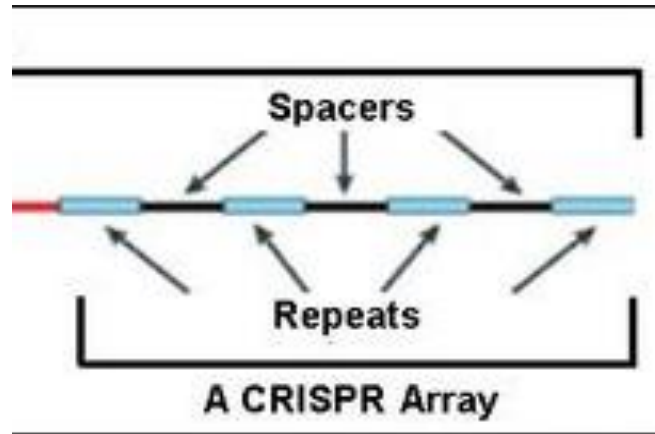


The existence of CRISPRs shared in common amongst Lactobacillus

Deoxyribonucleic acid (DNA) encodes many genetic functions of all organisms and viruses. The DNA is formed by the sequence of nucleotides. The main nucleotides are pyrimidine (Cytosine, Thymine, and Uracil) and purine (Adenine, Guanine). When we look at a sequence of a genome, the sequence is consisted of different nucleotides with random distribution. But in some organisms, there are such patterns exist within the sequence of a genome. One particular pattern that can be recognized easily is “Tandem Repeats”. Tandem repeats are when short units that repeat one after another.

For a special type of tandem repeats, CRISPR is possibly seen in a sequence of a genome. CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. Unlike tandem repeats, CRISPR has spacer sequences between each repeats. Many scientists claim that CRISPRs might be the keys to immunity of bacteria. According to Branger et al (2010), CRISPRs typically consist of 23 to 47 base pair (bp) with spacer sequences. For their study, they focused finding the CRISPRs in Escherichia Coli. Instead of using all strains of E.Coli, they performed on 27 fully sequenced strains. In result, they found that there are some CRISPRs shared among E. Coli strains: Repeat 1, 2, and 3.



- Repeat1: GGTTTATCCCCGCTGGCGCGGGGAACAC
- Repeat2: GGTTTATCCCCGCTGGCGCGGGGAACTC
- Repeat3: GTTCACTGCCGTACAGGCAGCTTAGAAA

These common repeats are present through different strains of E. Coli. The researchers have not found that what these that particular CRISPRs do.in order to study the organisms of Lactobacillus, there are some limitations. In Biobike, there are only 8 organisms available. Due to the storage space of biobike, we are going to consider the first three organisms.

–**Lactobacillus-acidophilus-NCFM**

–**Lactobacillus-brevis-ATCC-367**

–**Lactobacillus-casei-ATCC-334**

–Lactobacillus-delbrueckii-subsp-bulgaricus-ATCC-BAA-365

–Lactobacillus-gasseri

–Lactobacillus-johnsonii-NCC-533

Min Im
Research Report

–Lactobacillus-plantarum-WCFS1

–Lactobacillus-reuteri-F275

First step to find CRISPRs would be using “COUNTS-OF-K-MERS.BIKE” to see what kind of sequences tend to appear the most. Then take the coordinates for each hit and calculate the length of distance.

CRISPRs in each organism:

–Lactobacillus-acidophilus-NCFM:

