Within the Streptococcus genus, only Streptococcus thermophilus is considered to be non-pathogenic due to its adaptation to the milk environment. Streptococcus macedonicus is also an intriguing streptococcal species since its most frequent source of isolation is dairy. Sequencing of S. macedonicus ACA-DC 198 genome was performed using a combination of 454 GS-FLX pyrosequencing and HiSeq 2000 Illumina sequencing. The hybrid assembly between 454 and HiSeq2000 data (>400X coverage) resulted in a contiguous scaffold of 2,320,374 bp and in a plasmid of 41,370 bp. The genome assembly was validated against a Rhizobium optical map of the S. macedonicus genome. Sequences were annotated with the Bddy and the RAST pipelines and manually curated using Kofam. Final corrections were made based on the quality assessment of the annotation using Genomic2R. We found 2,125 protein-coding genes on the chromosomes, 152 of which were identified as potential pseudogenes, indicating an ongoing genome decay process. This hypothesis is also supported by the approximately 220 kb smaller genome size of S. macedonicus compared to the S. pyogenes genomes, despite the high level of gene synteny between the two species. Such a reducible evolutionary process is common for lactic acid bacteria domesticated to the food environment, which in the case of S. thermophilus was also accompanied by the loss of pathogenicity traits. With this in mind, we also attempted to investigate whether S. macedonicus shows traits that would support its adaptation to the dairy environment at the genomic level.

Sequencing the genome of S. macedonicus ACA-DC 198

• 1st step: shotgun pyrosequencing with 454 GS-FLX titanium (100 contigs)
• 2nd step: 36X paired-end pyrosequencing with 454 GS-FLX titanium (7 scaffolds)
• 3rd step: gap-clouse and polishing with Illumina sequencing using the HiSeq 2000 (1 chromosome and 1 plasmid)
• 4th step: validation of the overall assembly (>300X coverage) with an Rhizobium optical map

Annotating the genome of S. macedonicus ACA-DC 198

• 1st step: initial annotation was performed with the Bddy and the RAST pipelines
• 2nd step: annotations were manually compiled in one using Kofam software
• 3rd step: final corrections and quality assessment was performed using Genomic2R (including predictions for potential pseudogenes)

Comparative genomics of S. macedonicus ACA-DC 198

• The complete genome sequence of S. macedonicus offered new opportunities to investigate the properties of the species at the genomic level.
• The inclusion of S. macedonicus and S. pneumoniae as subspecies of S. pyogenes has been previously suggested (Schnegel et al., J Biol 3 Synt Colloq, 2003), but this taxonomic reappraisal has not been formally accepted due to low DNA-DNA hybridization values (<70%) (Shaw et al. J Biol 3 Synt Colloq, 2002).

Our findings closely suggest that not only S. macedonicus, but also S. pneumoniae and S. infantarius have evolved from S. pyogenes in their potential to catalyse complex plant carbohydrates and to cope with the harsh environment of the GI tract of mammals. Further analyses of S. gallolyticus showed that it harbours the pil3 accessory pilin genes (pil1, pil2, pil3) which may induce its binding to the extracellular matrix (ECM), although variations of pilus gene presence/absence within strains have also been reported. Each gene cluster consists of three genes. The first two genes encode the adhesins belonging to the MDRSMF (membrane-related surface recognizing adhesive matrix molecules) family, being the major one and the one being the minor (or accessory) plus a third one that is related to the major adhesin (pil1, pil2, pil3). This gene cluster is followed by a set of paralogs. The pilus functions as a scaffold recognized by the third gene of the cluster. pil1 and pil2 loci are absent in S. macedonicus, S. pneumoniae and S. infantarius, indicating a diminished ability to adhere to ECM that could probably influence their ability to colonise host tissues and to produce infections when compared to S. pyogenes.

Conclusions

1. S. macedonicus is most probably a separate species from S. pyogenes
2. In silico analysis of S. macedonicus ACA-DC 198 suggests that:
   • The strain is at the process of adapting to a rich in nutrients environment
   • It shows a diminished capacity to live and survive in the GI tract of herbivores
   • It has a diminished pathogenic potential compared to S. pyogenes

Acknowledgments: The present work is financially by the European Social Fund and the National Resources EPEAEK and VESPID of the Thessaly program.