

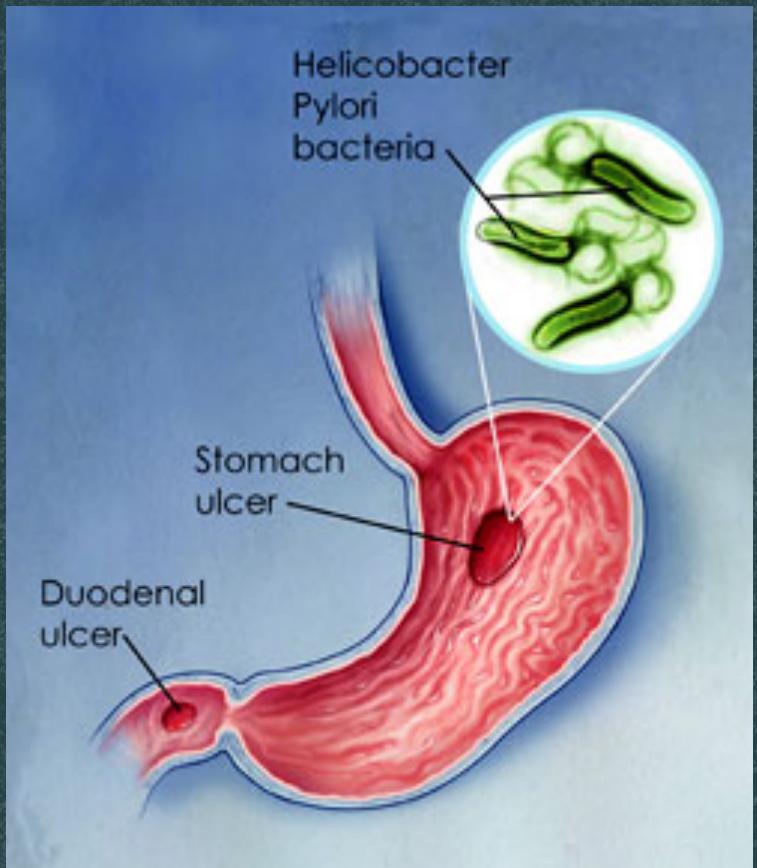
Genome Analysis: *Helicobacter pylori*

K i m b e r l y H a b o n

Introduction and Purpose

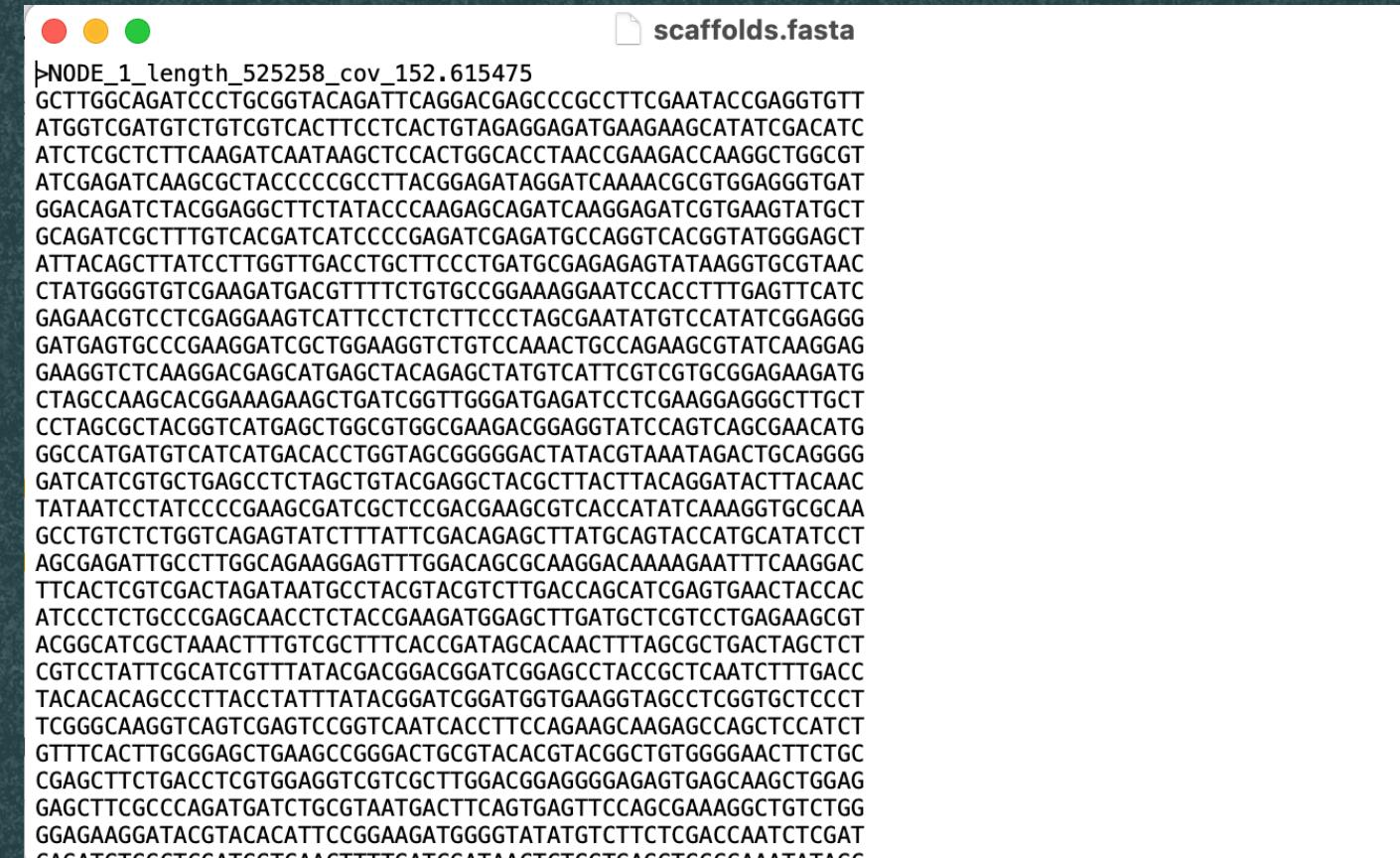
Helicobacter pylori is known for its role in various gastrointestinal diseases such as ulcers and gastric cancer.

Purpose: understand the genetic makeup of *Helicobacter pylori* by taking raw genomic data and assembling it into biological information.



