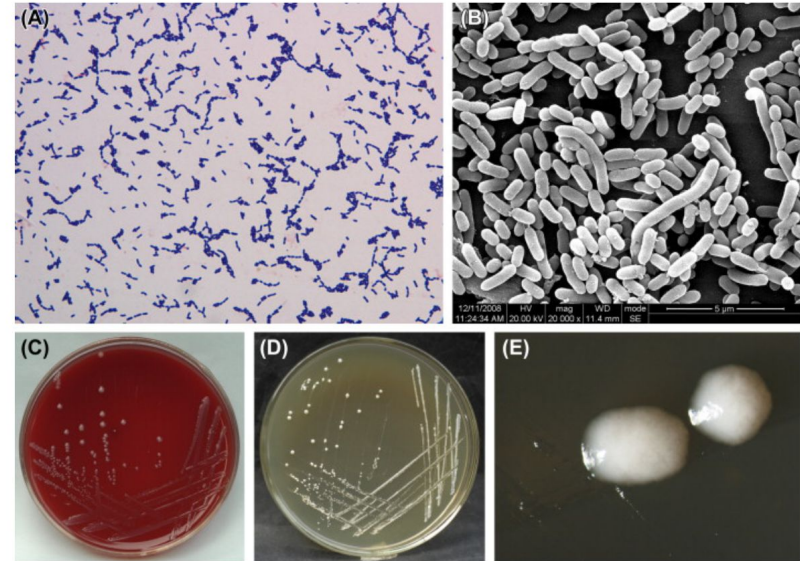


Genome Analysis of *Lactobacillus paracasei*

By: Analisa Coppens

Introduction

- Gut microbiome, probiotic
- **Importance:** digestion, immunity, gut balance
- **Goal:** Assemble, Annotate, & Compare



[Sign in to download full-size image](#)

Figure 3.8. (A) *Lactobacillus casei* cell (Gram stain). (B) *L. casei* cell (SEM). *L. casei* cells measure $0.7\text{--}1.1\text{ }\mu\text{m} \times 2.0\text{--}4.0\text{ }\mu\text{m}$ and are arranged in a chain. Individual cells usually have rounded ends. Cells stain gram-positive. (C) Colonies of *L. casei* (BHI blood agar). (D) Colonies of *L. casei* (Rogosa agar). (E) Cheese colonies of *L. casei* (stereomicroscope).

- Builds contigs from raw DNA reads

stats.tabular

| h | n:500 | L50 | min | N75 | N50 | N25 | E-size | max | sum | name |
|-------|-------|-----|-----|------|------|-------|--------|-------|---------|--------------------|
| 44572 | 1940 | 686 | 500 | 629 | 810 | 1121 | 967 | 4924 | 1576558 | abyss-unitigs.fa |
| 43677 | 1750 | 502 | 500 | 751 | 1202 | 1770 | 1488 | 8300 | 1832880 | abyss-contigs.fa |
| 33515 | 538 | 82 | 500 | 4711 | 9351 | 19357 | 14429 | 66971 | 2911730 | abyss-scaffolds.fa |

