Insights from sequencing of *Lactobacillus rhamnosus LGG®* from commercial fermentations
Mette Jørgensen PhD.
Today’s key points

- Who are Chr Hansen
- The production process of bacteria
- The probiotic bacteria LGG®
- Stability of the LGG® genome during production
- Detection of ‘In sample’ variants
Chr. Hansen A/S in brief

- Founded in 1874 in Copenhagen by Danish pharmacist Christian D.A. Hansen
- Products include bacteria cultures, dairy enzymes and natural colors
- Producing starter cultures since the 1890s
- Traded on NASDAQ OMX Copenhagen since 2010
- Turnover in 2015/16 of €949 million
- Spend >7% of turnover on R&D
- State-of-the-art production facilities on five continents
- Estimate >1.2 billion people consume our products on any given day
Commercially Grown Bacteria

Danish men are 5th tallest in world
Bacteria face many stresses
- Acid
- Oxidative
- Temperature
- Starvation

16,000 L of bacteria

Batch culture (closed system) – liquid medium

Graph showing the growth phases of bacteria:
- Lag Phase
- Exponential Phase
- Stationary Phase
- Decline/death Phase

Number of cells vs. Time (hours)
Lactobacillus rhamnosus LGG®

Origin, Safety & probiotic effect

- Isolated from the microbiota of a human intestine in 1985
- Used worldwide for more than 20 years without safety concerns
- >1000 scientific publications
- Positive effect on general gastro intestinal (GI) health, eczema and diarrhea
- LGG® acquired by Chr. Hansen in 2016
Intra- and extracellular components of LGG\textsuperscript{®} interact with various components of the gastro intestinal (GI) tract to influence host health.
Deletions observed when LGG® is grown in stressful conditions for 1000 generations

Douillard et al 2016 Applied and Environmental Microbiology

Deletions observed in LGG isolated from commercial products

Sybesma et al 2013 Applied and Environmental Microbiology

Genomic instability of LGG® has been reported in some cases
My project: Genomic stability of LGG® through the production process

Sequenced all batches of produced LGG® over a year

- The culture bank
- The Inoculation material
  - 1 culture sample and 10 isolates
- 7 production batches
No large insertions or deletions were observed in any of the production batches.

It was shown that the spaCBA and other important genes were retained.
Mapping of reads

- >99% of the reads map to the reference
- Coverage is equally distributed along the genome
Zoom in on spaCBA region
The expected number of mutational events

- From Douillard et al. 2016 Applied and Environmental Microbiology
  - Mutation rate of $2.65 \times 10^{-9}$ SNPs or single base deletions per nucleotide per generation for LGG® under normal growth conditions.
  - That corresponds to 2.0 mutational events in the 7 batches.

- From Drake 1991 Proc. Natl. Acad Sci. USA
  - Without selective pressure mutations appear in a population at a rate of 0.0033 changes/genome/generation regardless of genome size.
  - That gives 0.83 expected mutations in the seven batches.
Internal sample variation

- Billion of genomes in each sample
- One mutational event only change one of them
- If the mutation gives a advantage to the bacteria the frequency of the mutation can raise
- We want to find positions in the genome that vary within a strain/sample
  - We call these positions ‘in sample’ variants

A sample can contain several clone types
Novel method for quantifying the frequencies of ’in sample’ variants

– I am using Cortex which is build on multi coloured de Bruin graphs
– Normally used for detecting variants between samples
– With custom cleaning thresholds and strict manual filtering it can also be used to estimate the frequencies of ‘in sample’ variants
– Validated with a simulated study and the method can reliable detect the frequency of variants present in down to around 5% of the population
– A complement to sequencing isolates
Simulated study frequencies

Variants and reads simulated with Pirs
Simulated study sensitivity
Simulated study false positives
Estimated batch to batch variation

– Only one new detectable subclone do emerge during the final fermentations
– This corresponds to Drakes estimate under no selective pressure
– The frequency is <1% and it is an intergenic snp
Conclusion

– Cortex can be used to estimate frequencies of ‘in sample’ variants
– The LGG® genome is stable in our production environment
– The pili encoding genes are retained during industrial production under our conditions
THANK YOU!

- Acknowledgements
  - HHN Discovery at Chr Hansen
    - Marianne Stage for the experimental work
    - The rest of the group for useful discussions
  - Identification at Chr. Hansen
  - The Krogh group at the Bioinformatics centre UCPH

Innovationsfonden