VILNIUS UNIVERSITY

ALEKSANDRAS KONOVALOVAS

Molecular determinants of *Totiviridae* family viruses of *Saccharomyces sensu stricto* clade

Summary of doctoral dissertation
Physical sciences, biochemistry (04 P)

Vilnius, 2018

The work presented in this doctoral dissertation has been carried out at the Institute of Biosciences, Life Sciences Center, Vilnius University during 2013–2018

Supervisor:

prof. dr. **Saulius Serva** (Vilnius University, physical sciences, biochemistry – 04 P)

The dissertation is defended at the Council of Biochemistry Sciences of Vilnius University:

Chairman – prof. habil. dr. **Kęstutis Sasnauskas** (Vilnius University, physical sciences, biochemistry – 04 P)

Members:

dr. **Rūta Gerasimaitė** (Max Planck Institute for Biophysical Chemistry, physical sciences, biochemistry – 04 P)

dr. **Laura Kalinienė** (Vilnius University, physical sciences, biochemistry – 04 P)

doc. dr**. Eglė Lastauskienė** (Vilnius University, biomedical sciences, biology – 01 B)

prof. dr. **Arvydas Lubys** (Thermo Fisher Scientific, physical sciences, biochemistry – 04 P)

The dissertation will be defended at the public session of the Council of Biochemistry Sciences at 3:00 pm on 28th of September, 2018, in the R401 auditorium at the Life Sciences Center, Vilnius University (Sauletekio av. 7, LT-10257, Vilnius, Lithuania)

The summary of doctoral dissertation was sent on 28th of August, 2018

This dissertation is available at the Library of Vilnius University and at the website of Vilnius University: http://www.vu.lt/naujienos/ivykiu-kalendorius

VILNIAUS UNIVERSITETAS

ALEKSANDRAS KONOVALOVAS

Mielių *Saccharomyces sensu stricto Totiviridae* šeimos virusų ciklo ląstelėje molekulinės determinantės

Daktaro disertacijos santrauka
Fiziniai mokslai, biochemija (04 P)

Vilnius, 2018

Disertacija rengta 2013 – 2018 metais Vilniaus universiteto Gyvybės mokslų centro Biomokslų institute Biochemijos ir Molekulinės Biologijos katedroje

Mokslinis vadovas:

prof. dr. **Saulius Serva** (Vilniaus universitetas, fiziniai mokslai, biochemija – 04 P)

Disertacija ginama viešame disertacijos gynimo tarybos posėdyje:

Pirmininkas – prof. habil. dr. **Kęstutis Sasnauskas** (Vilniaus universitetas, fiziniai mokslai, biochemija – 04 P)

Nariai:

dr. **Rūta Gerasimaitė** (Max Planck biofizikinės chemijos tyrimų institutas, fiziniai mokslai, biochemija – 04 P)

dr. **Laura Kalinienė** (Vilniaus universitetas, fiziniai mokslai, biochemija – 04 P)

doc. dr. **Eglė Lastauskienė** (Vilniaus universitetas, biomedicinos mokslai, biologija – 01 B)

prof. dr. **Arvydas Lubys** (Thermo Fisher Scientific, fiziniai mokslai, biochemija – 04 P)

Disertacija ginama viešame Gynimo tarybos posėdyje 2018 m. rugsėjo mėn. 28 d. 15 val. Vilniaus universitete, Gyvybės mokslų centre, R401 auditorijoje. Adresas: Saulėtekio al. 7, LT-10257, Vilnius, Lietuva

Disertacijos santrauka išsiuntinėta 2018 m. rugpjūčio 28 d.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: https://www.vu.lt/naujienos/ivykiu-kalendorius

CONTENT

LIST OF ABBREVIATONS

aa – amino acids

bp – base pair

BSA – bovine serum albumin

DMSO – dimethyl sulfoxide

DTT – dithiothreitol

dsRNA – double-stranded RNA

EDTA – ethylenediaminetetraacetic acid

EVE – endogenous virus element

LC-MS/MS – liquid chromatography tandem-mass spectrometry

NatC – N-terminal acetyltransferase complex

NRC – Nature Research Centre (Vilnius, Lithuania)

nt – nucleotide

PAGE – polyacrylamide gel electrophoresis

PBS – phosphate buffer saline

PCR – polymerase chain reaction

PEG – polyethylene glycol

PMSF – phenylmethylsulfonyl fluoride

RdRp – RNA depended RNA polymerase

SDS – sodium dodecyl sulfate

ssRNA – single-stranded RNA

INTRODUCTION

Saccharomyces cerevisiae yeasts are among the most important organisms in biochemistry, molecular biology, cell biology and systems biology studies. With a wide selection of different genetic tools developed*, S. cerevisiae* is often used as a model organism for the study of eukaryotic cell processes. An important data were obtained in the studies of *S. cerevisiae* yeast on the metabolism of lipids in human cells (Petranovic *et al.*, 2010), mitochondrial metabolism and mitochondrial diseases (Rinaldi *et al.*, 2010), eukaryotic cell aging processes (Barros *et al.*, 2010), about prions (Wickner *et al.*, 2018) and prion-induced diseases mechanisms (Fruhmann *et al.*, 2017). By using *S. cerevisiae* yeast as a model organism, large number of host cell proteins important for the plant RNA viruses replication and transcription was discovered (Nagy and Pogany, 2012). Paradoxically, the interactions of genuine yeast L-A virus with host cell proteins has not been studied until now.

The beginning of yeast virology dates back to 1963, when Bevan and Makower found that certain yeast strains possess biocidal activity against other yeasts, so called killer phenotype. Subsequent studies have shown that this phenotype in the *S. cerevisiae* is caused by secretable protein, or killer toxin encoded by the *Totiviridae* virus family double-stranded RNR (dsRNA) L-A virus M satellite. *S. cerevisiae* viruses have no extracellular stage and thus are non-infectious. Only one L-A virus variant was known for a long time; however, over the last few years several new L-A variants were discovered in *Saccharomyces sensu stricto* clade (Rodríguez-Cousiño *et al.*, 2013; Konovalovas *et al.*, 2016; Rodríguez-Cousiño and Esteban, 2017; Rowley *et al.*, 2016).

Research aims

1. Identify new genome sequences of dsRNA viruses from *Saccharomyces sensu strico* clade and carry out the bioinformatic analysis.
2. Identify host proteins interacting with ScV-LA-1 virus-encoded proteins and determine the significance of these proteins for the cycle of the virus in the cell.

Scientific novelty

In this work, we are presenting an universal strategy which we developed to obtain full length L-A and M dsRNA genomes from yeast. We have identified previously undescribed L-A virus variants from *S. cerevisiae*. Moreover, we identified L-A viruses from *S. paradoxus* for the first time. By analyzing the sequences of newly detected and published dsRNA L-A virus-encoded proteins, we identified highly variable motifs. We detected that they are located in the spatial area near the virus replication and transcription regions. We have determined that the C-end of Gag protein of L-A virus is necessary for the formation of viral particles capable to replicate the virus. In this work, we identified 15 proteins important for virus replication, including those 13 interacting directly with virus-encoded proteins, for the first time.

Research significance

In this research, we have discovered that L-A virus is found not only among yeast from natural environment, but also is widespread among the most popular laboratory strains of *S. cerevisiae* yeast used in scientific research. The presence of L-A virus is not associated with changes in cell viability or growth rate, however the data obtained in recent years show that the presence of the virus in the cell changes a large part of the host's gene expression (McBride *et al.*, 2013; Lukša *et al.*, 2017). Certain gene deletions can remove the virus from the host cell (Tercero and Wickner, 1992), others can activate virus replication and translation at incredible rates, amount of viral protein in the yeast cell reaching up to 20% of all cellular proteins synthesized (Dihanich *et al.*, 1989). Transcriptomic or proteomic analysis data may be misinterpreted in the presence of a virus, because observed changes in gene expression may not necessarily be related to the studied process, but also may be related to the change in the amount and existence of the virus in the cell. Only a few proteins from the host cell interacting with dsRNA L-A virus have been described. Most often, they were identified accidentally by doing research on different cell processes (Chong *et al.*, 2004). In this work we have found new, unpublished host cell proteins interacting with proteins encoded by L-A virus. We have identified new L-A virus M satellite genomes from *S. paradoxus* killer strains. We have found the direct relationship between the phylogenetic groups formed by different variants of the L-A virus and the host species. Additionally, distribution in phylogenetic groups is based on strain origin and ability to provide a specific M satellite replication.

Findings presented for defense

1. Previously undescribed L-A variants along with brand new M satellite from different yeast strains was discovered by using universal strategy to obtain full length dsRNA virus genome sequences from *Saccharomyces sensu stricto* clade.
2. The variability of dsRNR L-A viruses from different *Saccharomyces sensu stricto* strains is the consequence of host and virus coevolution.
3. The insertion of M satellite into chromosome may be related to the activity of transposons in the yeast cell.
4. The C-terminal motif of the L-A virus Gag protein is significant for the dsRNR virus genome replication in the nucleocapsid.
5. 15 host cell proteins were identified modulating the level of replication of the dsRNA, among whom 13 directly interacting with the proteins encoded by L-A virus and two, involved in the a signal transduction in the cell.
6. Biogenesis of Gag protein likely involves its transport to the nucleus.

1 MATERIALS AND METHODS

1.1 Materials

1.1.1 Chemicals

All chemicals used in this study were of the highest available quality (molecular biology grade).

1.1.2 Enzymes

Restriction endonucleases, FastDigest restriction endonucleases, alkaline phosphatase FastAP, T4 polynucleotide kinase, T4 DNA ligase, DNaseI, T4 RNR ligase, Maxima reverse transcriptase, Phusion and DreamTaq DNA polymerases, DNA O'GeneRuler Ladder Mix, PageRuler Unstained Protein Ladder used in this study were obtained from Thermo Fisher Scientific and were used according to the manufacturer's protocols.

1.1.3 Molecular biology kits

GeneJet Gel Extraction Kit, GeneJet Miniprep Kitwere obtained from Thermo Fisher Scientific and were used according to the manufacturer's protocols.

1.1.4 Growth media

Table 1.1. Growth media used in this study.

Medium	Composition
LB	1 % peptone; 0,5 % yeast extract; 0,5 % NaCl. For preparation of agarized media, 2 % of agar was added.
YPD	2 % peptone; 1 % yeast extract; 2 % glucose. For preparation of agarized media, 2 % of agar was added.
SD	0,67 % yeast nitrogen base; 2 % glucose. According to the need, uracil (10 μg/ml) and amino acids: L-leucine (60 μg/ml), L-metionine (10μg/ml), L-histidine (10 μg/ml) has been added. For preparation of agarized media, 2 % of agar was added.
MD	0,5 % peptone; 0,5 % yeast extract; 2 % glucose; 1,05 % citric acid; 3,53 % Na_2HPO_4 x $12H_2O$; 0,003 % methylene blue; 2 % agar.

1.1.5 Microorganism strains

Table 1.2. Microorganism strains used in this study.