GGATCACCTCCACATACGTGGAGAAAA (30 repeats of 27-nt sequence with 34-nt spacers)

```
1541045 GTTACCTCCACTTTTCATAGAAAAAATAAAAAGGCTCTAAAAGCTACAGAGTTACCATCGAG
1541105 GATCACCTCCACTTTTCGTGGAGAAAAATTGGAATCTCATCGTAAGAAAAAAGTCGCATATA
1541165 GGATCACCTCCACATACGTGGAGAAAAATCCTTTTCCTAGGATCTTCATAAGCTTCTCGCC
1541225 AGGATCACCTCCACATACGTGGAGAAAAATATCGTAGTCAATCTCGTACTTAAAAACCACCC
1541285 TGGGATCACCTCCACATACGTGGAGAAAAATCCAGGTTGACTTGCGCTAGGTGTTGCATCA
1541345 ATAGGATCACCTCCACATACGTGGAGAAAAATTAGGCAAATAGCCAATTTTTATCATAACAT
1541405 TCCGGGATCACCTCCACATACGTGGAGAAAAATCGGCAATTTTTGAAACAAAACAATATGT
1541465 ATATAGGATCACCTCCACATACGTGGAGAAAAATAAATAAGGAAGATATTGCCACCCTCGG
1541525 TACCCAGGATCACCTCCACATACGTGGAGAAAAATACAAGTTTTGCTCTAACCATGATGTT
1541585 GTAAACAGGATCACCTCCACATACGTGGAGAAAAATACGTTAAAGCGGACAATAAGCTTCA
1541645 ACGTTTTAGGATCACCTCCACATACGTGGAGAAAAATCGTGCTTGAAATTGCTCTCGGGGT
1541705 TTCGCCTAAGGATCACCTCCACATACGTGGAGAAAAATTTTGCTGCGAGTAACTCTGACT
1541765 TGTTTACCCGGGATCACCTCCACATACGTGGAGAAAAATTTAGCTAAGTTTAAGACCGAA
1541825 GATGGCCAAAGGGATCACCTCCACATACGTGGAGAAAAATCGGCAATTTTTGAAACAAAACA
1541885 ACTATGTATATAGGATCACCTCCACATACGTGGAGAAAAATCGGCAATTTTTGAAACAAAAC
1541945 AACTATGTATATAGGATCACCTCCACATACGTGGAGAAAAATACAAGTTTTGCTCTAACCA
1542005 TGATGTTGTAACAGGATCACCTCCACATACGTGGAGAAAAATAGCTATCCAAATATTTAAA
1542065 TTTGCACTAGTTAAGGGGATCACCTCCACATACGTGGAGAAAAATGAAGAATTTTTATCTTC
1542125 TAGGTGGCTTTTTTTGTGGGATCACCTCCACATACGTGGAGAAAAATAGAAATATTTGATTT
1542185 TGATAGTGA AAAAAGAATAGGATCACCTCCACATACGTGGAGAAAAATCAAGATCAACCAT
1542245 TCATTTGCCACGCAAAATCGGGATCACCTCCACATACGTGGAGAAAAATATTTTTCTGGTGT
1542305 GAAAAATGCTGATGATATTGGGATCACCTCCACATACGTGGAGAAAAACCCAAGATATAAT
1542365 TAAAAGAAAAATATCGTAGTCAGGATCACCTCCACATACGTGGAGAAAAATTTGATTTGAG
1542425 TTAAGTACCGGTGCATCAATGAGGATCACCTCCACATACGTGGAGAAAAATCTTTAATGT
1542485 GACCAAAAAGCGCTCGTAGCAGCAGGATCACCTCCACATACGTGGAGAAAAACGGACGCAAAA
1542545 AGGGCAGAAAAGCTAATTTATGACGGGATCACCTCCACATACGTGGAGAAAAATCTGCTATG
1542605 TCCGCTCTTGCTTGCAATGCAAGTCAAGGATCACCTCCATATACGTGGAGAAAAACCCACGCC
1542665 ACATATTGTGTGCGGAATAGCCTAAGGGATCACCTCCACATACGTGGAGAAAAATCTGCC
1542725 AAAATAGTAGCCGAATCTGGAAAAATCAGGATCACCTCCACATACGTGGAGAAAAACAAAGT
1542785 GTCACAAAATAGCACCTTTTATTAAAAGGATCACCTCCACATACGTGGAGAAAAACAGAG
1542845 TTTAGTCGTCGCAATGTGTGTCTCAATGAGGATCACCTCCACATACGTGGAGAAAAATCTAA
1542905 TACGTAGGTGGCAAGCGTTGTCCGGATTTAGGATCACCTCCACATACGTGGAGAAAAACAT
1542965 AAAAATAAGAGGAAACCACCGTTTTCTCTTAGGATCACCTCCACATACGTGGAGAAAAACA
1543025 CTAAAGGAATCCCATCGTTAAGCTATTTTAAATTTACTAAAAAGCAAAATTTCTTTAGTT
1543085 TGCATTTCCCAACCAAAATCATCGACACAATTATATCAAAATTTCTTCTACATATATCTA
1543145 GCTCATATAATAATTTTTAAATTTGCTATTATATTACAGGATTACAAATAAAACAAAATATG
```