Methods

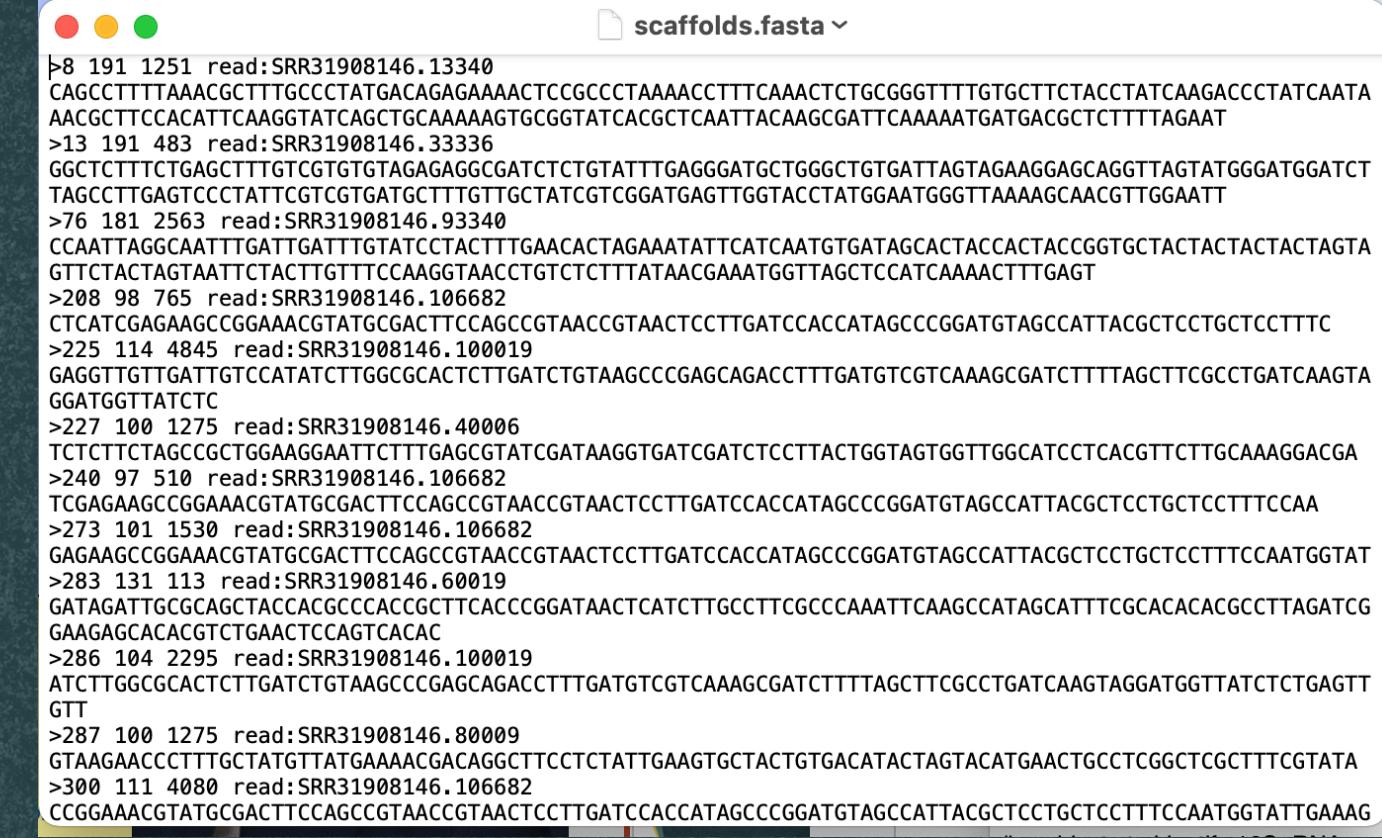
SPAdes Galaxy



The screenshot shows a file named "scaffolds.fasta" in a Galaxy tool interface. The file content is a FASTA sequence starting with the header: >NODE_1_length_525258_cov_152.615475.

```
>NODE_1_length_525258_cov_152.615475
GCTTGGCAGATCCCTCGGTACAGATTCAAGGACGAGCCGCTTCGAATACCGAGGTGTT
ATGGTCGATGTCGTCACTTCACTGTAGAGGGAGTAAAGAAGCATATCGACATC
ATCTCGCTTCAAGATCAATAAGCTTCACTGGCACCTAACGAAAGCCAAGGCTGGGT
ATCGAGATCAAGCCTACCCCGCCTACGGAGATAGGATCAAACGCGTGGAGGGTGT
GGACAGATCTACGGAGGCTTCTATACCAAGAGCAGATCAAGGAGATCGTAAGTATGCT
GCAGATCGCTTGTACGATCATCCCCAGATCGAGATGCCAGGTACCGTATGGAGCT
ATTACAGCTTATCTGGTGTGACCTGCTTCCCTGATGCGAGAGGATATAAGGTGCGTAAC
CTATGGGTGTCAGATGACGTTCTGTGCCGGAAAGGAATCCACCTTGAGTTCATC
GAGAACCTCTCGAGGAAGTCATTCTCTTCTCGAATATGTCATATCGAGGG
GATGAGTCCCCGAAGGATCGCTGGAAAGTCTGTCAAACTGCCAGAAGCGTATCAAGGAG
GAAGGTCTAAGGACGAGCATGAGCTACAGAGCTATGTCATTGCGTGCAGAGATG
CTAGCCAAGCACGGAAAGAAGCTGATCGGTTGGATGAGATCTCGAAGGGGGCTTGCT