Strain name	Species	Genotype	Reference
DH10B	*E. coli*	F^- *endA1 deoR*$^+$ *recA1 galE15 galK16 nupG rpsL* Δ*(lac)X74* φ80*lacZΔM15 araD139* Δ*(ara,leu)7697 mcrA* Δ*(mrr-hsdRMS-mcrBC)* Str^R λ^-	Invitrogen
BY4741	*S. cerevisiae*	MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0, ScV-LA-1*	Brachmann et al., 1998

BYΔLA	*S. cerevisiae*	MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0, LA-0*	This work
BYER	*S. cerevisiae*	MATa *leu2Δ0 met15Δ0 ura3Δ0 sec61-EGFP::HIS3, ScV-LA-1*	This work
α′1	*S. cerevisiae*	MATα, *leu2-2*	Čitavičius and Inge-Večtomov, 1972
21PMR	*S. cerevisiae*	MATα, *leu2 ura3-52*	Jokantaitė and Melvydas, 1996
K7	*S. cerevisiae*	MATa, *arg9, kil-K1*	Somers and Bevan, 1969
M437	*S. cerevisiae*	ScV-LA-lus, M-2	Naumova and Naumov, 1973
Rom K-100	*S. cerevisiae*	ScV-LA-2, M-2	Jokantaitė et al., 1982
KE	*S. cerevisiae*	ScV-LA-lus, M-lus	Dr. E. Servienė, unpublished
K28	*S. paradoxus*	SpV-LA-28, M-28	Pfeiffer and Radler, 1984
64b	*S. paradoxus*	SpV-LA-64b, M-64b	Dr. E. Servienė, unpublished
66	*S. paradoxus*	SpV-LA-66, M-66	Dr. E. Servienė, unpublished

BY4741 strain knock-out collection, where single ORF are replaced by KanMX4 module, was obtained from Open Biosystems.

1.1.6 Oligonucleotides

Table 1.3. Oligonucleotides used in this study.

Name	**Sequence (5′ → 3′)**	**Remarks**
AK-14+P	Ggatcccgggaattcggtaatacgactcactatatttttatagtgagtcgtatta	Primer used for ligation to dsRNA(Potgieter *et al.*, 2009)
AK-15	ccgaattcccgggatcc	Primer used toamplify cDNA of dsRNA genome (Potgieter *et al.*, 2009)
AK-60	cgccaaggaaggtgggtttactaagaacctcgttccaggattttctgatttgatgcgtacgctgcaggtcgac	Primers used for BYER strain construction
AK-61	gtgtggctaaatgcgatttttttttttctttggatattatttt catttttatattcaatcgatgaattcgagctcg	
AK-62	ccgggctagctccagcaccagcaccagc	Primers used to amplify pYM44 vector
AK-63	gactcgagtgcgcgccacttctaaataagc	
AK-65	cactcgagttacttgtacagctcgtccatgc	Primers used to amplify mCherry protein coding gene
AK-66	gagcggccgcgtccggagcaggtgctggtg	
AK-67	tcacgcggccgcctctactaaaacattgtcc	
AK-69	ctactagttaaaaatgctgaagcttcgtacgctgc	
YCFH-AK-1	gatccatggcggccgctgattacaaagacgatgacgataagggttctggtggatctcatcatcatcatcatcatta	FLAG and His tags containing adaptor
YCFH-AK-2	ggcctaatgatgatgatgatgagatccaccagaacccttatcgtcatcgtctttgtaatcagcggccgccatg	
YCFH-AK-3	Gcggatccttaaaaatgttgagattcgttaccaaaaactctcaag	Primer used to amplify *gag* gene

1.1.7 Plasmids

Table 1.4. Plasmids used in this study.

Plasmid	Description	Reference
pUC19	Cloning vector	Thermo Fisher Scientific
pUC19-k7L	DNA copy of dsRNA L-A virus genome from *S. cerevisiae* K7 strain cloned into pUC19 vector	This work
pUC19-romL	DNA copy of dsRNA L-A virus genome from *S. cerevisiae* Rom K-100 strain cloned into pUC19 vector	This work
pUC19-m437L	DNA copy of dsRNA L-A virus genome from *S. cerevisiae* M437 strain cloned into pUC19 vector	This work
pUC19-m4L	DNA copy of dsRNA L-A virus genome from *S. cerevisiae* KE strain cloned into pUC19 vector	This work
pUC19-64bL	DNA copy of dsRNA L-A virus genome from *S. paradoxus* 64b strain cloned into pUC19 vector	This work
pUC19-66L	DNA copy of dsRNA L-A virus genome from *S. paradoxus* 66 strain cloned into pUC19 vector	This work
pUC19-28L	DNA copy of dsRNA L-A virus genome from *S. paradoxus* K28 strain cloned into pUC19 vector	This work
pUC19-k7M	DNA copy of dsRNA M satellite genome from *S. cerevisiae* K7 strain cloned into pUC19 vector	This work
pUC19-romM	DNA copy of dsRNA M satellite genome from *S. cerevisiae* Rom K-100 strain cloned into pUC19 vector	This work
pUC19-m437M	DNA copy of dsRNA M satellite genome from *S. cerevisiae* M437 strain cloned into pUC19 vector	This work
pUC19-m64bM	DNA copy of dsRNA M satellite genome from *S. paradoxus* 64b strain cloned into pUC19 vector	This work
pUC19-m66M	DNA copy of dsRNA M satellite genome from *S. paradoxus* 66 strain cloned into pUC19 vector	This work
pYAK3	Yeast expression vector with constitutive promoter TEF1	This work
pYAK3-GagWT+CFH	ScV-LA-K7 Gag fused with C-terminal FLAG-tag and HIS-tag cloned into pYAK3 vector	This work
pYAK3-LAcds	ScV-LA-K7 genome cloned into pYAK3 vector	This work
pYAK3-GagWT	ScV-LA-K7 full length Gag protein coding sequence cloned into pYAK3 vector	This work
pYAK3-GagΔ	Truncated ScV-LA-K7 virus Gag version (1-642 aa) cloned into pYAK3 vector	This work
pYAK3-GagWT+mCherry	Sequence coding Gag protein with C-terminal mCherry tag cloned into pYAK3 vector	This work
pYAK3-GagWT-CFS-mCherry	This plasmid coding full length Gag protein and Gag protein fused with mCherry formed by ribosomal frameshifting	This work
pYAK3-mCherry	mCherry coding sequence cloned into pYAK3 vector.	This work

1.1.8 Solutions

Table 1.5. Solutions used in this study.

Solution	Composition
TB buffer	50 mM Tris-HCl (pH 9,3), 1% β-mercaptoethanol
TES buffer	10 mM Tris-HCl (pH 7,5), 100 mM NaCl, 10 mM EDTA, 0,2 % SDS
PBS buffer	80 mM Na_2HPO_4, 20 mM NaH_2PO_4 (pH 7,4), 200 mM NaCl

1xTE buffer	10 mM Tris-HCl (pH 7,5), 1 mM EDTA
1xLiAc/0.5xTE solution	100 mM lithium acetate, 5 mM Tris-HCl (pH 7,5), 0,5 mM EDTA
1xLiAc/40%PEG/1xTE solution	100 mM lithium acetate, 40 % PEG-4000, 10 mM Tris-HCl (pH 7,5), 1 mM EDTA
"Sodium" solution	5 mM Tris-HCl (pH 8,0), 100 mM NaCl, 5 mM $MgCl_2$
"Calcium" solution	5 mM Tris-HCl (pH 8,0), 100 mM $CaCl_2$, 5 mM $MgCl_2$
NaOH/SDS solution	100 mM NaOH, 1 % SDS
Yeast lysis buffer (for glass beads)	50 mM Tris-HCl (pH 7,4), 200 mM NaCl, 1 mM PMSF
Yeast lysis buffer (for liquid nitrogen)	PBS (pH 7,4), 2 mM PMSF, 1 mM EDTA
5 x Protein loading dye	0,313 M Tris-HCl (pH 6,8), 10 % SDS, 0,05 % bromophenol blue, 50 % glycerol, 100 mM DTT

1.2 Methods

1.2.1 DsRNA extraction

Yeast strains were cultivated in YPD medium for 20 hours at 30 °C. Cell were harvested by centrifugation at 6000 × g for 5 min at room temperature. After removing of the supernatant, the pellet is resuspended in 4 ml of 50 mM EDTA solution. Cells were pelleted by centrifugation under the same conditions. The cells were resuspended in 4 ml TB solution and incubated for 15 min at room temperature. Cells were pelleted by centrifugation under the same conditions. Cells were resuspended in TES solution and then "hot-phenol" (pH 5.2) was added. Samples were incubated by shaking and then centrifuged at 18000 × g for 45 min at 4 °C. Aqueous fraction of about 9 ml transferred to new tube, 1ml of sodium acetate solution (pH 5.2) and 10 ml of isopropanol was added. Samples were vigorously mixed. Nucleic acids were pelleted by centrifugation at 18000 × g for 10 min at room temperature. Pellet was washed with 70 % ethanol, drained and dissolved in 200 µl of nuclease free water.

Purified total RNA was incubated in 2.8 M LiCl overnight at –20 °C. Single stranded RNA pelleted by centrifugated at 18000× g for 45 min at 4 °C. DsRNA from supernatant was precipitated by using 0.1 volume of 3 M NaCl and 1 volume of isopropanol. DsRNA was washed with 70 % ethanol, drained and dissolved in 200 µl of nuclease free water.

1.2.2 Sequence independent cDNA synthesis of dsRNA

Single-stranded DNA primer ligation to dsRNA and reverse transcription

AK14+P primer was ligated to purified virus dsRNA genome (Table 1.6). Ligation was performed overnight at 37 °C. Samples of dsRNA with ligated $AK14^{-}$+P was purified by using GeneJet Gel Extract Kit, after adding equal sample volume of binding buffer another purification steps carried out according to the manufacturer's protocol.

DsRNA with ligated AK14+P primers was denaturated by addition of DMSO to a final concentration of 15 %. Samples incubated for 2 min at 98 °C, moved to ice bath for

5 min and then samples used for reverse transcription (Table 1.7). The reaction mix was incubated at 50 °C for 30 min.

Table 1.6. DNA primer ligation to dsRNA sample mix.

Regents	Amount
50 % PEG6000	12 µl
10x RNA ligase buffer	3 µl
100 % DMSO	3 µl
dsRNA	250 ng
100 µM AK-14+P primer	0,6 µl
RiboLock™ (25 u/µl)	0,4 µl
BSA (1 mg/1 ml)	3 µl
T4 RNR ligase (5 U/µl)	3 µl
Nuclease free water	to 30 µl

Table 1.7. Reverse transcription sample mix

Regents	Amount
dsRNAwith ligated AK14+P primer	12 µl
5x reverse transcription buffer	3 µl
10 mM dNTP mix	3 µl
RiboLock™ (25 u/µl)	0,4 µl
Maxima reverse transcriptase	3 µl
T4 RNR ligase (5 U/µl)	3 µl
Nuclease free water	to 20 µl

RNA hydrolysis and cDNA amplification

After cDNA synthesis, the residual RNA was hydrolyzed with 0.1 M NaOH. Before cDNA annealing, 0.1 M HCl was added. The cDNA was reannealed at 65 °C for 90 min followed by gradual cooling to 4 °C (Potgieter *et al.*, 2009). cDNA amplified using Phusion DNA polymerase with primer AK-15. The first PCR step was set for 60 s at 72 °C to ensure end repair was followed by the polymerase manufacturer's recommended cycling conditions: 98 °C for 10 s, 68 °C for 30 s and 72 °C for 40 s/kb.