Abyss summarizes genome quality

- contigs
- genome length
- N50

scaffolds.fasta

```
>NODE_1_length_424351_cov_542.194982
AATTTGTCCTAACTTGGAAATAACACGAGAACAACTTTTTAATATTTGAGCTGTCCACT
TGACGGGGCACCGCCCCCTGACTGCTTCGTTCTTTTTATTTTACTAAGCTTGCTCCTT
GAGTTTGAGCGAACCATCTCTCGACAAGCTTTGTGAGAATGCGATAGAGAAGTTCTCAGA
GAAATCGCGAGATCCGACAGTAGGTAAGTGTAGGTATCGGCTACACCGAGGATCAGGTAC
TTTTGCTAACCACTGTCGAGTCGTGACCTTGTCTGCGGTATAGAGTTCCAGAGAC
GAGTGCATCTGTCTGTTGGCTGATTAATGCGCGCAATGCCAAGTTTTCTGCCATGTTAG
CTTCTTTCTAACTTGATACCATTATACCCGGCAAAGATTTTTAGGCAAAGCTGTGCTGC
CTCTTTGATTTCTGTCGCGAGCTAGGCTAAGATGTTAAGCATGATGATACCATGGTTAGTA
ATCCAGCCAACAATTCAATCTGCGCAGCAATTGCTAGGGATGCAACTGAGTCTGGCTGGC
TGCAGTGGTTTTGATTGTTGAGACTGAGGCGTATTTAGGTGTAGAGGATAAAAGCCGCGCAT
GCGTTTGCGCGTCAAAAAACCCGCGTAATCACAGTCTGTACTTACCGCGCGGCAACGTC
TATGTTTATCAATGCGCCAGTACTGCTTATTGAATATTGTGACGCGAACTGCGGAGGTG
CCGGAGTGTATCTTGATTTCGGGCGCTGGAGCCGAAGCTGGTATGGCGAAAATGCAGCAA
CGTCGGCACCAAGAAGGCGTTGGGTTGACCAATGGCCCGGTAAGCTTTGTGACGGCATTG
GCAATTGACCGTGCCTTGACGGCAGTCGATTAACCGCTGGCGGCTGCAGCTGACGGCG
ACAACATTTTCAGCCGCGTCACATCAAAATAGGACCGCGCATTTGGCATTGCAAAACAAAGAA
CCTTGGACAAGCGCACCTTTGCGTTTTTTCGTCGAGGGGAATCCGTTTGTATCAGGCACA
GCTCGCGGAGATCAGGATTTAAAAACGAGAGGATGGCAATCATGAGCCAAATCAATATGT
TAGATGCAATTTGGTGACAAAATGGATGCGCTGGATTTGATGGTGATTTAAGTCTAACT
GGGATAAAGATGCCCATGTGATTGAACCTGAAATCAGATGACAGTACAAAGCGAGTCTG
GTATCGAGGTTGAAGATCAATCTGGTGAGACCGTTGACAATGGGCCAGTTGAATATCAAG
ATGCCATTTTATTTATGATGAAACCCGGCTGCACGGTGAAGAGTATGCAGATGATTATT
TAGCTGTTTTTCGGGTTTAAATGGTAAGAAGGGCATTGATAAAGCCACAGTGGATGCTTTGT
TTATCTATCTGACAGGACTTTTTAGACGATAGCGAAAGTGACCTCATGGATTTTGTGATG
GCACGCTGACGATGATACCTTCGTTTGAATTTTACGATGCGGCATTTGAACGGATTTC
TGACGGAACAGCCGAGGAGGATAACCGCATCTTTTTACCTATCCCAAGTACTAGTCAG
CCTGCCAAAGAAGAAGAAATATTGGCCGAAAGACCGGTGCGCTGACTTTTCAAGCCGGT
ATCTGGACAGTAGCAGTTTTTGATCAAAAGAACACGTTTAAAGGAGGCAACCATGAATGA
TTGGACAGCGTTAACGGTCACAACCACCACTGAAGCCATTGAGGCGGTGAGCAATATTTT
```

Actual DNA output

Report

| | SPAdes_on_data_7_and_data_8_Scaffolds | SPAdes_on_data_7_and_data_8_Scaffolds_broken |
|---------------------------|---------------------------------------|--|
| # contigs (>= 0 bp) | 127 | - |
| # contigs (>= 1000 bp) | 34 | 38 |
| Total length (>= 0 bp) | 3063500 | - |
| Total length (>= 1000 bp) | 3043248 | 3042848 |
| # contigs | 38 | 42 |
| Largest contig | 424351 | 424351 |
| Total length | 3046465 | 3046065 |
| GC (%) | 46.20 | 46.20 |
| N50 | 181098 | 181098 |
| N90 | 94254 | 61105 |
| auN | 210090.7 | 208278.5 |
| L50 | 6 | 6 |
| L90 | 15 | 16 |
| # N's per 100 kbp | 13.13 | 0.00 |

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).

Quast

- Quality Stats
- N50, contigs, GC content

Report

| | ABYSS_on_data_7_and_data_8_scaffolds | ABYSS_on_data_7_and_data_8_scaffolds_broken |
|---------------------------|--------------------------------------|---|
| # contigs (>= 0 bp) | 33515 | - |
| # contigs (>= 1000 bp) | 445 | 789 |
| Total length (>= 0 bp) | 7885312 | - |
| Total length (>= 1000 bp) | 2898376 | 2627138 |
| # contigs | 545 | 1355 |
| Largest contig | 67518 | 24250 |
| Total length | 2960830 | 2893639 |
| GC (%) | 46.18 | 46.16 |
| N50 | 9401 | 3646 |
| N90 | 2660 | 1067 |
| auN | 14533.8 | 5134.8 |
| L50 | 83 | 222 |
| L90 | 314 | 767 |
| # N's per 100 kbp | 1545.01 | 4.25 |

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).