CCTAGCGTACGGTCACTGAGCTGGCGAAGCGGAGGATCAGTCAGCGAACATG
GGCCATGATGTCATCATGGTACCTGGTAGACACCTGATGAAATAGACTGCAGGGG
GATCATGTCAGCCTCTAGCTGAGGCTACGCTTACTACAGGAACTTACAAC
TATAATCTATCCCGAAGCGATCGCTCGACGAACGCTACCATATCAAGGTGCGCAA
GCCTGCTCTGGTCAGAGTATCTTATTGACAGAGCTTATGCACTGATCATCTT
AGCGAGATTGCGCTGGCAGAAGGAGTTGGACAGCGAAGGACAAAAGAATTCAAGGAC
TTCACTCGTCACTAGATAATGCCCTACGTCAGTCTGACCGATCGAGTGAACCTACAC
ATCCCCTGCCCCAGCAACCTCACCGAAGATGGAGCTTATGCTCTGAGAAGCGT
ACGGCATCGCTAAACTTTCGCTTACCGGATAGACAACCTTAGCGCTACTAGCTCT
CGTCTATTGCACTTTATCGACGGACGGATGGAGCGCTACCGCTCAATCTTGACC
TACACAGCCCTTACCTATTACGGATCGGTAGGTAGCCTCGGTGCTCCCT
TCGGGCAAGGTCACTCGAGTCCGGTCAATCACCTTCAAGAAGAGCCAGCTCATCT
GTTTCACTTGGAGCTGAAGCCGGACTCGTACGGCTGGGAACCTTCTG
CGAGCTTCTGACCTCGTGGAGGTCTGCTGGACGGAGGGAGAGTGAAGCAAGCTGGAG
GAGCTTGGCCAGATGATCGCTAATGACTTCAGTGAAGTCCAGCGAAAGGCTGCTGG
GGAGAAGGATACTGACACATTCCGGAAAGATGGGGTATATGTCCTCGACCAATCTGAT
CAGATCGCTCCATGCTCAACTTTCGATGCTGAGCTGCCAATATGAG
```