1.2.3 Purification of viral particles

Preparation of yeast lysates with liquid nitrogen

Yeast strains were grown in SD medium overnight at 30 °C. Cell were harvested by centrifugation at 6000 × g for 10 min at 4 °C. Cell were resuspended in 0.1 volume water of initial volume and pelleted by centrifugation under the same conditions. Yeast were resuspended in "Yeast lysis buffer" (5 ml buffer for 1 g of yeast) and ground in a mortar chilled by liquid nitrogen for 15 min, then centrifugated at 12000 × g for 10 min at 4 °C.

Virus like particle ultracentrifugation

After cell lysis, selected supernatant was loaded onto a chilled 45 % sucrose cushion in centrifuge tubes and ultracentrifuged for 16 hours at 70000 × g at 4 °C (Kontron TST 28.38 rotor and Sanyo MSE MS60 ultracentrifuge are used). Supernatant was discarded and the pellet resuspended in 500 µl lysis buffer without PMSF for further analysis.

1.2.4 Yeast cell microscopy

Localization studies were performed with epi-fluorescence microscope Olympus AX70 equipped with 100× oil immersion objective. The images were recorded with

CDD camera „Orca 4742-95“. Further image processing was performed by the ImageJ program (Schneider *et al.*, 2012). For fluorescent detection of Hoechst 33342-stained DNA,U-WU filter cube was used, for mCherry tagged protein detection, U-MWG filter cube was used and for GFP tagged protein detection, U-MWIBA cube was used.

1.2.5 Bioinformatic analysis

Search for homologous sequences in genomes

The search of homologous sequences using the protein or gene sequences was performed by the BLAST+ program (Camacho *et al.*, 2009). Search was performed in genomes and *de novo* assembled transcriptomes mentioned in section 1.2.6. Also, the NCBI nucleotide collection (nr/nt) and non-redundant (nr) sequence databases were used.

Multiple sequence alignment

Multiple sequence alignments were constructed using MUSCLE (Edgar, 2004) or MAFFT (Katoh and Standley, 2013) with default parameters.

Construction of a phylogenetic tree

Maximum likelihood phylogenetic tree was constructed using IQ-Tree v1.6.3 (Nguyen *et al.*, 2015) with automatic selection of best-fit amino acid substitution and site heterogeneity models. LG+R3 proved to be the best-fitted model. Edge support was estimated with bootstrap test (1000 replicates).

Characterization of unknown proteins

Phobius web server (Käll *et al.*, 2007) was used for transmembrane domain and signal sequence prediction in the protein sequence. Sites of N-glycosylation in the protein sequences were identified using NetNGlyc web server (http://www.cbs.dtu.dk/services/NetNGlyc/). Conservative domains of the protein was determined using Pfam database (Finn *et al.*, 2016).

De novo transcriptome assembly

All *de novo* transcriptome assembly steps were performed on open access, high-performance https://usegalaxy.org/ and https://usegalaxy.eu/computing platforms (Afgan *et al.*, 2016). Prior to transcriptome assembly, RNA-seq reads were subjected to adapter and quality trimming using Trimmomatic (Bolger *et al.*, 2014). Transcriptome *de novo* reconstructed with Trinity assembler (Grabherr *et al.*, 2011).

Other used software

The UCSF Chimera Visualization software (Pettersen *et al.*, 2004) was used for protein structures visualization. The statistical analysis and graph plotting performed

using R v3.5.1 software (https://www.r-project.org/). The Shannon entropy is calculating using the Bio3D v2.2 package for R software.

1.2.6 Data and databases for bioinformatic analysis

Saccharomyces sensu stricto clade genomes obtained from Liti *et al.*, 2009; Scannell *et al.*, 2011; Yue *et al.*, 2017; Peter *et al.*, 2018 research were used in this work.

Raw sequence reads of RNA-seq experiments for *de novo* transcriptome reconstruction were obtained from SRA database.

2 RESULTS AND DISCUSION

2.1 Analysis of dsRNA virus genomes

2.1.1 Identification of new variants of dsRNA L-A viruses and M satellites

To identify new dsRNA viruses and their satellite genome sequences, we developed approach for viral dsRNA identification and amplification (Fig. 2.1 A). The approach consists of four stages: dsRNA extraction, sequence-independent amplification, cloning and sequencing of viral genome DNA and bioinformatic analysis of sequenced viral genomes. Using this method, the strains with killer phenotype and others from the collection of Genetics Laboratory from the Nature Research Centre (NRC) were analyzed.

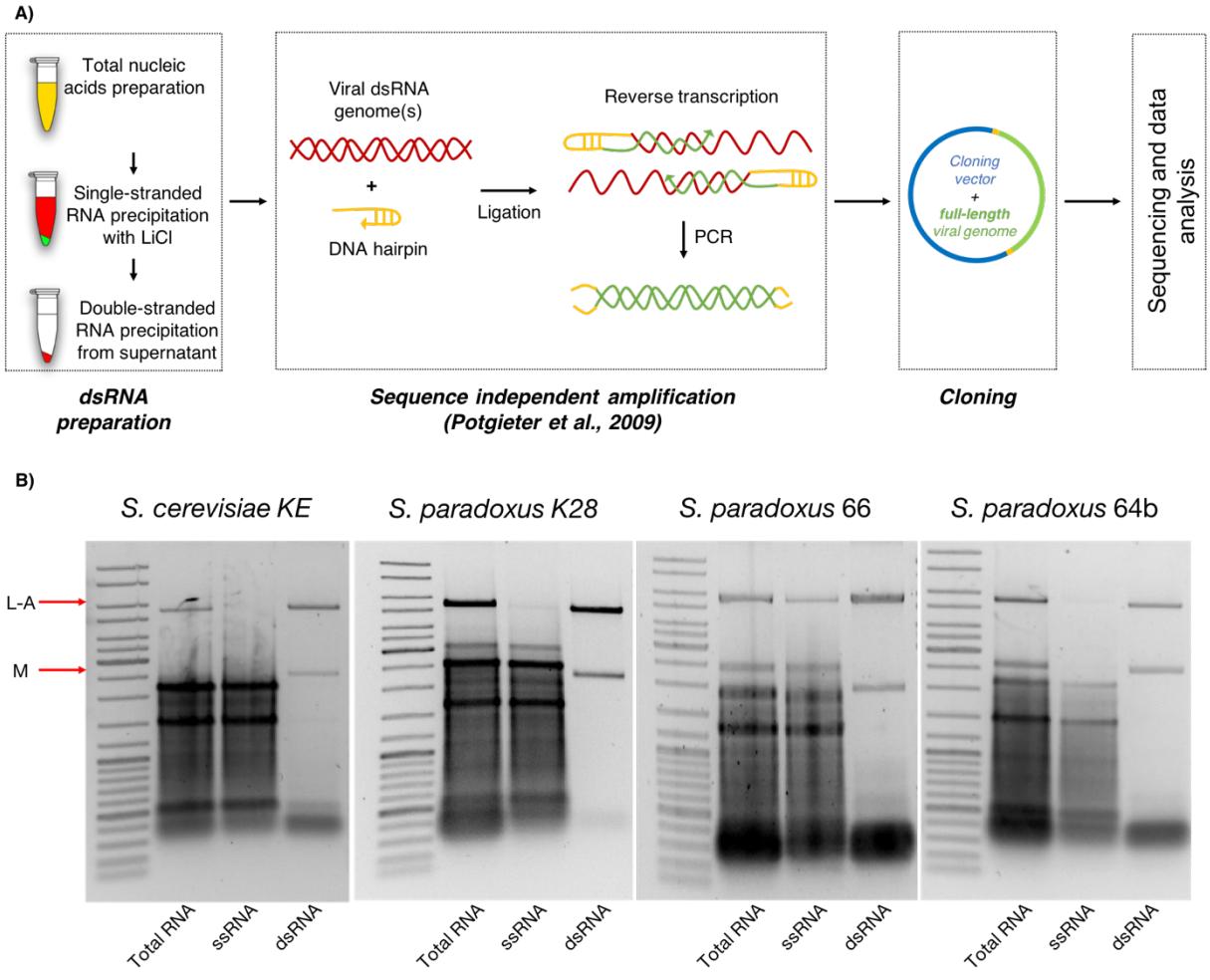

Fig. 2.1. Identification of double-stranded RNA viruses. A) Algorithm of dsRNA virus genome sequence identification; B) Total RNA from different strains fractionated with lithium chloride. Three fraction are separated: total RNA, single-stranded RNA (ssRNA) and double-stranded RNA (dsRNA).

DsRNA L-A viruses and M satellites were detected in all examined strains with established killer activity (Fig. 2.1 B). After we sequenced L (after Large) segment genomes, we identified different variants of L-A virus in analyzed strains. Some of them show high genome sequence similarity to known L-A virus variants, while others are

more variable (Fig. 2.2). By analyzing the available laboratory strains of *S. cerevisiae* (BY4741, α′1, 21PMR) we identified in each of them variants of the L-A virus, closely related to the ScV-LA-1 virus variant. We decided to appreciate prevalence of L-A virus in the most commonly used *S. cerevisiae* strains.

