***M and M***

* Where are the isolates from (*Herds and sampling)*
  + - Isolates of S. chromogenes and S. simulans collected in 2014 to 2015 in a study on prevalence of NAS in bovine subclinical mastitis in Sweden (Nyman et al.,2018) were used. In short, NAS isolates identified in routine milk samples from udder quarters with subclinical mastitis sent to the National Veterinary Institute(Uppsala, Sweden) were included in the study
    - From that study, a total of 783 NAS isolates from 671 cows with subclinical mastitis were recovered from 201 herds. A total of 148 isolates were identified as *S. chromogenes* and 169 isolates as *S. simulans* using MALDI-TOF
  + SCC
  + Culture (*Isolates and NAS identification)*
  + Identification (MALDI etc.)
* Selection of isolates
  + For the current study…
  + For persistent *S. chromogenes* IMI which had 3 associated quarter-day observations, the middle isolate was submitted for WGS. For persistent *S. chromogenes* IMI which had 2 associated quarter-day observations, the first of 2 isolates was submitted 7 times and the second of 2 isolates submitted 10 times for WGS.
  + Matching
    - When possible, Isolates from the low group were matched to isolates in the high group within cow (different quarter) when possible; when matching for a given cow was not possible, isolates from both groups were matched on farm (or facility type if farm was not possible)
    - Isolates from the low group matched to the high group for the same cow when possible (different quarter), or same farm if same cow was not possible, or same facility type when same farm was not possible; when matched isolates were paired between low and high groups were for different cows, an isolates were matched on DIM and lactation number
  + RAPD
* Sequencing process from SeqCoast
  + *DNA extraction and whole-genome sequencing, assembly, and annotation*
    - *For each isolate, colony material was collected from the agar plates for DNA extraction for Gram-positive bacteria using Qiagen EZ1 DNA Tissue Kit (Qiagen).Nextera Library preparation (Illumina) and paired-end sequencing (2 × 150 bp) was then performed at Clinical Genomics Stockholm, SciLifeLab (Solna, Sweden) using an Illumina Novaseq 6000 instrument. The raw reads for each sample were quality checked using FastQCvo.11.9 (Andrews, 2010), trimmed using Trimmomaticvo.39 (Bolger et al., 2014) and assembled using SPAdesv3.14.0 (Prjibelski et al., 2020). The assemblies wereerror-corrected using Pilon v1.23 (Walker et al., 2014) and annotated using Prokka v1.12 (Seemann, 2014).*
* Bioinformatics process
  + *Bioinformatic analyses and in-silico analysis of AMR and VF*
  + Some from SeqCoast
  + MLST typing using the WGS fasta files with DTU online tool
    - Did PCR and traditional Sanger sequencing to confirm any novel alleles
    - Compared Sanger and WGS data to make sure that differences found were not sequencing errors
  + AMR
    - The different pipelines used? Databases searched?
    - Deduplication, cut-offs for percent identity? (look at Nobrega)
  + Virulence (*Identification of virulence factors in NAS genomes*)
    - The different pipelines used? Databases searched?
    - Deduplication, cut-offs for percent identity? (look at Naushad)
    - *Classification and distribution of virulence factors*
* Statistical analysis
  + Compared high and low SCC groups to see if systematically different
  + ST and SCC
    - Association of ST or CC with being in HIGH or LOW group
  + AMR (*Presence of antimicrobial resistance determinants)*
    - Association of AMR with ST or CC
    - Association between the presence of AMR and HIGH vs. LOW
  + Virulence
    - Association of virulence with ST or CC
    - Association between the presence of virulence factors and HIGH vs. LOW
* Data availability
  + The raw reads from ONT and Illumina for all 62 genomes are available under NCBI Bioproject accession number PRJNA1130504 (Table XX).
  + NCBI submission data for your 30 isolates, Bioproject PRJNA1130504 and Biosamples SAMN42232476 to SAMN42232505
  + "The data have been deposited with links to BioProject accession number PRJNA1130504 (Biosamples SAMN42232476 to SAMN42232505) in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/)."

***Results***

* Descriptive statistics for… farms?
  + Number of persistent chromogenes IMI
    - How many fell into low, high, how many went back and forth
      * 15 stayed high
      * 61 go back and forth
      * 60 stayed low
      * In total, 136 apparently persistent chromogenes infections
        + 91 for 3 visits
        + 45 for 2 visits
  + For LOW
    - 60 total
    - 45 three visits (135 isolates)
    - 15 for 2 visits (30 isolates)
  + For HIGH
    - 15 total
    - 3 for three visits (9 isolates)
    - 12 for 2 visits (24 isolates)
  + How many persistent chromogenes IMI were there after actually checking them with RAPD
    - 74/75
    - 198 isolates were RAPD typed
  + Isolates sequenced representative of 30 IMI
    - That the isolates come from 7 organic dairy farms in VT
    - 16 from BP, 14 from TS
    - 17 from persistent IMI associated with 2 visits
    - 13 from persistent IMI associated with 3 visits
* Descriptive statistics on isolates
  + Compare between HIGH and LOW group
    - Facility type
    - Farm
    - Lactation number
    - DIM
    - Lactation group
    - Physical location of quarter
    - SCC
  + Taking out, bc not normally dist., but wanted to keep:
    - The average parity and DIM of the cow from which the isolate originated was 2.5 (SD: 1.6) and 253.4 days (SD: 99) for the HIGH group, and 2.5 (SD: 1.5 days) and 206.8 days (SD: 111 days) for the LOW group, respectively.
* MLST and clusters (phylogeny)
  + How many ST were identified
  + Table showing number of isolates in each ST, if that ST got assigned to a cluster, num, high and low in each group
    - How many diff ST in high vs LOW
    - ST which are unique to HIGH vs. low?
  + How many were novel ST
  + How many clusters were formed
  + Link to phylogenetic tree in Supplemental
* blaZ carriage
  + Compare HIGH vs. LOW group
    - blaZ carriage did not differ between groups using Pearson’s Chi-squared (p = 0.70) (unconditional analysis)
    - blaZ carriage did not differ between groups using generalized linear mixed model, with outcome as blaZ+/- and herd as a random effect
    - 
    - *the dependent variable is simply a count of events, for example a number of animals that test positive for tuberculin in a group of animals and there is no denominator to consider (i.e. the number of animal-times at risk does not vary by study unit)*
  + Compare carriage by ST or CC

***Figures and tables***



**n = 9**

**n = 11**

**n = 6**

**n = 4**

**Figure XX.** Carriage of blaZ gene by SCC category.