Abyss Galaxy



The screenshot shows a file named "scaffolds.fasta" in a Galaxy tool interface. The file content is a FASTA sequence starting with the header: >8 191 1251 read:SRR31908146.13340.

```
>8 191 1251 read:SRR31908146.13340
CAGCCTTTAACGCTTGCCTATGACAGAGAAAACCTCCGCTAAACCTTCAAACCTCTGCGGTTTGTGCTTCTACCTATCAAGACCCATCAATA
AACGCTTCCACATTCAAGGTATCAGCTGCAAAAGTGCCTGATCACGCTCAATTCAAGCGATTCAAAATGATGACGCTCTTTAGAAT
>13 191 483 read:SRR31908146.33336
GGCTCTTCTGAGCTTGTGCTGAGAGGGAGTCTCTGATTGAGGGATGCTGGCTGTGATTAGTAGAAGGGAGCAGGTTAGTATGGGATGGATCT
TAGCCTTGAGTCCCTATTGCTGATGCTTTGTTGATCGTGGATGAGTTGGTACCTATGGAATGGGTTAAAGCAACGTTGAAATT
>76 181 2563 read:SRR31908146.93340
CCAATTAGGCAATTGATTGATTCTACTTTGAAACTAGAAATATTCAATGTGATAGCACTACCACTACCGGTGCTACTACTACTAGTA
GTTCTACTAGTAATTCTACTGTTCAAGGTAACCTGCTCTTTATAACGAAATGGTAGCTCATCAAAACTTGTAG
>208 98 765 read:SRR31908146.106682
CTCATCGAGAACCGGAAACGTATCGCACTTCCAGCGTAACCGTAACCTTGATCCACCATAGCCGGATGTAGCCATTACGCTCTGCTCTTTC
>225 114 4845 read:SRR31908146.100019
GAGGTGTTGATTGTCATCTTGGCGACTCTGATCTGTAAGCCGAGCACCTTGTAGTCGCAAAGCGATCTTAGCTCGCTGATCAAGTA
GGATGTTTATCTC
>227 100 1275 read:SRR31908146.40006
TCTCTCTAGCCGCTGGAAAGGAATTCTTGAGCGTATCGATAAGGTGATCGATCTCTACTGGTAGTGGTGGCATCCTCACGTTCTGCAAAGGACGA
>240 97 510 read:SRR31908146.106682
TCGAGAACCGGAAACGTATCGCACTTCCAGCGTAACCGTAACCTTGATCCACCATAGCCGGATGTAGCCATTACGCTCTGCTCTTCAATGGTAT
>273 101 1530 read:SRR31908146.106682
GAGAACCGGAAACGTATCGCACTTCCAGCGTAACCGTAACCTTGATCCACCATAGCCGGATGTAGCCATTACGCTCTGCTCTTCAATGGTAT
>283 131 113 read:SRR31908146.60019
GATAGATTGCGCAGCTACCGCCACCGCTTACCCGGATAACTCATCTTGCTCTGCCCCAAATTCAAGCCATAGCATTTCGACACACGCCCTAGATCG
GAAGAGCACACGCTGAACCTCCAGTCAC
>286 104 2295 read:SRR31908146.100019
ATCTGGCGCACTTGTAGTGTAAAGCCGAGCAGACCTTGTAGTCGCAAAGCGATCTTAGCTGCTGATCAAGTAGGATGGTATCTGAGTT
GTT
>287 100 1275 read:SRR31908146.80009
GTAAGAACCTTGTATGTTGAAAACGACAGGCTCTTATTGAAGTGTACTGTGACATACTAGTACATGAACGTGCTCGCTCGCTTGTATA
>300 111 4080 read:SRR31908146.106682
CCGGAAACGTATCGCACTTCCAGCGTAACCGTAACCTTGATCCACCATAGCCGGATGTAGCCATTACGCTCTGCTCTTCAATGGTATTGAAAG
```