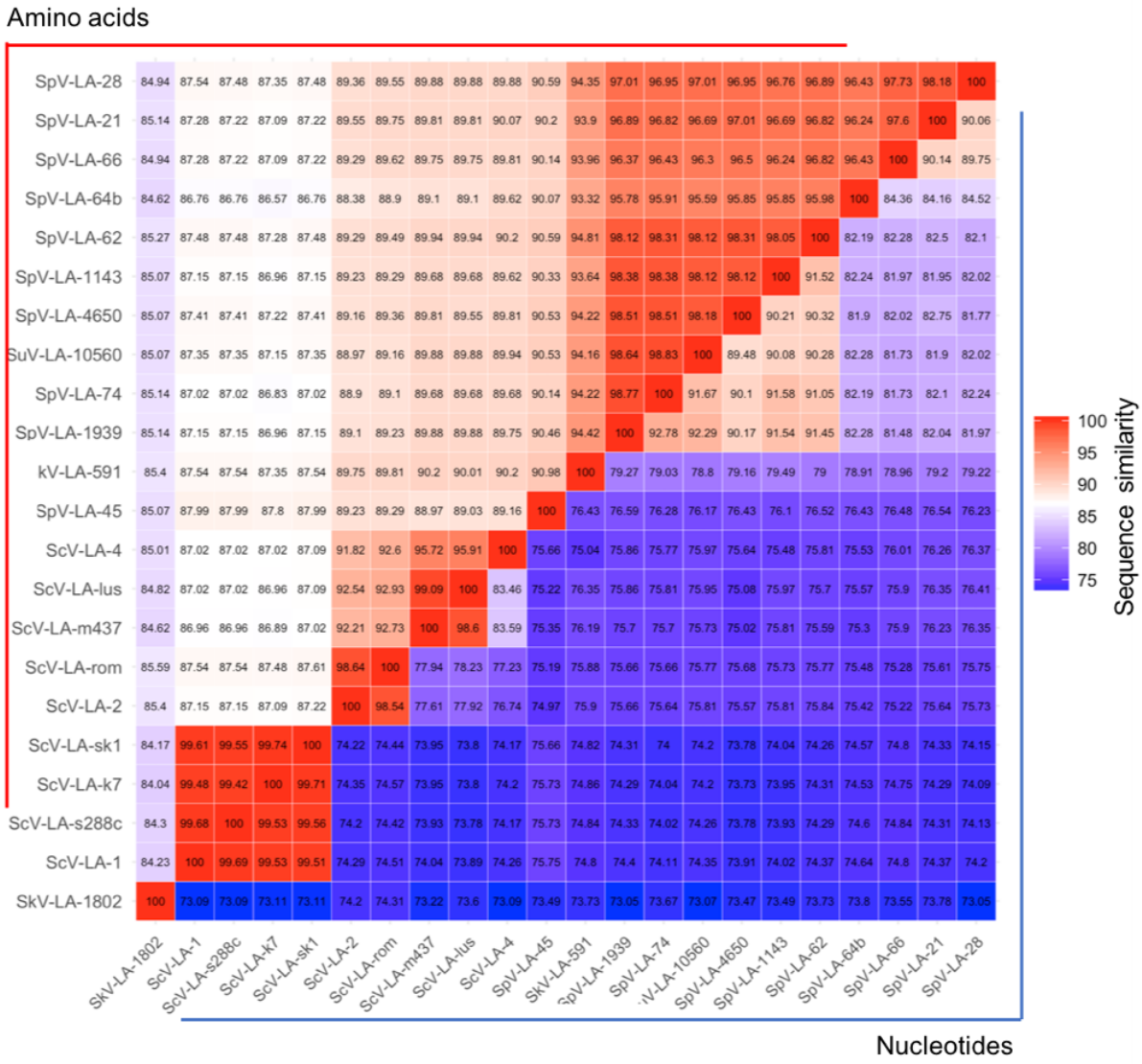

Fig. 2.2. The similarities of known dsRNA L-A viruses.

First we searched the NCBI SRA database for the raw data of RNA-seq experiments of the most commonly used *S. cerevisiae* strains (Louis, 2016). Then we performed *de novo* transcriptome reconstruction and searched for L-A virus sequences. Similar sequences to ScV-LA-1 variant were detected in eleven strains: S288C, BY4741, BY4742, BY4743, FY4, W303, YPH499, YPH500, Σ1278b, CEN.PK, SK1.We have found that the most commonly used laboratory strains of the *S. cerevisiae* are infected with the ScV-LA-1 variant of L-A virus exclusively.

2.1.2 Phylogenetic analysis of L-A virus sequences

We used all available sequences of L-A virus published in literature (Konovalovas *et al.*, 2016; Rowley *et al.*, 2016; Rodríguez-Cousiño and Esteban, 2017; Rodríguez-

Cousiño *et al.*, 2017) and discovered in this work (section 2.1.1) for the phylogenetic tree construction. The phylogenetic tree is constructed by using Gag and Gag-Pol proteins sequences of L-A virus. The phylogenetic tree shows that virus sequences from *S. cerevisiae* and *S. paradoxus* form two separate clades (Fig. 2.3). Moreover, we identified formation of two subclades (Scer-1 and Scer-2) in the *S. cerevisiae* L-A virus phylogenetic clade, based on maintenance of M satellite (Fig. 2.3).

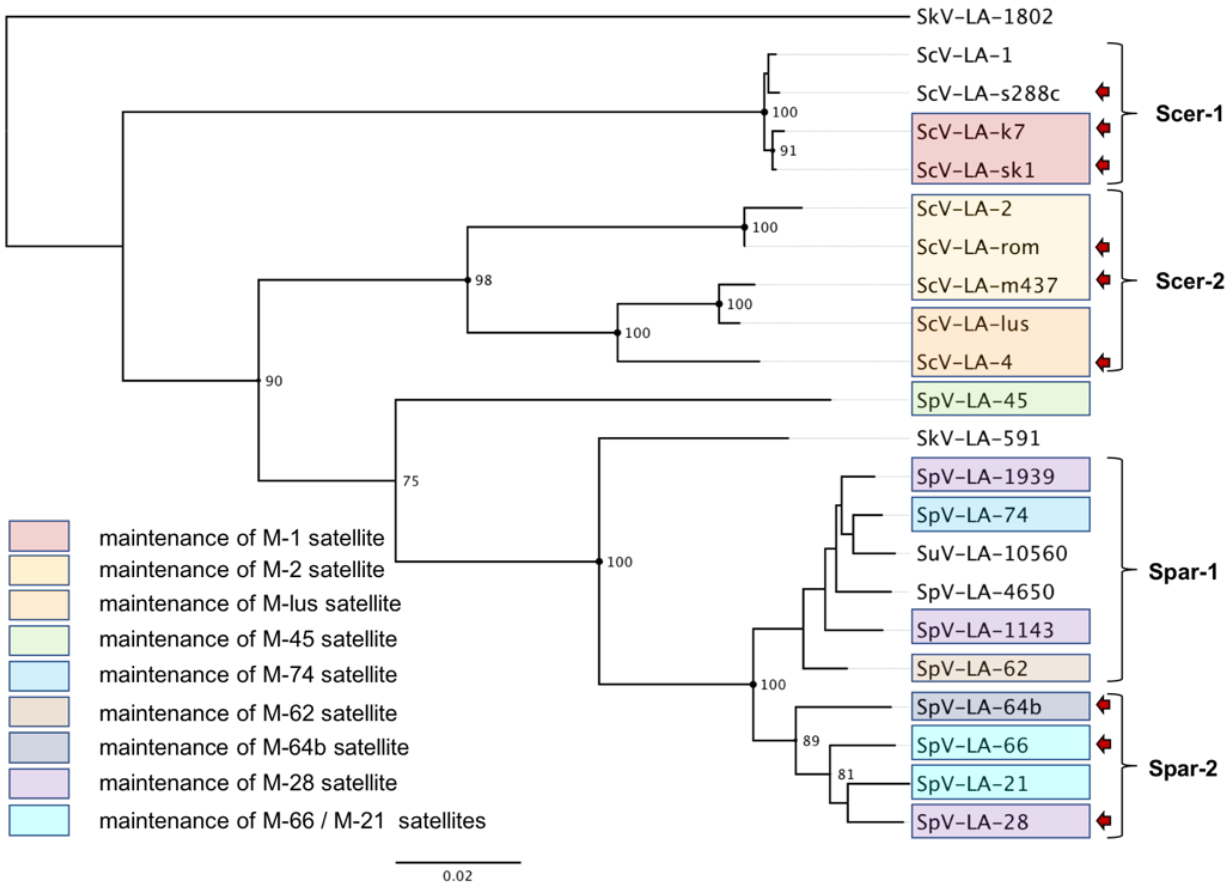

Fig. 2.3. Phylogenetic tree of dsRNA L-A virus variants found in *Saccharomyces sensu stricto* clade.

Historically, yeast strain K28 has been classified as *S. cerevisiae* (Pfeiffer and Radler, 1984). Following the identification of L-A virus sequences from K28 strain and two *S. paradoxus* strains (64b and 66), it was found that they form a phylogenetic clade separate from L-A viruses from *S. cerevisiae*. Obtained results allowed us to hypothesize that K28 strain belongs to *S. paradoxus* and previously was mistakenly attributed to *S. cerevisiae*. This hypothesis was confirmed by RFLP analysis and sequencing of internal transcribed spacer regions of ribosomal DNA (Dr. Elena Servienė, unpublished).

2.1.3 Relationship between coding sequence variability and structure of the L-A virus coding proteins

The L-A viruses found in *Saccharomyces sensu stricto* clade are very conservative in their sequences (Fig. 2.4 B). However, we have found that certain genomes of L-A virus are particularly variable, where drastic changes in amino acids sequences are observed - some amino acids is replaced by others with strikingly different physicochemical

properties (Fig. 2.4 B). We discovered that the most variable regions are located on outer surface of capsid, meanwhile inner capsid surface is highly conservative (Fig. 2.4 A).

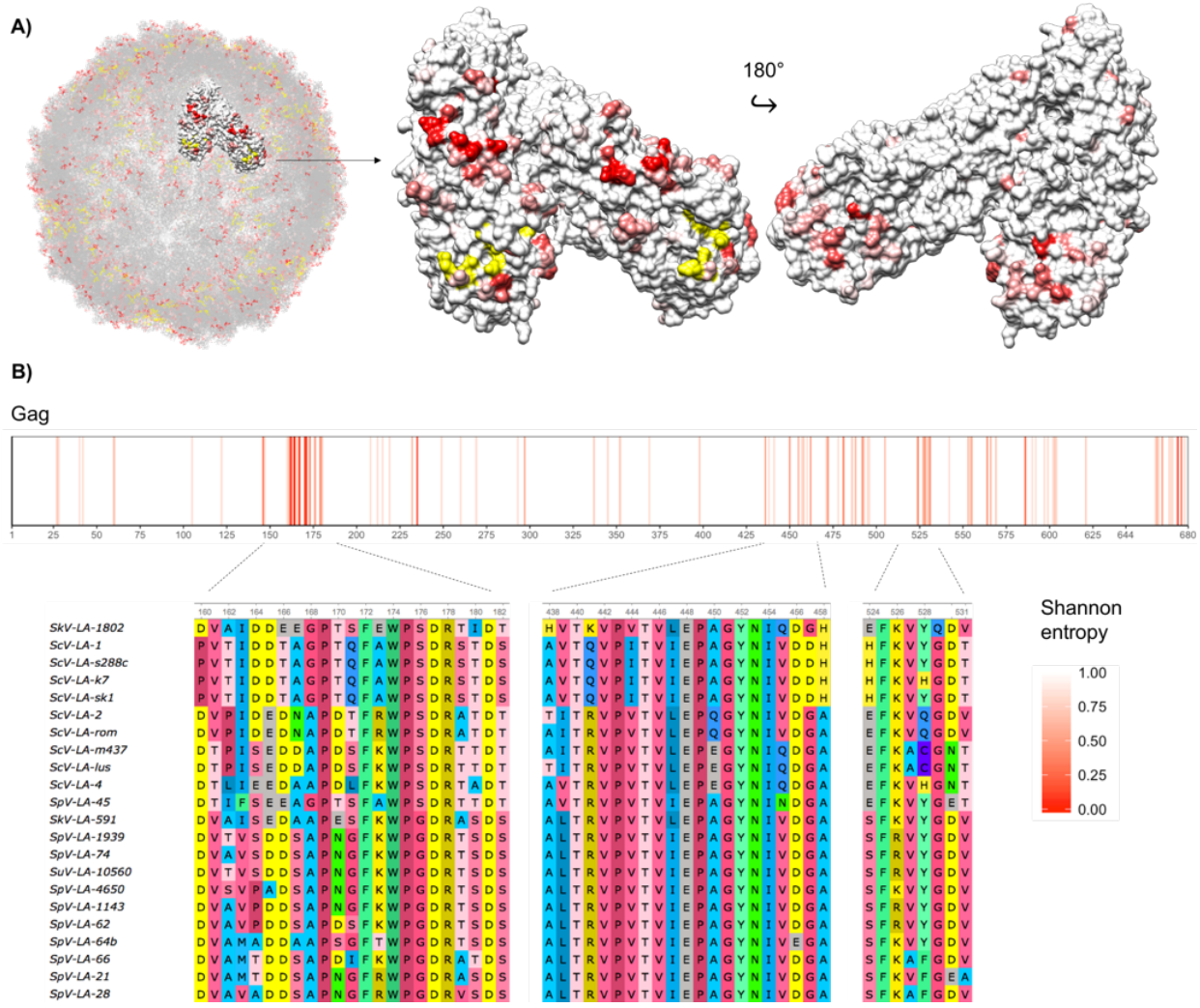

Fig. 2.4. Relationship between Gag protein of L-A virus sequence variability and structure. A) Gag protein structure (PDB archive: 1M1C): viral particle; the outer and inner sides of Gag asymmetric dimer. The white-red gradient shows the estimate Shannon's entropy, which is calculated using a multiple alignment of the Gag proteins sequences. Shannon's entropy estimate of one denotes the conservatism of the position (white), while the value approaching zero show the increase in variability (red color). The amino acids involved in the cap-snatching are marked in yellow; B) Multiple sequence alignment and graph of conservative-variable positions of the Gag protein sequences.

