plants
...
5126	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5127	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5128	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5129	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes
5130	Saccharomyces cerevisiae	Eukaryota;Fungi;Ascomycetes

	Strain	BioSample	BioProject	Size(Mb)	GC% \
0	CCMP1516	SAMN02744062	PRJNA77753	167.6760	64.5000
1	NaN	SAMN03081427	PRJNA10719	119.6680	36.0528
2	A17	SAMN02299339	PRJNA10791	412.9240	34.0470
3	NaN	SAMN00002965	PRJNA19861	978.9720	35.1208
4	NaN	SAMN02981290	PRJNA119	823.7860	35.7097
...
5126	NaN	SAMEA3110475	PRJEB7675	12.0392	38.1000
5127	ZP848	SAMEA3110493	PRJEB7675	11.7086	38.1000
5128	VKMY373	SAMEA3110473	PRJEB7675	11.8194	38.0000
5129	245	SAMEA3110476	PRJEB7675	11.7081	38.1000
5130	M2ON0800_1A	SAMEA3110463	PRJEB7675	11.8189	38.1000

	Replicons	WGS	Scaffolds \
0	NaN	AHAL01	7795
1	chromosome 1:NC_003070.9/CP002684.1; chromosom...	NaN	7
2	chromosome 1:NC_016407.2/CM001217.2; chromosom...	APN001	2187
3	chromosome 1:NC_016088.2/CM000834.2; chromosom...	ACUP02	1191
4	chromosome 1:NC_015438.2/CM001064.2; chromosom...	AEKE02	3224
...
5126	NaN	FXDE01	1925
5127	NaN	FXCK01	590
5128	NaN	FXC001	2013
5129	NaN	FXCP01	1028

5130

NaN FXCG01

799

		CDS Release Date	GenBank FTP \
0	38554.0	2013-04-19	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
1	48350.0	2001-08-13	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
2	57661.0	2011-08-12	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
3	71525.0	2010-01-05	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
4	36010.0	2010-12-10	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000...
...
5126	0.0	2017-04-21	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5127	0.0	2017-04-21	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5128	0.0	2017-04-21	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5129	0.0	2017-04-21	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...
5130	0.0	2017-04-21	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900...

	RefSeq FTP
0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
3	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
4	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000...
...	...
5126	NaN
5127	NaN
5128	NaN
5129	NaN
5130	NaN

[5131 rows x 14 columns]

```
[212]: categories = [("Eukaryota;Protists", "blue"), ("Eukaryota;Plants;Land Plants", "green"), ("Eukaryota;Fungi;Ascomycetes", "black"), ("Eukaryota;Fungi;Basidiomycetes", "red")]
```

```
[213]: categories
```

```
[213]: [('Eukaryota;Protists', 'blue'),
        ('Eukaryota;Plants;Land Plants', 'green'),
        ('Eukaryota;Fungi;Ascomycetes', 'black'),
        ('Eukaryota;Fungi;Basidiomycetes', 'red')]
```