Methods

QuastSpades

report.txt

```
All statistics are based on contigs of size >= 500 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs).  
Assembly scaffolds  
# contigs (>= 0 bp) 5694  
# contigs (>= 1000 bp) 71  
# contigs (>= 5000 bp) 53  
# contigs (>= 10000 bp) 45  
# contigs (>= 25000 bp) 34  
# contigs (>= 50000 bp) 21  
Total length (>= 0 bp) 5027165  
Total length (>= 1000 bp) 3512591  
Total length (>= 5000 bp) 3471582  
Total length (>= 10000 bp) 3415084  
Total length (>= 25000 bp) 3227620  
Total length (>= 50000 bp) 2808116  
# contigs 160  
Largest contig 525258  
Total length 3568865  
GC (%) 46.48  
N50 102577  
N90 25960  
auN 218915.5  
L50 7  
L90 34  
# N's per 100 kbp 33.90
```

Quastabyss

report.txt

```
All statistics are based on contigs of size >= 500 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs).  
Assembly scaffolds  
# contigs (>= 0 bp) 16841  
# contigs (>= 1000 bp) 215  
# contigs (>= 5000 bp) 137  
# contigs (>= 10000 bp) 108  
# contigs (>= 25000 bp) 42  
# contigs (>= 50000 bp) 12  
Total length (>= 0 bp) 5838538  
Total length (>= 1000 bp) 3597317  
Total length (>= 5000 bp) 3396960  
Total length (>= 10000 bp) 3188766  
Total length (>= 25000 bp) 2067366  
Total length (>= 50000 bp) 1028712  
# contigs 240  
Largest contig 158068  
Total length 3616194  
GC (%) 46.45  
N50 30155  
N90 7923  
auN 45272.3  
L50 33  
L90 116  
# N's per 100 kbp 100.44
```