Pol domain of Gag-Pol protein stands out by the sequence variability in comparison with Gag domain. Analysis of multiple sequence alignments revealed that there are at least one substitution in 14% of the amino acid positions in the Gag protein, while in the Pol domain this number reaches 30%; non-conservative amino acid changes are at least twice as frequent. The most variable region of Pol domain is located between RdRp catalytic center and C-terminal ssRNA binding domain (1321-1325 aa) (Fig. 2.5). The function of this region is not yet known.

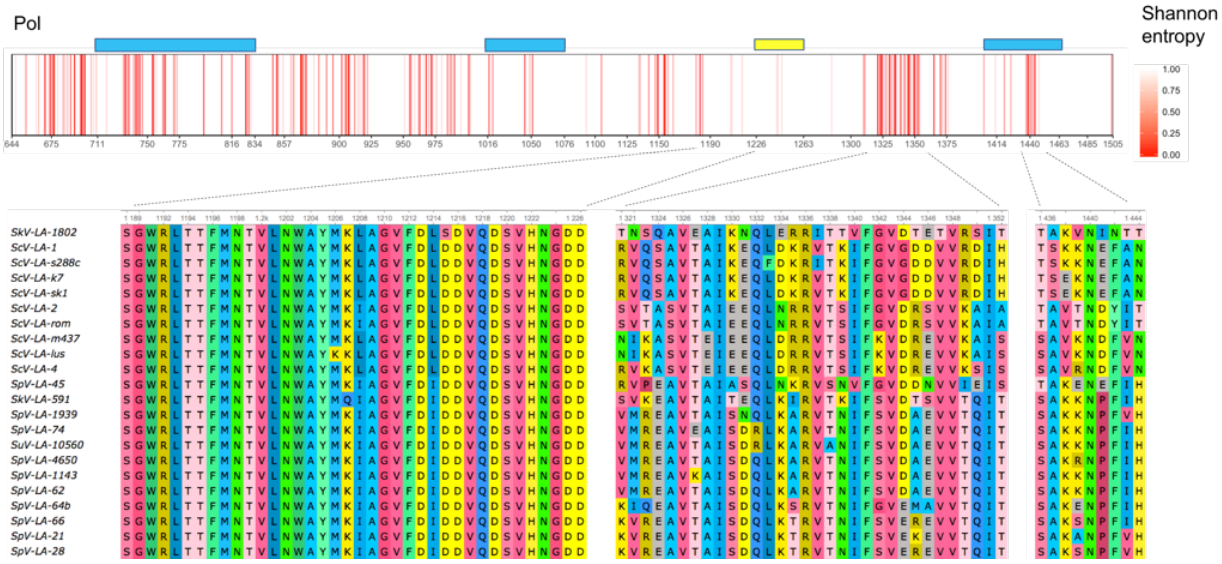

Fig. 2.5. Multiple sequence alignment and graph of conservative-variable positions of L-A virus-encoded Gag-Pol protein sequences. Blue rectangles above the graph denote single-stranded RNA binding domains, yellow rectangle – RNR dependent RNA polymerase catalytic domain.

2.1.4 Analysis of new L-A virus M satellite sequences

Two new M satellites SpV-M66 and SpV-M64 were identified in *S. paradoxus* killer strains. Killer strains were isolated from natural environment by Dr. Elena Servienė from NRC Genetic Laboratory. Genome architecture coincides with other known *Saccharomyces sensu stricto* clade L-A virus M satellites, which consist of short 5′ untranslated region (UTR), open reading frame, polyA sequence and long 3′ UTR. SpV-M66 satellite genome sequence observes 86 % identity with genome sequence of SpV-M21 satellite (Rodríguez-Cousiño *et al.*, 2017) at the nucleotide level and 92 % identity at the coding-protein level. We expressed satellite-coded killer toxin in BY4741 strain and confirmed, that protein encoded by SpV-M66 satellite possess killer activity (data not shown). We identified three transmembrane helixes and conservative Pfam family domain DUF5341 (Fig. 2.6 A). The function of DUF5341 domain in the cell is not yet known. Proteins with this domain are mostly found in *Ascomycota*.

We detected many homologues of the ScV-M64 satellite encoded protein in *Schizosaccharomyces* genus and *Saccharomyceyaceae* family species genomes. We discovered similar to ScV-M64 genome transcript sequence (N17-M64) in *S. paradoxus* N-17 yeast strain transcriptome, which was reconstructed *de novo*. SpV-M64 and N17-M64 shows 83 % sequence similarity at the nucleotide level and satellite coded proteins show 89 % similarity. Also, we did not find N17-M64 sequence in the N-17 strain genome, so we believe that *S. paradoxus* N-17 strain have M satellite similar to M64 satellite.

Homologous M satellite variants from *S. paradoxus* show higher variability then homologous M satellites variants, found in *S. cerevisiae*. For example, SpV-M66 and SpV-M21 satellites show 86 % similarity at the nucleotide level and 92 % at the encoded

protein level, respectively SpV-M64 and M64-N17 observe 83 % and 89 % similarity. The most divergent *S. cerevisiae* M-1 satellite-like variants of encoded protein show 98.73 % similarity either most divergent *S. cerevisiae* M-2 satellite-like variants 97.5 % similarity at the encoded protein level. The variability of M-satellites in *S. paradoxus* yeasts can indeed be a factor in adapting the species to changing environmental conditions.

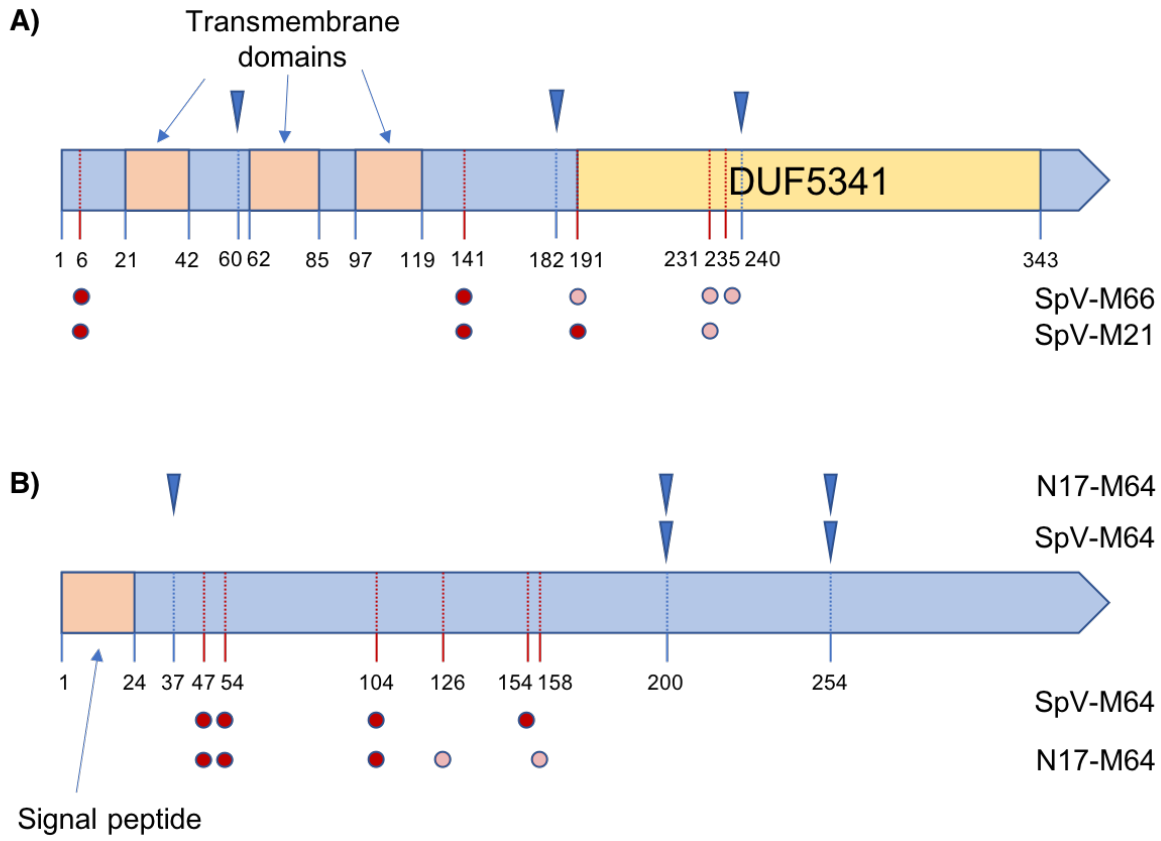

Fig. 2.6. Features of proteins coded by SpV-M66 and SpV-M64 satellites. A) Diagram of SpV-M66 satellite encoded protein, predicted transmembrane domains and DUF5341 conservative Pfam family domain is marked; B) Diagram of SpV-M64 satellite encoded protein. Triangles indicated predicted KEX2 protease sites and circles indicated predicted N-glycosylation sites, N-terminal hydrophobic probable signal sequence was marked.

2.1.5 M satellites and endogenous virus elements in the yeast

We have found homologues to all known *Saccharomyces sensu stricto* yeast M-encoded toxins in different yeast species genomes. Homologues of *S. cerevisiae* M1 and M2 satellites and *S. paradoxus* M45 and M74 satellite-encoded proteins are found in various yeast species, but in *Saccharomyces sensu stricto* clade these homologues are not found. Meanwhile, homologues of *S. cerevisiae* Mlus and *S. paradoxus* M64, M66, M62 satellite-encoded proteins are detected only in *Saccharomyces sensu stricto* clade genomes. A significant similarity of the sequences makes it possible to assume the common origin. However, the similarity of the coding sequences at the nucleotide level in the sequences listed above is not observed.