–Lactobacillus-brevis-ATCC-367

GGATCACCCCCACACCTGTGGGGAATAC (4 repeats of 28-nt sequence with 33-nt spacers)

```
2249675 TAGTTCTTTGCTTGTGACAGCGCTACTTCCGCTAATACTAGTAGTTTATGGAAGTAGCGG
2249735 AAATGCCCCCACACCTGTGGGGTATACGCTTAAAAATCAAGCAGTCGTTGGGCTACGTCAA
2249795 GGATCACCCCCACACCTGTGGGGAATACTGACAACCAAAAAATATACGACTTTAACTTCT
2249855 GGGATCACCCCCACACCTGTGGGGAATACCTTTCTGGTATTTCTTCTCTATGTCATGATT
2249915 AAGGATCACCCCCACACCTGTGGGGAATACGATGAATAAGAAAAAGAAAAACAAAATAATAC
2249975 GCAGGATCACCCCCACACCTGTGGGGAATACACTAAAGAAAAACCTATTATACCGGCATT
2250035 TCATTTTAGCAATACCGCGATTCTCATCACTTTGCTTCTGAGTGCAGGTGCTCATGGTTG
```

Blue region is a similar repeat

–Lactobacillus-casei-ATCC-334

GTTTTCCCCGCACATGCGGGGGTGATCC (21 repeats of 28-nt sequence with 33-nt spacers)

```
356754 TAACGAGCAACCGCATCACGCGTTGTCTGATGCGACGGCAACAATGGCACTGGCGGACAA
356814 ACTCATAAAAAAGGGATCTTAACGATTGGCAAACGTTGATGGGATAGGGATTTTTTAGT
356874 GTTTTCCCCGCACATGCGGGGGTGATCCCTTCTCTGACTGGCTTACCGCTTCTTTGCATT
356934 TGTTTTCCCCGCACATGCGGGGGTGATCCCGCTTTTATAACATTTATGCCGCATAAAAAAC
356994 ACGTTTTCCCCGCACATGCGGGGGTGATCCCTGGCGACAATTCGATTTCTAAACGTAATA
357054 CCGTTTTCCCCGCACATGCGGGGGTGATCCCTGGCGACAATTCGATTTCTAAACGTAATA
357114 ACCGGTTTTCCCCGCACATGCGGGGGTGATCCCGCATTGTAATTCCTCCTAATGCGA
357174 ATTTGTTTTCCCCGCACATGCGGGGGTGATCCCGCGTTTGATGATGGCAGCGCGGCAAC
357234 TTTAACGTTTTCCCCGCACATGCGGGGGTGATCCTCAAAATCGCGACCTGGTTCAAAATG
357294 CGGCTGAGTTTTCCCCGCACATGCGGGGGTGATCCTGCTCGTGCGGTGTCTGGAGTTGCT
357354 GGCAAGTAGTTTTCCCCGCACATGCGGGGGTGATCCCTATATTTAGCTCAATTGCAACAA
357414 GAGGGCAAAGTTTTCCCCGCACATGCGGGGGTGATCCCGACAATAACGAGCTTGCGCTGA
357474 TGGCAACTTTGTTTTCCCCGCACATGCGGGGGTGATCCCAAGCGAGTCCGATCCTCCCGT
357534 ATTGTGAGCATGTTTTCCCCGCACATGCGGGGGTGATCCTGGCTGTTAGTTTGGCCAGTG
357594 TTATAAGCCGAGTTTTCCCCGCACATGCGGGGGTGATCCCAACTGGTGCGGTAATCGGG
357654 CGGAAATTGTTAAGTTTTCCCCGCACATGCGGGGGTGATCCCGATATTATACCGGTGTTT
357714 TTACCGGTTCTCCTGTTTTCCCCGCACATGCGGGGGTGATCCTGAGTTTGGTATCACGCT
357774 CTTTTGTGGTGTGTCAGGTTTTCCCCGCACATGCGGGGGTGATCCTAAATGTACTTCTTCC
357834 ACCACGTGGCACTATAAGTTTTCCCCGCACATGCGGGGGTGATCCTGCCGAAGTTGAAGA
357894 CATGGGTGAGGCTTTTGCCTTTTTCCCCGCACATGCGGGGGTGATCCCGACCAAAGCCAAA
357954 CCAACACAGGCGAATACAAGTTTTCCCCGCACATGCGGGGGTGATCCTAAGGTTGTCAAC
358014 ACGGCTTATGCCAACAAGGAGTTTTCCCCGCACATGCGGGGGTGATCCCGACATCATCAA
358074 GGATTGCCAGTTCAACTTGGGGTTTTCCCCGCACATGCGGGGGTGATCCTACCAGCACGG
358134 GGTTACCATGTCAACCCTTTTTGTTTTCCCGTATGTGTCAAGTCTTTCTCGGTTCCACTA
358194 ATTGAATATTCTTTTTTGTCTCGACTGCCGGTTCATCCTGGTCTACTGAGTGGGTTCAGTT
```

