Identification of genomic suppressor mutations by high-throughput sequencing of yeast genomes

K. Arens, Y. Meier, T. Biswakarma, E. Hamza

Introduction

Ribosomes are fundamental macromolecular machines which allow the conversion of genetic information encoded within messenger RNAs (mRNA) into proteins [1]. In order to catalyse protein synthesis, ribosomes must be assembled from ribosomal RNA (rRNA) and proteins (r-proteins). The biogenesis of ribosomes is linked to other fundamental cellular processes, including growth and cell division. Recent studies have demonstrated that defects in ribosome biogenesis are associated with several hereditary diseases.

Ribosome biogenesis takes place successively in the nucleolar, nucleoplasmic, and cytoplasmic compartments. Most of the current knowledge concerning this highly dynamic multi-step process comes from studies with the yeast *Saccharomyces cerevisiae* which shows that mature ribosomes (80S) are composed of two subunits, the large 60S and the small 40S subunits. The 60S subunit comprises three rRNAs (25S, 5.8S, 5S) and 46 r-proteins, while the 40S subunit has one rRNA (18S) and 33 r-proteins [2]. The assembly and maturation of pre-ribosomal particles upon travel from the nucleolus to the cytoplasm, depends on a multitude (>200) of biogenesis factors. Among these are many energy-consuming enzymes such as GTPases, protein kinases, ATP-dependent RNA helicases and three AAA-type ATPases (ATPases associated with various cellular activities). This suggests that the energy derived by these enzymes is required for ribosomes assembly. The absence of one of these proteins might stall ribosome biogenesis and terminate cell growth even under optimal growth conditions [4]. To date, a role in ribosome assembly could be attributed to three essential AAA-ATPases, namely Rix7 (ribosome export), Rea1/Mdn1 (ribosome export associated/midasin) and Drg1/Afg2 (diazaborine resistance /ATPase family), which act at distinct steps during 60S subunit biogenesis in the yeast *S. cerevisiae*. Interestingly, all three AAA-ATPases promote the release of distinct biogenesis factors from nucleolar (**Nsa1** by **Rix7**), nucleolar and nucleoplasmic (**Ytm1-Erb1-Nop7 and Rsa4** by **Rea1**) and cytoplasmic (**several shuttling factors** by **Drg1**) pre-60S intermediates. The release of these factors from pre-60S particles ensures their recycling and likely triggers conformational changes that are critical determinants for the progression of ribosome assembly, e.g. promoting export or subunit-joining competence.

NSA1 (nucleolar protein seven-associated protein1) is a conserved WD repeat protein that is required for the yeast pre-60S ribosome assembly. *NSA1* has also been found up-regulated in mammalian cancer cells, suggesting it may also play a role in cell proliferation. *NSA1* null mutation (*Δnsa1)* leads to lethality. Rix7 interacts genetically with Nsa1 and is targeted to the Nsa1-defined pre-ribosomal particle. In vivo, Nsa1 cannot dissociate from pre-60S particles in r*ix7* mutants, causing nucleolar Nsa1 to escape to the cytoplasm, where it remains associated with aberrant 60S r-subunits [1].

The mechanistic details of the Rix7-mediated release of Nsa1 from pre-60S particles was studied [1] and revealed that the lethality associated with the absence of Nsa1 (*Δnsa1)* can be suppressed by mutant alleles of the biogenesis factors **Ebp2, Mak5, Nop1 and Nop4,** thus genetically unravelling possible structural changes within pre-60S particles that influence their assembly kinetics and/or path in order to compensate for the lack of Nsa1 recruitment. Moreover, genetic experiments indicate that Rix7 may have other nuclear substrates besides Nsa1 and that it may recognize and target defective pre-60S subunits for degradation.

The aim of the present study is to define novel suppressor genes using high-throughput sequencing.

Materials and Methods

Yeast strains and genetic methods

The S. *cerevisiae* strains used are derivatives of W303 or DS1-2b. For yeast two-hybrid analyses the reporter strain PJ69-4A was used. Preparation of media, yeast transformation, and genetic manipulations were done according to established procedures [1].