Method

Barrnnap and bedtools

File rRNAsequences.fasta — Edited

>NODE_63_length_1794_cov_470.447292:90-1615
AAAGGAGGTGTTCCAGCCGACCTTCCGGTACGGTACCTTGTACGGACTTAGCCCCAGTCACCTGTATTACCCTAGGCCGCCCTCGCAGGTACGGACT
TAGGTACTCCAGACTCCCATGGCTTGACGGCGGTGTACAAGGCCGGAACGTATTCACCGGCCATGGCTGATGCGGATTACTAGCGAATCCAG
TCACGGAGTCGAGTTGACACTCGATCGAACGTGTATAAGGTTAGAGATTGATCCTGTGCCAGGTAGCTGCCCTTGTCTTACATTGTAAC
GTGTGTCGCCCGGATGTAAGGGCCGTGCTGATTTGACGTACATCCCCACCTTCTCACGCCCTACGCCAGTCTACGGAGTCCAGCATTACCT
TGGCAACCGTAGACAGGGGTTGCCTCGTTATGGCACTTAAGCCGACACCTCACGGCACGAGCTGACGACAACCATGCGACCTACATAGATGCCCCG
GGAAAGCCGACTTCACCGGCCGTACATCTACATTCAATCCGGTAAGGTTCTCGGTATCATCGAATTAAACACATGTTCTCCGCTTGTGCCGG
CCCGTCAATTCTTGTAGTTCACCGTTGCCGGTACTCCCCAGGTGGATAACTTAACGCTTCGCTTAGGAGCATACTGTATATCGCATACGCCAG
ATCATCGTTACGGCGTGGACTACCAAGGTATCTAATCTGTTGATACCCACGCTTCGCTCACGCTCAGTTGACCTTAGTAAGCTGCCCTCGA
TCGGAGTTCTCGGTGATATCTATGCATTCCACCGCTACACCACGATTCCGCCTACCTCATCTACACTCAAGCCGCCAGTATCAATGGCAATTAGGA
TAAGCTCCTAGATTTCACCGCTGACTTAACAGGCCCTACGCACCCCTAAACCCAATAACCGGATAACGCTCGATCCTCGTATTACCGCCGCT
TGGCACGGAGTTAGCCGATGCTTATCGTAGGGTACATTACTACAGCACACGTGCTGTTCTTATTCCCCCTACAAAAGAAGTTACAATCCATAAGAC
TCATCCTCACCGGACTTGGCTGGTCAGGCTCTGCCATTGACCAATTCTCACTGCTGCCCTGGTAGGAGTCTGGTCCGTGTCAGTACCG
TGGGGGTTCAACCTCTCAGTCCCTACCCATCGTGCCTGGTAGGCCGTTACCTCACCAACTAGCTAATGGGACGCATGCCATCTCACAGCGAAATT
ATTTCTCCTACCTTGGATGCCAAGGGGAGTGTACGCCGTATTAGTCCGTCTTCGACGGGTTATCCCCCTCTGTGAGGAAGTTGACATCGCCTAAC
ACCGTGCGCCGGTCGCCATCAGTCTTAGCAAGACCATGCTGCCCTGACTTGCATGTGTTAACGCTATCGTAGCCTACGTTATCCTGAGCCAGG
CAAACCTTC
>NODE_64_length_1710_cov_291.294550:492-1710. *** 16sRNA used this one to blast
AAAAACAAGCCAAACGCTCTATTAGTAGTAGTCAGCTAACGCTTACACATCTACCCCTATCAAGCACATAGTCTTGTGCGAGCTCAGG
AAGTTTATCTTGGAGTTGGCTCTGCTTAGATGCTTCAGCAGTTACATCCGTGTTAGCTACCCAGCGATGCTTGGCAGAACAACTGGTCA
AGTGACACGTCATCCGGCTCTCGTACTAGGGACAGCTCTCTCAACTTCCCTACGCCACGGCAGATAGGGACCGAAGTGTCTACGACGTTCTG
CCCAGCTCGGTACCGCTTAAATGGGAACAGCCATACCCCTGGACCTGCTCAGGCCAGGATGCGATGAGCCACATCGAGGTGCCAAACCTCC
TCGATGTGAGCTTGGGGGAGATCAGCCTGTTATCCCCGGGTACCTTTATCTTGTGAGCGATGGCCCTCACAGAACCCAGGATCACTATGA
GACTTCTGCTCTGCTTGACTTGTATGTCTTACAGTCAGGCTGGCTGTGCCATTACACTCAACTTGCATTTCAACCGCAATGAGCCAACCTTG
CCTCCGTTACTTTAGGAGGCGACCCAGTCAAACTACCCACCAAGCATTGCTCTGCCGTGGATAACACAGGCCAGTTAGCTAACAGAACATC
GGGTGGTATCTCAAGGATGGCTCCATAAGAGCCAAGGCCCTACTTCAAAGCCTCCACCTATCTGCGCATGATATTCCGTTAGCAGTGTAAAGTTG
GTAAAGGTCCACGGGGTCTTCCGTCTGCCGGTAGGAGGAATTTCACCTCCACTACAATTCACTGAATCTCTGGTTGAGACAGCTCCATCTC
TACGCCATTATGCAGGTGGTATTAACGACAAGGAATTGCTACCTTAGGACCGTTAGTTACGGCCGCCGTTACTCGGGCTCAATTCAACG
TCATCTTGCAGCTGACGCATCCTCTTAACCTCGAGCACCCGGCAGGCCGTACACCTTACTTCCCTTACGAGTTGGCAAAGTGTGTTTTGGT
ACACTCCCCACCCACTCTTCCGATACCCCAATTCTGCCACACCTTATCCCAACTTACCCCTTACTTCCGATACACTTCCGATACACTTCCGTTAAC
CCACACTTCCGATACACTTCCGATACACTTCCGATACACTTCCGTTAAC