- Identified 16S, 23S, & 5S rRNA genes
- Confirms quality & species ID

● ● ●  rRNA.fasta

```
>16S_rRNA::NODE_28_length_4918_cov_3771.755009:3287-4853(-)
TATGAGAGTTTGATCTGGCTCAGGATGAACGCTGGCGCGTGCCTAATACATGCAAGTGAACGAGTTCTCGTTGATGATCGGTGCTTGACCGAGATTCAACAT
GGAACGAGTGGCGGACGGGTGAGTAACACGTTGGGTAACCTGCCCTTAAGTGGGGGATAACATTTGGAACAGATGCTAATACCGCATAGATCCAAGAACCGCATGG
TTCTTGGCTGAAGATGGCGTAAGCTATCGCTTTTGGATGGACCGCGCGTATTAGCTAGTTGGTGAAGTAAATGGCTCACCAGGCGATGATACGTAGCGGAAT
GAGAGGTTGATCGGCCACATTGGGACTGAGACACGGCCAAACTCTACGGGAGGCGAGTAGGGAATCTTCCACAATGGACGCAAGTCTGTGAGGACACGCGC
CGTGAGTGAAGAAGGCTTTCCGGTCTGTAATACTCTGTTGTTGGAGAAGAATGGTGGCAGAGTAACGTGTGTCGGCGTGACGGTATCCAACAGAAAGCCACGGCT
AACTACGTGCCAGCAGCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGATTATTGGGCGTAAAGCGAGCGCAGGCGGTTTTTAAAGTCTGATGTGAAAGCCCT
CGGCTTAACCGAGGAAGCGCATCGGAACTGGGAACTTGAAGTGCAGAAAGGACAGTGGAACTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAAGAACC
AGTGGCGAAGGCGGCTGTCTGGTCTGTAACCTGACGCTGAGGCTCGAAAGCATGGGTAGCGAAGCAGGATTAGATACCCCTGGTAGTCCATGCCGTAACGATGAATGC
TAGGTGTTGGAGGTTTCCGCCCTTCAAGTCCGAGCTAACCGATTAAGCATTCCGCTGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCC
CGCACAAAGCGGTGGAGCATGTGGTTAATTCGAAGCAACGCGAAGAACCCTACCAAGTCTTGACATCTTTGATCACCTGAGAGATCAGGTTTCCCTTCCGGGGC
AAAATGACAGGTGGTGCATGGTTGTGTCAGCTCGTGTGTCGATGTTGGTTAAGTCCCGCAACGAGCGCAACCCCTTATGACTAGTTGCCAGCATTTAGTTGGG
CACTCTAGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAATCATATGCCCTTATGACCTGGGCTACACACGCTGTACAATGGATGGTACA
ACGAGTTGCCGAGACCGGAGGTCAAGCTAATCTCTTAAAGCATTCTCAGTTGCGACTGTAGGCTGCAACTCGCTACACGAAGTGGGAATCGCTAGTAATCGCGG
ATCAGACCGCGCGGTGAATACGTTCCCGGGCTTGTACACACGCGCCGTACACCATGAGAGTTTGTAAACCCGAAAGCGGGTGGCGTAACCCCTTTAGGGAGCG
AGCGCTTAAGTGGGACAAATGATTAGGGTGAAGTCTGAACAAGGTAGCGTAGGAGAACCCTGCGGCTGGATCACCTCCTT
>23S_rRNA::NODE_28_length_4918_cov_3771.755009:7-3020(-)
TTAAGTTACAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCGGATGAAGGACGGAACATAACCGATATGCTTGGGGAGCTATAAGTAAGCTTTGATCCGGA
GATTTCCGAATGGGGAAACCCAGTACACATCAGTGTATTGCTTGTGTCAGTGAATACATAGCTGGCCGGCGGAGACGCGGGGAACTGAAACATCTCAGTACCCGAG
GAAGAGAAAGAAAATCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAGAGCCCAAACCGAGAAGCTTGCTTCTCGGGGTTGTAGGACTGGACATTGGAGTTACC
AAAGTCCGACGTAGTCGAAGTCAGCTGGAAGGTGCGCCACAGAAGGTGAAGGCCCTGTAAGCGAAACGTCGAACCCCTCCGTCAGGATCCTGAGTACGGCGGAAC
ACGTGAAATTCGCTCGGAATCCGGAGGACCATCTCCAAGGCTAAATACTCCCTAGTGACCGATAGTGAACGAGTACCGTGAGGGAAGGTGAAAAGCACCCCGG
AAGGGGAGTGAACAGTTCCTGAAACCGTGTGCTACAATACTAGTCAAAGCCCGTTAATGGGTAAATGGCGTGCTTTTGTAGAATGAACCGCGAGTTACGTTTGTG
TGCGAGGTTAAGATGAAAGTCGGAGCGGTAGCGAAAGCGAGTCTGAACAGGGCGAATGAGTAAGCAGATGTAGACCCGAAACCAAGTGACCTACCATGACCAAGG
TTGAAGGTGTGGTGAACACACTGGAGGACCGAACCATGTATGTTGAAAAATGCTGGGATGAGTTGTGGGTAGCGGTGAAATCCAAACGAACCTGGAGATAGCT
GGTCTCTCGAATAAGCTTATAGGTTAGCTTCGGAGGATGGATCATGGAGGTAGAGATCTTTTGAACCTAGGGGCGATCAAGGGTTACTGAATTCAGATAAAT
```