Plasmid constructs

Recombinant DNAs were prepared using *Escherichia coli* DH5α for cloning and plasmid propagation. All cloned DNA fragments were generated by PCR amplification and verified by sequencing. Human NSA1 was PCR-amplified from a cDNA library, cloned into pBSKS(-) (Stratagene) and then followed by a fusion PCR to eliminate mutations within the hNSA1 gene and further cloned under the control of the *ADH1* promoter in a YCplac111-based vector.

Isolation and cloning of spontaneous Δnsa1 bypass suppressors

The *NSA1* shuffle strain Y3900 was transformed with pADH111-hNSA1.N328, cells that have lost the *NSA1* shuffle plasmid pHT4467Δ-*NSA1* (*URA3-ADE3*) were selected and then spotted on YPD plates. To ascertain that the suppressor strains carried mutations in the genes complementing the growth and suppression phenotypes, *mak5*, *nop4* and *nop1* were amplified by PCR from genomic DNA of the respective suppressor strains and the type of mutation was identified by sequencing.

Synthetic lethal screens

Synthetic lethal (sl) screens with the *mak5.G218D* and *mak5.R728** alleles were performed as previously described. The *mak5* alleles were cloned and the resulting plasmids were transformed into the *MAK5* shuffle strains YDP1 and YDP4, which bear *MAK5* on the instable *URA3 ADE3* marker containing plasmids pHT4467Δ. Cells were grown to exponential phase, plated on SC-Trp plates and then mutagenized. Cloning of the genes complementing the respective sl-mutations was done using a yeast genomic library. The slscreen with *mak5.R728** yielded the following candidates whose complementing genes were cloned: six times *NOP16* (SL2, SL4, SL63, SL67, SL85 and SL113), two times *RPF1* (SL3 and SL33) and one time *RPL14A* (SL103). The sl-screen with *mak5.G218D* yielded one sl-candidate for which the complementing gene: *EBP2* (SL19) was cloned. To confirm that the sl-mutant carry mutations in the complementing genes, genomic DNA was prepared from these strains and the corresponding mutant alleles were amplified by PCR and sequenced.

Yeast genomic DNA isolation and sequencing

Genomic DNA libraries were prepared and sequenced on the Illumina (platform of University of Berne). The output Illumina files with base calls and qualities were converted into FASTQ file format.

Quality control of reads

The reads in FASTQ format were processed for quality checks using the FastQC tool. The statistics on the number of reads per library, base quality scores, read length, GC content, number of missing base calls, and number of unique 15-mers was collected separately for the first and second template end in each of the lanes.

Mapping and SNPs visualization

Data were filtered using sickle and good quality reads were then mapped to the reference genome (*S.cerevisiae* R64-1-1) using BWA mapper. The Sam file created from the mapping step was then converted to BAM using samTools which was also used to sort and index the BAM file. The SNPs and indels were then retrieved using samTools which generated VCF file format. This was followed by calling the SNP effect using SNPeff to narrow down the search space and allowing for identification of putative mutations. Finally, the SNPs were visualized (Figure 1) using the integrative genomics viewer (IGV) software (http://www.broadinstitute.org/igv).

Results

Seven yeast clones were examined 22a, 29b, 34a, 56a, 82b and 90a for the presence of mutations. The quality of the reads was high – the mean base quality score in the Phred scale [5] was 37.5 indicating that the base call accuracy was above 99.97%. The reads were characterized by high variation in the sequence content, and there were no indications of high level of duplication (on average 30.76%) or GC-bias (on average almost 37.166 % GC content per sample).

Figure 1: Visualization of the observed mutations and their positions on different chromosomes using Integrative genomic viewer (IGV) program.

Table1: Overview of the observed mutations arranged according to the clone numbers. The table describes types of mutations and the affected proteins. The important mutations are marked with bold. * indicates high alteration.