DFAST

✓ scaffolds.fasta

```
>8 191 1251 read:SRR31908146.13340
CAGCCTTTAACGCTTGCCTATGACAGAGAAAACTCCGCCCTAAAACCTTCAAACCTGCGGGTTTGCTTACCTATCAAGACCTATCAATA
AACGCTTCCACATTCAAGGTATCAGCTGAAAAAGTCGGTATCACGCTCAATTACAAGCATTAAAGTACGCTCTTTAGAAT
>13 191 483 read:SRR31908146.33336
GGCTCTTCTGAGCTTGTGAGAGGGCATCTGTATTTGAGGGATGCTGGCTGTGATTAGTAGAAGGAGCAGGTTAGTATGGGATGGATCT
TAGCCTTGAGTCCCTATTGTCGTGATGCTTGTGCTATCGTCGGATGAGTTGGTACCTATGGAATGGTTAAAGCAACGTTGGAATT
>76 181 2563 read:SRR31908146.93340
CCAATTAGGCAATTGATTGATTGATTCCTACTTGAACACTAGAAATATTCAATGTGATAGCACTACCACTACCGGTGCTACTACTACTAGTA
GTTCTACTAGTAATTCTACTTGTCCAGGTAACCTGTCTTTATAACGAAATGGTTAGCTCCATAAAACTTGAGT
>208 98 765 read:SRR31908146.106682
CTCATCGAGAACGCCGAAACGTATGCGACTTCCAGCGTAACCGTAACCTGATCCACCATAGCCGGATGTAGCCATTACGCTCCTGCTCCTTC
>225 114 4845 read:SRR31908146.100019
GAGGTTGTTGATTGTCATATCTGGCGCACTTGTGATCTGTAAGCCGAGCAGACCTTGATGTCGTCAGCGATCTTAGCTCGCTGATCAAGTA
GGATGGTTATCTC
>227 100 1275 read:SRR31908146.40006
TCTCTCTAGCCGCTGGAAGGAATTCTTGAGCGTATCGATAAGGTGATCGATCCTTACTGGTAGTGGTGGCATCCTCACGTTCTGCAAAGGACGA
>240 97 510 read:SRR31908146.106682
TCGAGAACGCCGAAACGTATGCGACTTCCAGCGTAACCGTAACCTGATCCACCATAGCCGGATGTAGCCATTACGCTCCTGCTCCTTCAA
>273 101 1530 read:SRR31908146.106682
GAGAACGCCGAAACGTATGCGACTTCCAGCGTAACCGTAACCTGATCCACCATAGCCGGATGTAGCCATTACGCTCCTGCTCCTTCAAATGGTAT
>283 131 113 read:SRR31908146.60019
GATAGATTGCGCAGCTACCACGCCACCGCTTACCCGGATAACTCATCTTGCCTCGCCAAATTCAAGCCATAGCATTTCGACACACGCCCTAGATCG
GAAGAGCACACGTCTGAACCTCCAGTCACAC
>286 104 2295 read:SRR31908146.100019
ATCTTGGCGCACTTGTGATCTGTAAGCCGAGCAGACCTTGATGTCGTCAGCGATCTTAGCTCGCTGATCAAGTAGGATGGTTATCTGAGTT
GTT
>287 100 1275 read:SRR31908146.80009
GTAAGAACCCCTTGCTATGTTATGAAAACGACAGGCTTCTCTATTGAAGTGCTACTGTGACATACTAGTACATGAACGCTCGGCTCGCTTGTATA
>300 111 4080 read:SRR31908146.106682
CCGGAAACGTATGCGACTTCCAGCGTAACCGTAACCTGATCCACCATAGCCGGATGTAGCCATTACGCTCCTGCTCCTTCAAATGGTATTGAAAG
```