Next step is to see any of these CRISPRs appear in other organism amongst Lactobacillus. For this step, the Biobike function SEQUENCE-SIMILAR-TO would be used.

Searching Lactobacillus-acidophilus-NCFM's CRISPR to other organisms:

SEQUENCE-SIMILAR-TO each "GGATCACCTCCACATACGTG..." IN SEQUENCE-OF lactobacillus-brevis... Options Options
SEQUENCE-SIMILAR-TO each "GGATCACCTCCACATACGTG..." IN SEQUENCE-OF lactobacillus-casei-... Options Options

>NIL, NO RESULTS FOUND!

Searching Lactobacillus-brevis-ATCC-367's CRISPR to other organisms:

SEQUENCE-SIMILAR-TO each "GGATCACCCCCACACCTGTG..." IN SEQUENCE-OF lactobacillus-casei-... Options Options
SEQUENCE-SIMILAR-TO each "GGATCACCCCCACACCTGTG..." IN SEQUENCE-OF lactobacillus-acidop... Options Options

>NIL, NO RESULTS FOUND!

Searching Lactobacillus-casei-ATCC-334's CRISPR to other organisms:

SEQUENCE-SIMILAR-TO each "GTTTTCCCCGCACATGCGGG..." IN SEQUENCE-OF lactobacillus-acidop... Options Options
SEQUENCE-SIMILAR-TO each "GTTTTCCCCGCACATGCGGG..." IN SEQUENCE-OF lactobacillus-brevis... Options Options

GGATCACCTCCACATACGTGGAGAAA (Lacto-A)

GGATCACCCCCACACCTGTGGGAATAC(Lacto-B)

GTTTTCCCCGCACATGCGGGGTGATCC (Lacto-C)

Although the Biobike functions did not find any matches, it seems like the first two sequences have some similarity. This might be somewhat significant but not so sure.

After analyzing the first three organisms of Lactobacillus, there are no commonly shared CRISPRs. In future studies, maybe more organisms needed to study using NCBI website. Also if COUNTS-OF-K-MERS.BIKE improves to find sequences that are bigger than the window size 30, it might be helpful. It is always possible to come up with own code to find the results rather than relying on some Biobike function.

References

Branger, C. CRISPR Distribution within the Escherichia coli Species Is Not Suggestive of Immunity-Associated Diversifying Selection. *Journal of Bacteriology*, 2460-2467.

Hugenholtz, P. CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, 181-186.