Clone	Chromosome / position	Type of mutation	Alteration (Nucleotide)	Alteration (A.A. / position)	Protein
22a	**XIII / 804604**	**stop gained high nonsense***	**G / A**	**Trp 50***	**TMA23**
22a	VII / 493762	moderate non-synonymous (n.s.)	G / A	Gly / Asp (138)	CDH1
29b	**XV / 724612**	**moderate n.s.**	**G / C**	**Gly /Arg (568)**	**DED1**
29b	X / 715176	moderate n.s.	C / T	Pro / Ser (189)	DAN4
29b	**VIII / 188042**	Stop gained **high nonsense***	**C / A**	**Ser40***	**BCD1**
34a	**XII /547161**	**stop gained high nonsense***	**G / C**	**Tyr355***	**NOP56**
56a	**VIII / 188891**	**frameshift high***	**deletion of G at position 970**	**Glu324 frameshift**	**BCD1**
56a	III / 209960	moderate n.s	G / A	Arg / Lys (155)	IMG1
82b	**XV / 252481**	**stop gained high nonsense***	**G / T**	**Glu 56***	**NOP12**
90a	**IV / 571870**	**moderate n.s.**	**C / A**	**Gln / Lys (408)**	**MAK21**
90a	X / 190760	moderate n.s.	G / C	Prol / Al (111)	RPE1
90a	XV / 70244	moderate n.s	A / G	His / Arg (101)	Yol134

1. Clone 22a shows a stop codon mutation in *TMA23* gene

As shown in table 1, clone 22a contains two mutations, one at position 804604 on chromosome XIII which encodes the region for ***TMA23* gene**. The mutation is of the stop codon high nonsense type which occurred at the G-patch domain of *TAM23* gene. The G-patch domain has been suggested to mediate an RNA-protein interaction [6]. The second mutation was found at position 493762 on chromosome VII encoding for *CDH1* gene and it is of the moderate non synonymous type which replaced glycine by aspartate and occurred at rADc domain.

2. Clone 29b contains two mutations in *BCD1* and *DED1* genes

Three mutations were found in Clone 29b, one is high nonsense and two are moderate non synonymous. The high nonsense took place in chromosome VIII at position 188042 which encodes the region of ***BCD1*** gene. This gene is known to be required for C/D-box **snoRNAs** accumulation and the mutation is located at Zinc finger domain.
One of the other two mutations was found on chromosome XV at position 724612 encodes for the ***DED1* gene** which is an **ATP-dependent RNA helicase.** The mutation lead to a change in glycine into arginine at position 568, no domain is yet known at that position. The second mutation occurred on chromosome X at position 715176 in the region of DAN4 gene which encodes for yeast cell wall protein.

3. Clone 34a contains one stop codon mutation in *NOP56* gene

The mutation occurred in Clone 34a is one of the important mutations. It took place on Chromosome XII at position 547161 in the region of ***NOP56*** gene and is one of the stop codon high non sense type which appeared at position 355 of the gene in the Nop domain.

4. Clone 56a and the presence of mutation in *BCD1* gene

The Clone 56a contains two mutations, one is of high effect and the other is of moderate non - synonymous. The mutation with high effect was influencing the ***BCD1*** gene and characterized by a deletion of guanine at position 970 which leads to a frameshift at Glu324.
The moderate non-synonymous mutation result in changing arginine into lysine on the region encoding for the 54S ribosomal protein *IMG1* on Chromosome III at position 209960. The helical turn of the SH3-like domain of this protein might be affected by that mutation.

5. Stop codon mutation of *NOP12* in Clone 82b

A stop codon mutation was found in Clone 82b in the region of ***NOP12*** gene on chromosome XV at position 252481. ***NOP12*** is a nucleolar protein involved in pre-25S rRNA processing. The mutation is located at the nucleotide binding alpha beta plait and RNA recognition motif domain.

6. Clone 90a contains mutation for *MAK21*

The Clone 90a was found to contain 3 non-synonymous moderate mutations. The most important one is at position 571870 on chromosome IV which encodes for the ***MAK21* (*NOC1*) gene.** ***MAK21*** is a constituent of 66S pre-ribosomal particles that is required for biogenesis of the large (60S). The domain affected is not well defined but thought to play a role in protein-protein interaction. The other two mutated genes were the *RPE1* and *YOl134* which encode for a sugar converter and yeast membrane proteins respectively.