Methods

Blastn

An official website of the United States government. [Here's how you know.](#)

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastn suite » results for RID-1RACBWP1013

Home Recent Results Saved Strategies Help

< Edit Search Save Search Search Summary ▾

How to read this report? BLAST Help Videos Back to Traditional Results Page

Job Title Nucleotide Sequence
RID [1RACBWP1013](#) Search expires on 05-09 02:42 am [Download All](#) ▾
Program BLASTN Citation
Database core_nt See details ▾
Query ID lcl|Query_4381675
Description None
Molecule type dna
Query Length 1218
Other reports Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear exclude
Type common name, binomial, taxid or group name
+ Add organism

Percent Identity E value Query Coverage
[] to [] [] to [] [] to []
Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100 ?

select all 100 sequences selected GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Helicobacter pylori strain HpGP-JAP-026 chromosome, complete genome	Helicobacter pylori	2250	4500	100%	0.0	100.00%	1573045	CP079467.1
<input checked="" type="checkbox"/>	Helicobacter pylori strain HpGP-CHN-010 chromosome, complete genome	Helicobacter pylori	2250	4500	100%	0.0	100.00%	1650269	CP078994.1
<input checked="" type="checkbox"/>	Helicobacter pylori strain P07353 chromosome, complete genome	Helicobacter pylori	2244	4489	100%	0.0	99.92%	1568376	CP141774.1
<input checked="" type="checkbox"/>	Helicobacter pylori strain AL02 chromosome, complete genome	Helicobacter pylori	2244	4489	100%	0.0	99.92%	1575789	CP058250.1
<input checked="" type="checkbox"/>	Helicobacter pylori VN1291 DNA, complete genome	Helicobacter pylori	2244	4146	100%	0.0	99.92%	1702481	AP026444.1
<input checked="" type="checkbox"/>	Helicobacter pylori strain Hpfe113 chromosome	Helicobacter pylori	2244	4238	100%	0.0	99.92%	1570128	CP094174.1
<input checked="" type="checkbox"/>	Helicobacter pylori strain HpGP-USA-422 chromosome, complete genome	Helicobacter pylori	2244	4489	100%	0.0	99.92%	1602979	CP078194.1
<input checked="" type="checkbox"/>	Helicobacter pylori strain Hpfe048 chromosome, complete genome	Helicobacter pylori	2244	4434	100%	0.0	99.92%	1613194	CP094134.1
<input checked="" type="checkbox"/>	Helicobacter pylori OK310 DNA, complete genome	Helicobacter pylori OK310	2244	4489	100%	0.0	99.92%	1591278	AP012601.1
<input checked="" type="checkbox"/>	Helicobacter pylori strain AL02 chromosome, complete genome	Helicobacter pylori	2244	4489	100%	0.0	99.92%	1575789	CP058250.1

Feedback

Methods

FastANI

Column 1	Column 2	Column 3	Column 4	Column 5
scaffolds_fasta_query	acinonychis_fasta_ref	89.4921	431	1137
scaffolds_fasta_query	cetorum_fasta_ref	82.0794	393	1137

Conclusion

After reconstructing the genome by using the programs that was mentioned, the identification of species was *Helicobacter pylori*.

Knowing the assembly of a genome will allow researchers and scientists to develop new treatments for diseases.