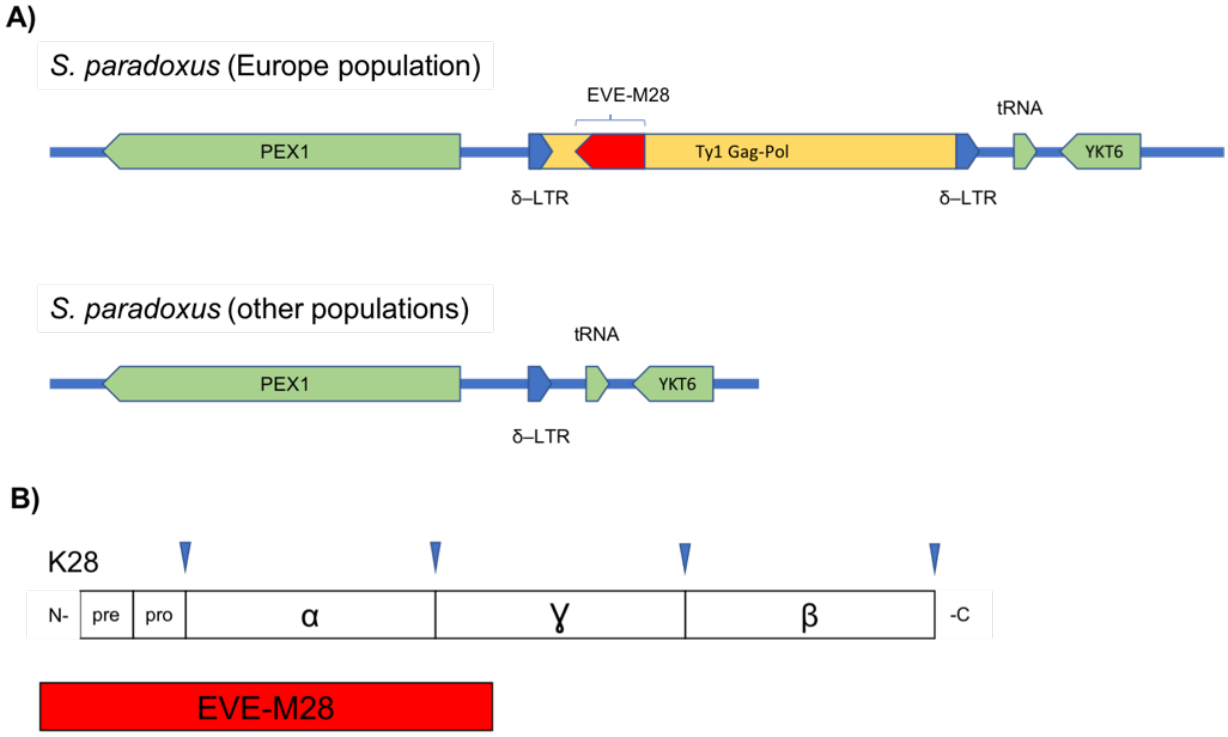

Fig. 2.7. M28 satellite analogues – endogenous virus elements. A) A comparison of XI chromosome region between genes encoding PEX1 and YKT6 proteins of *S. paradoxus* Europe population and other *S. paradoxus* populations; B) A comparison of M28 satellite encoded K28 toxin with EVE-M28 encoded protein. The blue triangles show the KEX2 protease cleavage sites.

We detected homologues of M28 satellite-encoded killer toxin EVE-M28 only in the European population of *S. paradoxus* yeast. Observed sequence similarity is about 88 % at the nucleotide and protein levels. EVE-M28 is located between PEX1 and YKT6 protein-encoding genes (Fig. 2.7 A). Among those genes, Ty1 transposon integration site is also found. In other *S. paradoxus* genomes, this region is conservative and Ty1 transposon integration is not observable (Fig. 2.7 A). Endogenous virus elements (EVEs) occur frequently during evolution and are thought to be mediated primarily by reverse transcriptases, encoded by host transposons (Feschotte, 2008; Bruenn *et al.*, 2015; Geuking *et al.*, 2009). In the case of EVE-M28, we are observing precisely this type of integration. By comparing the EVE-M28 encoded protein and the K28 killer toxin sequences, we found that the EVE-M28 encoded protein coincides with the amino acids in positions 34-152 of the K28 toxin (α subunit of K28 killer toxin) (Fig. 2.7 B). The α subunit of K28 toxin is responsible for the killer activity, while the β subunit is necessary for killer toxin transport into the target cell (Schmitt and Breinig, 2006). It has been demonstrated that yeast becomes significantly more resistant to K28 toxin, when they produce the recombinant α subunit protein (Breinig *et al.*, 2006). Analysis of *S. paradoxus* N-17 strain (Europe population (Liti *et al.*, 2009)) *de novo* reconstructed transcriptome confirmed transcription of EVE-M28. Future studies will clarify the precise function of the EVE-M28 in the yeast cell.

2.2 ScV-LA-1 virus interaction with host cell proteins

2.2.1 Expression of ScV-LA-1 virus-encoded proteins

To detect the effect of recombinant ScV-LA-1 virus variant-encoded proteins on the native L-A virus in the BY4741 strain, we constructed four plasmids (pYAK3-LAcds, pYAK3-GagWT, pYAK3-GagΔ, pYAK3-Gag+CFH) which codes different genetically modified L-A virus proteins (section 1.1.7). Transformation of *S. cerevisiae* BY4741 strain with pYAK3-LAcds plasmid leads to a high biosynthesis level of Gag protein and an increased level of the Gag-Pol protein synthesis, compared to the BY4741 transformed with the empty expression vector pYAK3 (Fig. 2.8 A). Also, the analysis of total RNA show that the increase in the synthesis level of recombinant Gag and Gag-Pol (formed by ribosomal frameshifting) proteins promotes the replication of a native L-A virus dsRNA genome (Fig. 2.8 B). Same replication promotion is observed during the expression of the recombinant Gag protein only (Fig. 2.8). The data obtained from these experiments suggest that increase in the amount of recombinant Gag protein biosynthesis is sufficient for replication stimulation of viral genome.

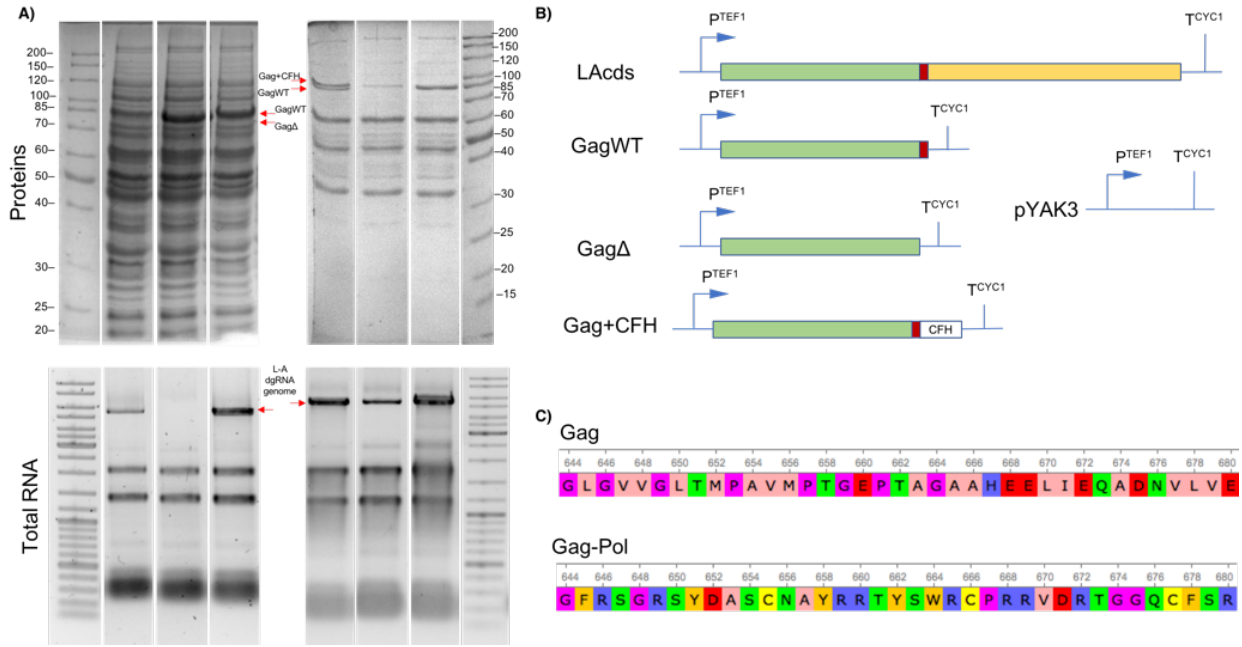

Fig. 2.8. The expression of recombinant ScV-LA-1 virus encoding proteins in BY4741 yeast strain and the effect of encoded protein biosynthesis on native L-A virus replication. A) SDS-PAGE gel (top) and agarose gel of total RNA (bottom) from BY4741 strain transformed with different genetically modified L-A virus proteins; B) ScV-LA-1 virus protein expression cassette in the pYAK3 expression vector. P^{TEF1} – TEF1 promoter. T^{CYC1} – CYC1 terminator, Gag domain – green, Pol – yellow, ribosomal frameshift region – red; C) Peptide sequences encoded by overlapped conservative ribosomal frameshift region of Gag and Gag-Pol proteins.

Two L-A virus open reading frames encoding Gag and Gag-Pol proteins have overlapped conservative 108 bp length ribosomal frameshift region. This region encodes two physiochemically by different polypeptides (Fig. 2.8 C). During the translation of the Gag protein, the end of the Gag protein C-terminal end (644-680 aa) consists of hydrophobic and negatively charged amino acids, while during the ribosomal

frameshifting, the C-terminal end of the Gag domain of the Gag-Pol fusion protein (644-680 aa) consists of polar and positively charged amino acids (Fig. 2.8 C). Such distribution of amino acids between the Gag and the Gag-Pol proteins at 644 to 680 aa positions and high conservatism of the sequence between different variants of the L-A virus suggests that peptide encoded by this ribosomal frameshift motif may have an important role in virus replication and/or replication cycles. To confirm this hypothesis, we constructed a pYAK3-GagΔ plasmid that encodes the Gag protein without overlapping region (GagΔ) (Fig. 2.8 B). Analysis of the total RNA from BY4741 transformed with pYAK3-GagΔ plasmid show that expression of truncated Gag aborts native virus replication (Fig. 2.8 A). We thus assume that the peptide encoded by the L-A virus region of the ribosomal frameshift is necessary for the replication of the native L-A dsRNA virus genome.

2.2.2 Purification of ScV-LA-1 nucleocapsid and virus like particles

To evaluate the ability of proteins coded by constructed plasmids (section 2.2.1) to produce virus-like particles, high molecular weight protein complexes were isolated by ultracentrifugation (section 1.2.3). Following ultracentrifugation of protein sample of a strain expressing a full-length Gag protein, a significant enrichment of Gag protein is observed (Fig. 2.9 A); also, virus dsRNA genome was detected in this sample (Fig. 2.9 B). Meanwhile, by analyzing of the protein sample from a strain expressing truncated Gag protein we found that GagΔ formed dsRNA-free virus like particles (Fig. 2.9 A and B), since expression of GagΔ eliminates native L-A virus dsRNA genome from the cell (section 2.2.1).