Discussion

Our study reveals the presence of 12 mutations. Seven of them are important as they are located in regions influencing expression of nucleolar, ATP-helicase or ribosomal proteins. Clone 22a shows a stop codon mutation in the G-patch domain of *TMA23* gene. Tma23 is a nucleolar protein implicated in ribosome biogenesis via interaction with DNA polymerase which synthesizes ribosomal RNA. Interference of the expression of *TMA23* was previously shown to extend the chronological lifespan of yeast [7, 8]. This suggests that Tma23 might act as competitors for Nsa1 in ribosome biogenesis. We also found a mutation in *CDH1* associated with *TAM23* in clone 22a. Cdh1 is responsible for conserved m6(2)Am6(2)A dimethylation in 3'-terminal loop of 18S rRNA, part of 90S and 40S pre-particles in nucleolus, involved in pre-ribosomal RNA processing [4]. It is not clear whether the suppression of the absence of *NSA*1 is a result of a synergetic effect of the mutations in *TAM23* and *CDH1*.

The second mutation of a nucleolar protein gene was in *NOP56* that was also a gain of a stop codon. This was the only mutation detected in Clone 56a. Nop56 is required for assembly of the 60S ribosomal subunit and is involved in pre-rRNA processing. A possible explanation for the suppressive effect of that mutation is the participation of Nop56 in a regulatory mechanism for the 60S ribosomal subunit. This is supported by the fact that Nop56 interacts with Nop1and Nop58 and that Nop1 and Nop58 were found recently to be involved in suppression of the effect of *Δnsa1* [1].

The third nucleolar protein mutation was also a stop codon mutation detected in the RNA recognition motif of *NOP12* in clone 82b. Nop12 is involved in pre-25S rRNA processing and involved in ribosome biogenesis. We suggest that deletion of *NOP12* compensates for the lack of NSA1 via inhibiting $27SA_3$ pre-rRNA processing, resulting in the formation of an alternative yet unproductive helix 5 when cells are grown at low temperatures [7].

Furthermore, our analysis showed a mutation of the ATP-dependent RNA helicase *DED1* in clone 29b. Ded1 is involved in translation initiation. Mutations in *DED1* were found before to complement the *NSA1* deletion as their functions are related [1]. Moreover, Ded1 has been shown to interact with Nop58 which has a major role in pre-18s rRNA- processing. This might explain the survival of the clone carrying that mutation.

The *DED1* mutation was found together with a high nonsense at the Zink finger domain in *BCD1*. Bcd1 is required for C/D-box snoRNA accumulation involved in snoRNA processing. Therefore, a lack of Bcd1 might alter the functionality of the involved snoRNAs. This confirms previous study [1] showed that dysfunction of the components of C/D-box RNPs alters the RNP structure and or folding of pre-rRNA within pre 60S and hence allows efficient progression of downstream assembly in the absence of Nsa1. snoRNA act as guide to rRNA. However, it is still not clear whether a mutation in Bcd1 has a relation with a mutation in C/D-box snoRNA snR24 which was shown previously to suppress the effect of *Δnsa1* [1].

Interestingly, a frameshift mutation in *BCD1* was also found in clone 56b together with a mutation in the SH3-like domain of *IMG1*. Img1 is a 54S ribosomal protein. Several prokaryotic and eukaryotic proteins that are involved in the translation process contain an SH3-like domain.

An interesting mutation was in the *MAK21* gene in clone 90a which acts as part of a Mak21-Noc2-Rrp5 module that associates with nascent pre-rRNA during transcription and has a role in biogenesis of the large ribosomal subunit. It is also involved in nuclear export of pre-ribosomes. Accordingly, we suggest that the partial dysfunction of Mak21 results in conformational alterations in the rRNA during processing that increases the efficiency of other genes involved in ribosome biogenesis and hence compensates for the lack of Nsa1.

In conclusion, we have detected two new suppressor mutations in *TMA23* and *BCD1*. These two mutations were also found in 3 clones (S21a, S87b and S63as respectively) examined by group E. The presence of these two mutations in more than one clone highlights their importance in compensating for the absence of Nsa1. The striking finding is that the *BCD1* mutation in clone S63as from group E was also associated with *IMG1* mutation which then indicates a probable synergetic effect of these 2 mutations on compensating for the absence of Nsa1. Further studies are then required to investigate the biological role as well as the mechanism of action of *TMA23* and *BCD1*.