Blast

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

DownloadSelect columnsShow100?

☐ select all

3 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"/> | Lactobacillus paracasei strain CAUH35, complete genome | Lacticaseibacill... | 2892 | 14442 | 100% | 0.0 | 100.00% | 2770411 | CP012187.1 |
| <input checked="" type="checkbox"/> | Lacticaseibacillus paracasei strain TK-P4A chromosome, complete genome | Lacticaseibacill... | 2892 | 14464 | 100% | 0.0 | 100.00% | 3020185 | CP045567.1 |
| <input checked="" type="checkbox"/> | Lacticaseibacillus paracasei strain L14 chromosome, complete genome | Lacticaseibacill... | 2892 | 14464 | 100% | 0.0 | 100.00% | 2997905 | CP129524.1 |
| <input type="checkbox"/> | Lacticaseibacillus paracasei strain MGL98 chromosome, complete genome | Lacticaseibacill... | 2892 | 14437 | 100% | 0.0 | 100.00% | 3022416 | CP168648.1 |
| <input type="checkbox"/> | Lacticaseibacillus paracasei strain IB1 chromosome, complete genome | Lacticaseibacill... | 2892 | 14448 | 100% | 0.0 | 100.00% | 3147666 | CP162491.1 |
| <input type="checkbox"/> | Lacticaseibacillus paracasei strain VHProbi OF10 chromosome, complete genome | Lacticaseibacill... | 2892 | 14464 | 100% | 0.0 | 100.00% | 3084431 | CP092498.1 |

1. Lactobacillus paracasei strain CAUH35
2. Lacticaseibacillus paracasei strain TK-P4A
3. Lacticaseibacillus paracasei strain L14

- Annotated coding sequences, rRNAs, & tRNAs
- Insights to predicted functions

```

genome.txt
>Feature NODE_1_length_424351_cov_542.194982
354      106      CDS
                    inference      ab initio prediction:Prodigal:002006
                    locus_tag      LLPCHJPJ_00001
                    product hypothetical protein

451      1065     CDS
                    EC_number      3.2.2.-
                    db_xref COG:COG2094
                    inference      ab initio prediction:Prodigal:002006
                    inference      similar to AA sequence:UniProtKB:P9WJP7
                    locus_tag      LLPCHJPJ_00002
                    product Putative 3-methyladenine DNA glycosylase

1062     1556     CDS
                    inference      ab initio prediction:Prodigal:002006
                    locus_tag      LLPCHJPJ_00003
                    product hypothetical protein

1673     2620     CDS
                    EC_number      2.1.1.-
                    db_xref COG:COG2264
                    gene      prmA
                    inference      ab initio prediction:Prodigal:002006
                    inference      similar to AA sequence:UniProtKB:P45558
                    locus_tag      LLPCHJPJ_00004
                    product Ribosomal protein L11 methyltransferase

2753     3493     CDS
                    EC_number      2.1.1.193
                    db_xref COG:COG1385
                    gene      rsmE
                    inference      ab initio prediction:Prodigal:002006
                    inference      similar to AA sequence:UniProtKB:P54461
                    locus_tag      LLPCHJPJ_00005
                    product Ribosomal RNA small subunit methyltransferase E

```

| Column 1 | Column 2 | Column 3 | Column 4 | Column 5 |
|--|---------------------|----------|----------|----------|
| SPAdes_on_data_7_and_data_8__Scaffolds_query | casei_fasta_ref | 80.8248 | 413 | 996 |
| SPAdes_on_data_7_and_data_8__Scaffolds_query | rhamnosus_fasta_ref | 79.8852 | 425 | 996 |

- Compared genome to references species
- Identity: 80.8% & 79.8%
- Confirms related but distinct species

Conclusion

- Goal: Assemble, annotate, & compare genomes
- Genome successfully assembled & annotated
- Different species confirmed by FastANI