Based on the obtained results, we believe that the C-terminal end of the Gag protein is responsible for interacting with Gag-Pol fusion protein. In the literature it is mentioned, that after removing the C-terminal end from the Gag such protein produces virus like particles of the same structure as the full length Gag protein (Fujimura *et al.*, 1992). Also, it has been described that after removing of 33 aa from Gag protein C-terminal end, such protein is unable to form the viral like particles with the Gag-Pol fusion protein (Ribas and Wickner, 1998). Therefore, in excess of the GagΔ protein in the cell, it may contain asymmetric dimers with Gag protein encoded by native L-A virus and forming Gag-Pol-free virus-like particles. The consequence would be the elimination of a native L-A virus genome from the cell. Meanwhile, in the presence of a full-length Gag protein overexpression, the promotion of native L-A virus replication is observed. The recombinant, full-length Gag protein contains a sequence possibly responsible for interactions at the C-terminal end and can form asymmetric dimers with Gag Pol protein, what results in the formation of virus particles with the incorporated Gag-Pol fusion protein necessary for virus replication and transcription.

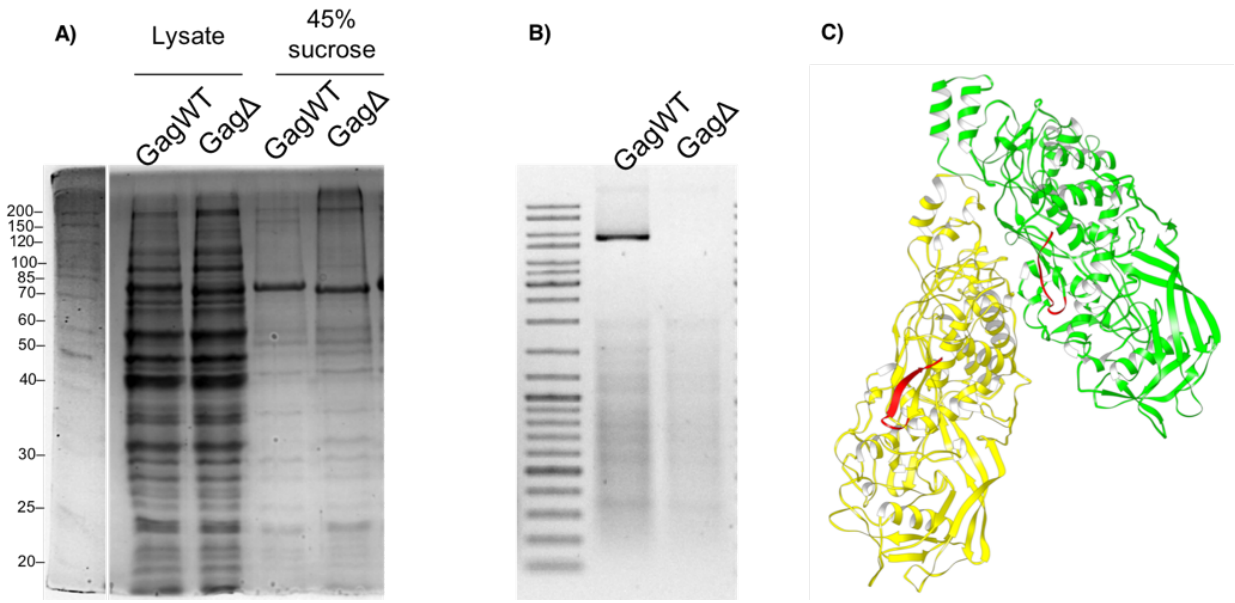

Fig. 2.9. Purification of L-A virus nucleocapsid and other high molecular weight derivatives, structure of asymmetric Gag protein dimer. A) SDS-PAGE gel: lysate – protein samples prior to ultracentrifugation, 45% sucrose – protein samples after ultracentrifugation. LAcds – BY4741 strain transformed by pYAK3-LAcds plasmid, GagΔ – BY4741 strain transformed by pYAK3-GagΔ plasmid; B) Nucleic acids extracted from protein samples after ultracentrifugation; C) The asymmetric dimer structure (PDB archive: 1M1C) of Gag protein. Dimer forming monomers are green and yellow, ribosomal frameshift region – red.

2.2.3 Interaction between virus-encoded proteins and host cell proteins

Based on the *S. cerevisiae* BY4741 strain, we constructed four strains to identify host cell proteins interacting with the proteins encoded by the L-A virus. These strains are BY4741 GagWT (BY4741 transformed with pYAK3-GagWT plasmid), BYΔLA GagWT (BY4741 strain without native LA virus (BYΔLA) transformed with pYAK3-GagWT plasmid), BY4741 GagΔ (BY4741 transformed with pYAK3-GagΔ plasmid) and BYΔLA pYAK3 (BYΔLA strain transformed with empty expression vector pYAK3). High molecular weight cell complexes from these strains were purified on sucrose cushion (Fig. 2.10 A and B). Proteins forming high molecular weight complexes were determined by LC-MS/MS proteomic analysis (proteomics analysis was performed by Dr. A. Kaupinis and Dr. M. Valius, VU LSC BChI Proteomics center).

In four samples, 864 unique proteins were identified during proteomic analysis. Proteins detected in the sample from BYΔLA pYAK3 strain were removed from further analysis. This strain does not contain any genetic material related to L-A virus and proteins detected in this sample are considered in a comparative proteomic analysis as non-viral high molecular weight protein complexes. The remaining 329 identified proteins are divided into three interacting groups by the type of interaction: the proteins interacting with Gag protein, the proteins interacting with Gag protein C-terminal end and the proteins interacting with inner capsid structures.

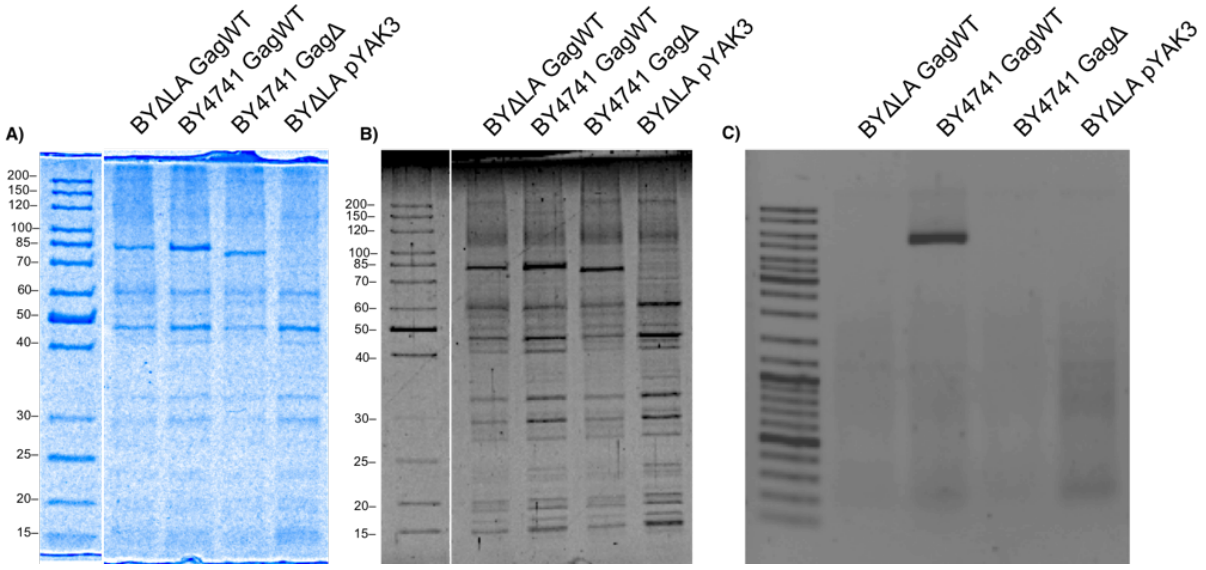

Fig. 2.10. The proteins and nucleic acids compositions of the samples used for the proteomic analysis. A) SDS-PAGE gel stained with PageBlue^{TM} dye; B) SDS-PAGE gel stained with SYPRO^{TM} Ruby dye; C) Purified nucleic acids from the protein samples after ultracentrifugation.

Proteins that interact with Gag are those that coincide in proteomes of protein samples prepared from BY4741 GagWT, BYΔLA GagWT and BY4741 GagΔ strains. This group of proteins includes proteins that are potentially interacting with capsid formed by Gag protein. We identified 70 such proteins (Fig. 2.11 A). Among these proteins, 52 proteins are linked to cell membranes, among which 35 proteins are linked to the endoplasmic reticulum. Moreover, four Ty1 transposon Gag-Pol proteins (YER138C, YDR365W-B, YBL005B-W and YNL284C-A) were also identified.

Proteins that interact with C-terminal end of Gag protein are those that coincide in proteomes of protein samples from BY4741 GagWT and BYΔLA GagWT strains but do not identified in proteome of BYΔLA GagΔ strain. We identified 24 such proteins. Three proteins of Ino80 complex are found among these proteins (ARP8, IES1 and INO80). The remaining proteins do not have common features, furthermore the molecular function of the part of the identified proteins in the cell is unknown.

The proteins interacting with the inner virus capsid structures are those that are unique for BY4741 GagWT proteome. BY4741 GagWT is the only strain from comparative proteomic analysis in which the dsRNA virus replication occurs and we presume that unique proteins identified in analysis from this strain potentially interact with the Pol domain of the LA virus fusion protein Gag-Pol and/or the dsRNA virus genome. In total, 46 proteins are assigned to this group. Among these proteins four proteins of Ino80 complex (IES3, IES5, NHP10 and TAF14) were found, as well as the ubiquinone biosynthesis-related proteins (COQ3, COQ5, COQ6, COQ11 and CAT5) were found.

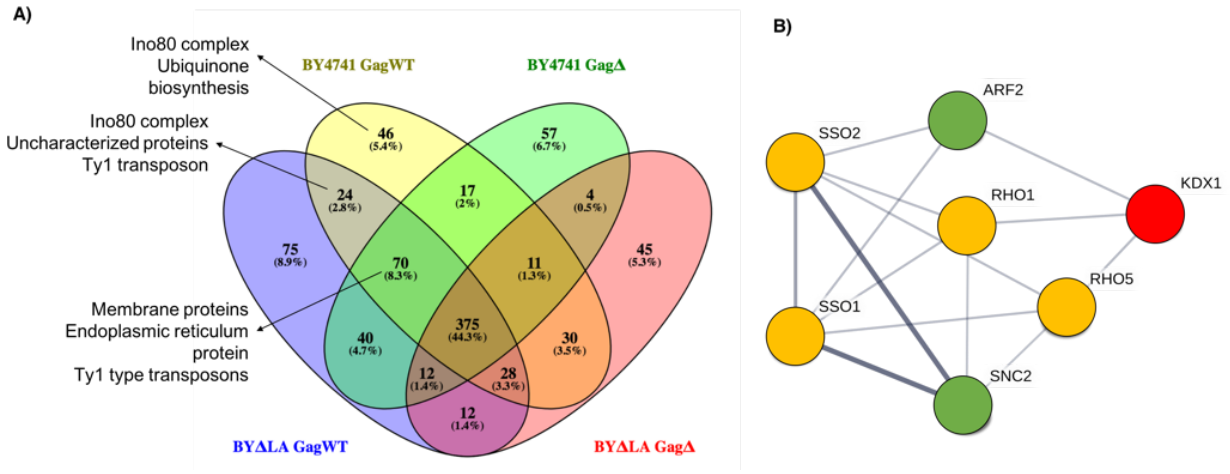

Fig. 2.11. Comparative proteomic analysis and relationship with transcriptomic analysis data. High molecular weight protein complexes and their relationships which identified by analysis of the four strain proteomes shown in the Venn diagram. Arrows show three protein groups interacting with proteins encoded by L A virus or with virus genetic material; B) Proteins interaction network of proteins interacting with LA virus and proteins identified during transcriptional analysis. (The interaction network is constructed using https://string-db.org/ database version 10.5). The proteins identified during the proteomic analysis – green, the proteins identified during transcriptomic analysis – red (KDX1) and the proteins identified during interaction network analysis – yellow.