References

1. D. Pratte, U. Singh, G. Murat, D. Kressler (2013). Mak5 and Ebp2 act together on early pre-60S particles and their reduced functionality bypasses the requirement for the essential pre-60S factor Nsa1. Plos One 8 (12), e82741.
2. B. Strunki and K. Karbstein (2009). Powering through ribosome assembly. RNA 15:2083-2104.
3. D. Kressler, D. Roser, B. Pertschy and E. Hurt (2008). The AAA ATPase Rix7 powers progression of ribosome biogenesis by stripping Nsa1 from pre-60S particles. JCB 18(6): 935-944.
4. http://www.uniprot.org
5. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-Calling of Automated Sequencer Traces UsingPhred. I. Accuracy Assessment. Genome Res 8: 175–185.
6. M. Buchhaupt, P. Kötter & D. Entian (2007). Mutations in the nucleolar proteins Tma23 and Nop6 suppress the malfunction of the Nep1 protein. FEMS Yeast Res 7(6):771-81.
7. C. Burtner, C. Murakami, B.Olsen, B. Kennedy, M. Kaeberline (2011). A genomic analysis of chronological longevity factors in budding yeast. Cell Cycle 10(9):1385-96.
8. J. Talkish, I.Cambell, A.Sahasranaman, J.Jakovlietic, JLWoolford (2014). Ribosome assembly factors Pwp1 and Nop12 are important for folding of 5.8S rRNA during ribosome biogenesis in Saccharomyces cerevisiae. Mol cell biol 34 (10):1863-77.

De novo assembly and annotation of seven *Pediococcus acidilactici* strains

K. Arens, T. Biswakarma, E. Hamza, Y. Meier

Introduction

Cheese is an important consumption and export product for Switzerland. There exist about 450 types of cheese with unique properties and around half of the milk produced in Switzerland is processed into cheese [1,2]. There are five main steps in the manufacturing process of cheese: preparation of the milk, acidification, coagulation, dehydration and salting and pressing of the cheese. Bacterial cultures are used for the acidification of milk by producing lactic acid from the lactose present in the milk. Since the 1960's all the bacterial strains used in the process of making cheese in Switzerland were combined by Agroscope (Swiss institute for agriculture) into one national strain collection, which contains more than 10.000 strains and is used for food production [2]. Many of the bacterial strains are thought to produce peptides with antimicrobial properties (AMPs) of which those synthesized in the ribosomes are generally referred to as bacteriocins. Antibiotics are not included in the group of bacteriocins as they are not ribosomally synthesized. Due to their applications in food preservation, many studies have led to a growing range of potential bio-preservatives with most promising the pediocins [3]. Pediocin is produced by the Gram-positive lactic acid bacterial genus *Pediococcus* and contains a highly conserved N-terminal region that was proposed to be responsible for its anti-listerial activity [4].

The species *Pediococcus acidilactici* belongs to the pediocin producing *Pediococcus* genus although it is not known whether all species in this genus do produce pediocin. *P. acidilactici* is extracted from dairy products and is believed to have an effect on the bacterial community through bacteriocins. The aim of this study is to assemble and annotate seven *P. acidilactici* strains (strains: B1, B3, C3, D1, D3, E3 and G1, obtained from the national collection of Agroscope) in order to study whether pediocin is present in those seven strains and/or if there are any other bacteriocins produced.

Materials and Methods

Bacterial strains

The pediococcus acidilactici strains used were obtained from the national strain collection of Agroscope (Switzerland).

Illumina sequencing

Genomic DNA libraries were prepared and sequenced on the Illumina (platform of University of Berne). Genomic DNA libraries were prepared using Nextera XT DNA Sample Preparation Kit (www.illumina.com). The input DNA was tagmented (fragmented and tagged with adapter sequences) by the Nextera XT Transcriptome. The tagmented DNA was then amplified via limited-cycle PCR program, followed by PCR-Clean-up to purify the library DNA and remove very short fragments. The quantity of each library was further normalized to ensure equal library representation in the pooled sample using normalization beads. The libraries were then pooled and loaded in Miseq.

Assembly and annotation of sequencing reads

The reads in FASTQ format were processed for quality checks using the FastQC tool. Based on these results the reads were trimmed and filtered using sickle. The flash tool (Fast Length Adjustment of Short reads, http://ccb.jhu.edu/software/FLASH) was used, as a pre-processor for genome assembly , to merge paired-end reads when partially overlapping. Subsequently the reads were assembled and scaffolded with the help of SPAdes.

The quality of the assemblies was validated with the QUAST online tool (http://quast.bioinf.spbau.ru/). The best assembly for each strain was subsequently annotated with

prokka (for prokaryotic genomes, http://www.vicbioinformatics.com/software.prokka.shtml). After annotation the strains were screened for the presence of pediocins and other bacteriocins (e.g. nisin, lacticin, enterocin and colicin). Contigs (>500bp) of all strains were aligned against the "best" of those strains (the strain with the least number of contigs) as reference genome using Mauve software in order to visualize differences/similarities in genome arrangement.

Results

Read quality for all seven strains was improved by trimming some of the sequences due to relatively short insert sizes. Assembly of the reads resulted in a minimum of 12 (B3) to a maximum of 61 (E3) contigs after scaffolding (Table 1). The assemblies have an average of ± 35 times coverage and the sum of the contigs corresponds to an average size of 1.9 Mb as is expected for bacterial strains.

those

Table 1. Quality statistics of SPAdes scaffolds.

Strain	n (contigs)a	n:N50	min	median	mean	N50	max	sum
B1	31	4	254	13680	58605	213230	271156	1816778
B3	13/12	4	254	105297	151408	291272	339732	1968305
C3	38/37	6	254	28374	52158	104447	249040	1982023
D1	40/39	5	254	13681	45418	160162	223986	1816756
D3	42/41	6	254	21664	46316	84625	309842	1945299
E3	62/61	9	254	16479	31403	64080	235784	1947012
G1	46/41	5	238	9437	43773	138241	394488	2013572

a Number of contigs present in SPAdes scaffold/number of contigs after Quast validation.

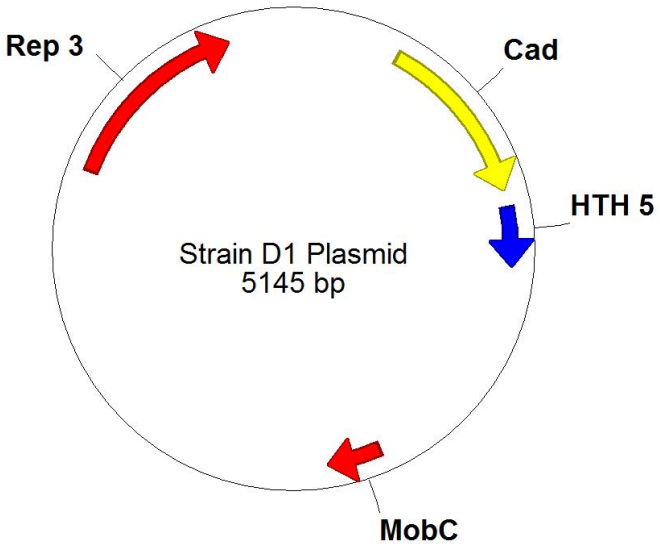

Figure 1. Illustration of the plasmid on the D1 strain, thought to be involved in cadmium resistance. Annotation based on the PFAM database. Cad: Cadmium resistance transporter family, HTH 5: bacterial regulatory protein domain from the arsR family, MobC: bacterial mobilisation protein family and Rep 3: initiator replication protein family.

Figure 2. Alignment of the seven *P. acidilactici* strains (in Mauve) with B3 as the reference strain. Contigs <500bp were excluded. Similarity is depicted in gray within the blocks from 0-100%.

The seven *P. acidilactici* strains were aligned against strain B3 as a reference (the strain with the least contigs) and visualized in Mauve (Figure 2). Contigs shorter than 500bp were excluded from the alignment. The results of the alignment show that although most blocks can be found back on all strains with or without some dissimilarities. Moreover, some rearrangement of the blocks along the chromosome, including some inversions (examples in red box) were found. From the annotated assemblies we could conclude that no (putative) pediocins were present in any of the seven strains. For strain B1 no other bacteriocins were present either, however for the other six strains (B3, C3, D1, D3, E3 and G1) a Colicin V production protein (cvpA) was found. cvpA was first found in *E. coli* [5], consists out of 162 amino acids and has 4 transmembrane (TM) domains. The proteins found in the six *P. acidilactici* strains were blasted against the NCBI database where they aligned with the predicted cvpA gene in the *P. acidilactici* strain 7-4 with E-values between $3e^{-109}$ to $4e^{-111}$. The cvpA gene from *P. acidilactici* 7-4 is 175 amino acids long of which only the first 13 amino acids are not found back in our six strains (Figure 3).

```
Sbjct  14   AFFSGYRRGLVAEIIYLVGYLIVFAAAKNFTAPFAEFLSRTFGNGSHDPLTNLTTMNAVS  73
G1     489  AFFSGYRRGLVAEIIYLVGYLIVFTAAKNFTAPFAEFLSRTFGNGSHDPLTNLTTMNAVS  310
B3     40   AFFSGYRRGLVAEIIYLVGYLIVFAAAKNFTAPFAEFLSRTFGNGNHDPLTNLTTMNAVS  219
C3     40   AFFSGYRRGLVAEIIYLDGYLIVFAAAKNFTAPFAEFLSRTFGNGNHDPLTNLTTMNAVS  219
D1     489  AFFSGYRRGLVAEIIYLVGYLIVFAAAKNFTAPFAEFLSRTFGNGSHDPLTNLTTMNAVS  310
D3     40   AFFSGYRRGLVAEIIYLDGYLIVFAAAKNFTAPFAEFLSRTFGNGNHDPLTNLTTMNAVS  219
E3     40   AFFSGYRRGLVAEIIYLDGYLIVFAAAKNFTAPFAEFLSRTFGNGNHDPLTNLTTMNAVS  219

Sbjct  74   FIFLMLVGWIVIRLIVRFSQMITWLPIIHQVNGLAGAVAGFVISYLITFIVLSISQFVPN  133
G1     309  FIFLMLVGWIVIRLIVRFSQMITWLPVIHQVNGLAGAVAGFVISYLITFIVLSISQFVPN  130
B3     220  FIFLMLVGWIAIRLIVRFSQMITWLPIIHQVNGLAGAIAGFVISYLITFIVLSISQFVPN  399
C3     220  FIFLMLVGWIAIRLIVRFSQMITWLPIIHQVNGLAGAVAGFVISYLITFIVLSISQFVPN  399
D1     309  FIFLMLVGWIVIRLIVRFSQMITWLPVIHQVNGLAGAVAGFVISYLITFIVLSISQFVPN  130
D3     220  FIFLMLVGWIAIRLIVRFSQMITWLPIIHQVNGLAGAVAGFVISYLITFIVLSISQFVPN  399
E3     220  FIFLMLVGWIAIRLIVRFSQMITWLPIIHQVNGLAGAVAGFVISYLITFIVLSISQFVPN  399

Sbjct  134  DFYQEQLSQSIVAQSILAKTPAISSKVINNYILDTPQTRDVL  175
G1     129  DFYQEQLSQSIVAQSILAKTPAISSKVINNYILDTPQTRDVL  4
B3     400  DFYQEQLSQSIVAQSILAKTPAISSKVINNYILDTPQTRDVL  525
C3     400  DFYQEQLSQSIVAQSILAKTPAISSKVINNYILDTPQTRDVL  525
D1     129  DFYQEQLSQSIVAQSILAKTPAISSKVINNYILDTPQTRDVL  4
D3     400  DFYQEQLSQSIVAQSILAKTPAISSKVINNYILDTPQTRDVL  525
E3     400  DFYQEQLSQSIVAQSILAKTPAISSKVINNYILDTPQTRDVL  525
```

Figure 3. Alignment of cvpA (colicin) gene of the seven studied P. acidilactici strains with the 7-4 strain as subject. The subject is labelled in amino acid number, while the other strains are labelled with the nucleotide coordinates as they are the result of a blastx search in the NCBI database. The cvpA gene on strains D1 and G1 are found in inverted orientation.

Discussion

Overall the seven assemblies are of good quality, ranging from 12-61 contigs with an average of ± 35 fold coverage. The seven annotated files were screened for pediocins and other bacteriocins as they might have a big influence on the taste of the cheese by eliminating other bacteria and on the shelf-life of cheese. However, no pediocin gene was found in any of the seven strains. Pediocins have been found before in the *Pediococcus* family and even in some *P. acidilactici* strains, for example pediocin AcH and pediocin PA-1. Pediocins are known to be plasmid localized [3], however as we found a plasmid on strain D1, conferring cadmium resistance, it can be concluded that there are no pediocins present in these seven strains as plasmids are correctly sequenced. The only bacteriocin found was colicin V, which was detected in six of the strains (B3, C3, D1, D3, E3 and G1). All of them are highly similar to the colicin V production protein previously predicted in the *P. acidilactici* strain 7-4. With the exception of strain B3, which has no amino acid exchange, all strains have only one or two exchanges in amino acid sequence. The amino acid exchanges might have a possible effect on the protein structure and functionality as they are all non-synonymous exchanges. They are different in size and/or expected location (e.g. helix, strand or turn). However, the colicin V protein from *P. acidilactici* 7-4 has not been functionally characterized yet either. As for none of the *P. acidilactici* colicin V protein has been extracted so far, the colicin V production proteins found in our strain can only be annotated as hypothetical protein. With the characterized colicin V production protein from *E. coli* it has been shown through pull down studies that it interacts with ribD (Riboflavin biosynthesis protein)[6] and so far no other interactors are known. RibD is a hydrolase responsible for the biosynthesis of vitamin B2, which plays an important role in humans in the release of energy from carbohydrates, protein and fats. The interaction of colicin V with ribD in Bacteria suggests that either colicin might attack other bacteria throught ribD and hence reduce energy release and kill the bacteria, or hydrolysis of Vitamin B3 via rib D. Furthermore, rib D might also be involved in the final processing steps for colicin V production. The strains C3, D1, D3, E3 and G1 have the colicin V production protein and the ribD protein, while strain B3 has only the colicin V production protein and B1 lacks both colicin V production protein and ribD. Based on the hypothesis chosen for the interaction of the two protein strain B3 might either be protected from degradation by colicin V or it might not be able to produce a fully functional colicin V protein.

Based on the colicin V alignment, but mainly the alignment of all strains against strain B3 as a reference it can be concluded that the seven strains are highly similar. However certain regions of the genome are rearranged between several strains and in some cases entire regions are inverted or missing. Overall the strains are thus highly similar as they are from the same species, but still have some differences between their genomic sequences.

References

9. Switzerland cheese marketing: http://www.cheesesfromswitzerland.com/
10. Agroscope – cheese quality and authencity: http://www.agroscope.admin.ch/lebensmittelqualitaet/05853/index.html?lang=de
11. Papagianni, M. and Anastasiadou, S., 2009; *Pediocins: The bacteriocins of Pediococci. Sources, production, properties and applications*. Microbial cell factories 2009, volume 8, issue 3. doi:10.1186/1475-2859-8-3
12. Miller K., Schamber R., Osmangaoglou O., Ray B., 1998; *Isolation and characterization of pediocin acH chimeric protein mutants with altered bactericidal activity*. Applied environmental microbiology, volume 64 pages 1997-2005
13. Fath, M. J., Mahanty H. K., Kolter, R., 1989; *Characterization of a purF operon mutation which affects colicin V production*. Journal of Bacteriology 171 (6):3158-61
14. Butland, G. *et al*., 2005; *Interaction network containing conserved and essential protein complexes in Escheichia coli*. Nature, 433 (7025):531-7