In previous studies, we found that the presence or absence of L-A virus in the cell changes the expression of many cellular genes (Lukša *et al.*, 2017). After selecting genes which expression level differ by more than four times among the yeast that contains or do not have L-A virus, we performed analysis of proteins interaction networks to determine the relationships between proteins interacting with LA virus and proteins identified during transcriptional analysis. We determined the interactions network between proteins which interacts with L-A and KDX1 protein (Fig. 2.11 B). The KDX1 encoding gene expression is increased by approximately 12 times in the case of the elimination of L-A virus (Lukša *et al.*, 2017). The KDX1 protein is a protein kinase that activates stress-response proteins in cellular stress associated with cell wall damage (Chang *et al.*, 2013; García *et al.*, 2004). We have found five intermediate proteins (SSO1 and SSO2 – t-SNARE proteins, SCS2 v-SNARE protein and two GTP binding proteins of RAS-family RHO1 and RHO5) which can be an interaction bridge between KDX1 and host cell proteins, which interact with L-A virus directly.

2.2.4 Impact of host proteins on L-A virus replication

We selected 27 candidates of proteins identified in proteomic analysis that are potentially interacting with L-A virus encoded proteins (section 2.2.2) for further analysis. To evaluate the effect of the lack the expression of these proteins on the replication of L-A virus, analysis of total RNA and dsRNA from BY4741 strain-based YKO (yeast single-gene-knockout library) strains has been performed.

We discovered that native L-A virus replication is affected by 15 analyzed host genes. Single deletion of four genes reduced native virus replication, while single deletion of 11 genes totally inhibit native virus replication (Fig. 2.12). The reduction of

native L-A virus genome replication is observed in strains with deleted SSO1, RHO5, YNK1 and CAT5 protein-encoded genes. SSO1 and RHO5 proteins were identified by analyzing protein interaction networks of the proteomic and transcriptomic analysis. Both of these proteins are likely regulated by KDX1 protein kinase. Observed reduction of the replication of the dsRNA virus may be associated with cellular signaling abnormalities, which suggest that the host cell may have yet undiscovered signal pathways to coordinate the replication of L-A virus genome.

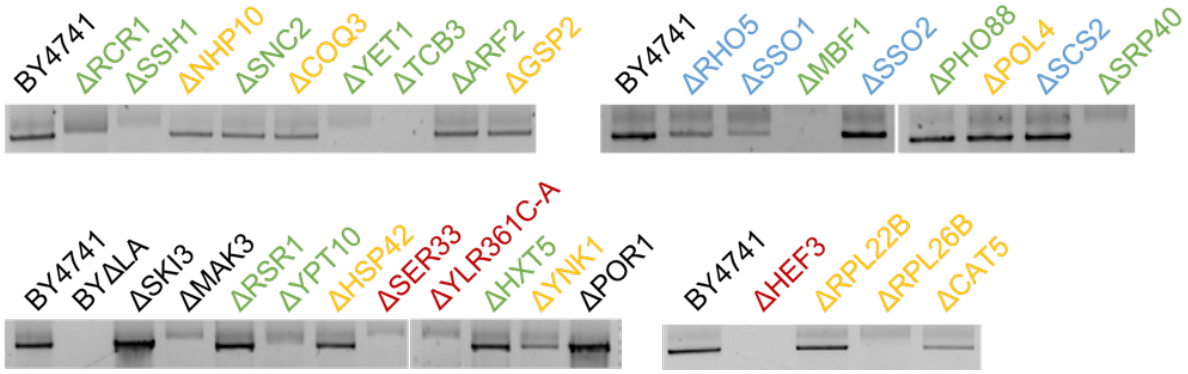

Fig. 2.12. The influence of host gene deletion on genome replication of native L-A virus dsRNA. Name the strain is according to the name by the deleted gene-encoded protein name. The group of proteins that interact with Gag – green, the group of proteins that interact with C-terminal end of Gag protein – red, the group of proteins interacting with the inner virus capsid – yellow , proteins identified by analysis of interaction networks – blue.

RPL26B was detected in the group of proteins interacting with the inner virus capsid structures. The deletion of RPL26B protein-encoding gene totally inhibit native L-A virus replication (Fig. 2.12). RPL26B is non-essential ribosomal 60S subunit protein. The deletion of this protein-encoding gene does not affect the growth of yeast cell or ribosome biogenesis (Babiano *et al.*, 2012). However, we found that the sequence of RPL26B protein among all *Saccharomyces sensu stricto* clade yeast are particularly conservative – in fact, identical. An ultimate conservatism in the sequence is usually associated with an essential conservative protein function in the cell. We believe that this protein is necessary for virus genome replication by interaction with conservative viral genomes regions (for example, ribosomal frameshift region or encapsidation signals) and/or with Pol domain of Gag-Pol protein.

Other 10 proteins, identified during comparative proteomic analysis, whose gene deletion inhibits native virus replication, are from groups of proteins that interact with Gag and proteins that interact with C-terminal end of Gag protein. Five of those proteins (RCR1, TCB1, SSH1, YET1 and YPT1) are implicated in cell transport and endoplasmic reticulum. Other three proteins (HEF3, MBF1 and SRP40) are involved in interaction with nucleic acids. YKL361C-A protein-encoding gene deletion inhibit L-A virus replication. The function of protein YKL361C-A is unknown. We detected homologues of this protein only in yeast genomes from the *Saccharomycetaceae* family.

2.2.5 Localization of ScV-LA-1 virus-encoded proteins in the host cell

Most RNA viruses perform their replication and assembly in cell membranes by forming the viral replication complexes – virus factories (Nagy and Pogany, 2012; de Castro *et al.*, 2013). We discovered that part of proteins interacting with L-A virus are implicated in cell transport and endoplasmic reticulum (section 2.2.4), so we decided to identify localization of the Gag protein. To evaluate this, we fused a red florescent protein (mCherry) to the Gag protein C-terminal end in two different ways. First, mCherry was fused to Gag protein C-terminal end in such a way that during translation of this construct full length Gag protein with C-terminal mCherry domain (Gag+mCherry) was synthesized (Fig. 2.13 A1). Second, mCherry was fused to Gag protein C-terminal end after native Gag stop codon (Fig. 2.13 A2). Wild type Gag protein was synthesized from this construction (Fig. 2.13 A2), while Gag and mCherry fusion protein is synthesized only during the ribosomal frameshift (Gag-CFS-mCherry) (Fig. 2.13 A2). According to the literature, the frameshift rate is about 2 % (Dinman *et al.*, 1991).

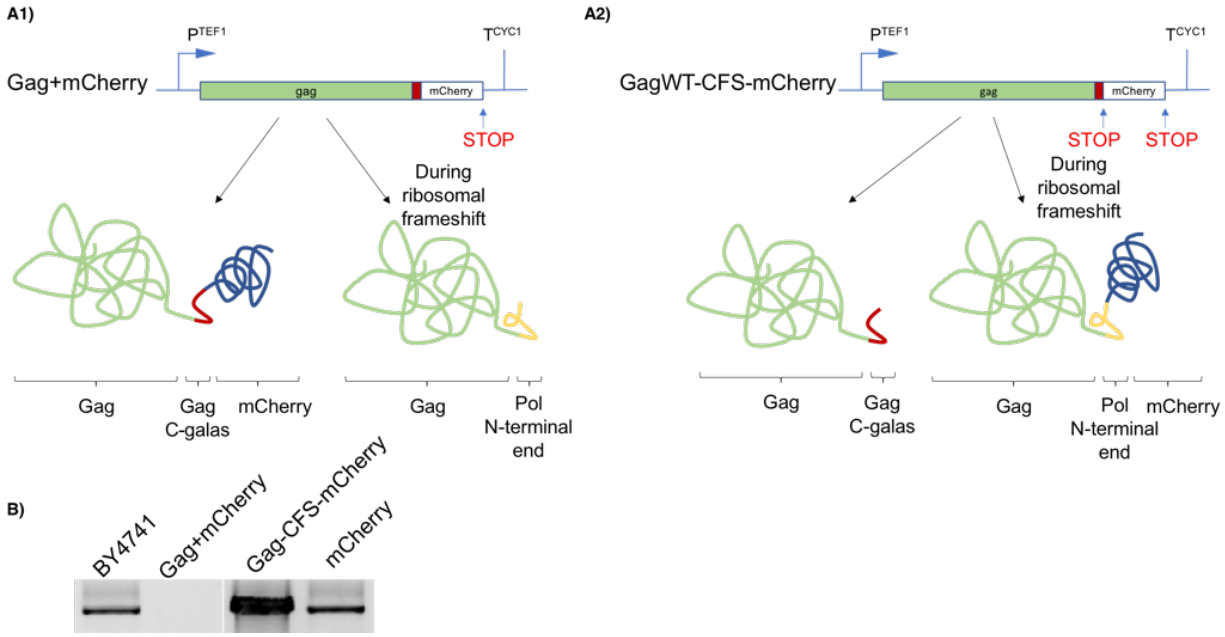

Fig. 2.13. Different constructions of fusion protein of Gag and mCherry and influence of these gene expression on native L-A virus genome replication. A1) Architecture of Gag+mCherry fusion protein and translated proteins; A2) Architecture of Gag-CFS-mCherry fusion protein and translated proteins; B) The influence of Gag+mCherry and Gag-CFS-mCherry protein expression on native L-A virus dsRNA genome replication.

During the expression, Gag+mCherry construct inhibits replication of native L-A virus (Fig. 2.14 B), while expression of Gag-CFS-mCherry construct-coding proteins stimulate L-A dsRNA virus genome replication. This result was expected, because 98 % of the translated protein was full-length Gag protein, while Gag-CFS-mCherry is synthesized in the case of only 2%. We identified that Gag+mCherry protein is localized in cytoplasm, while Gag-CFS-mCherry protein observes nuclear localization (Fig. 2.14). Gag-CFS-mCherry protein differs from Gag+mCherry protein only in peptide encoded

by ribosomal frameshift region (Fig. 2.8 C), still their localization in the cell is radically different. Altogether, we assume that the peptide at the Gag protein C-terminal end, which is synthesized during ribosomal frameshift, can redirect Gag protein to the nucleus.

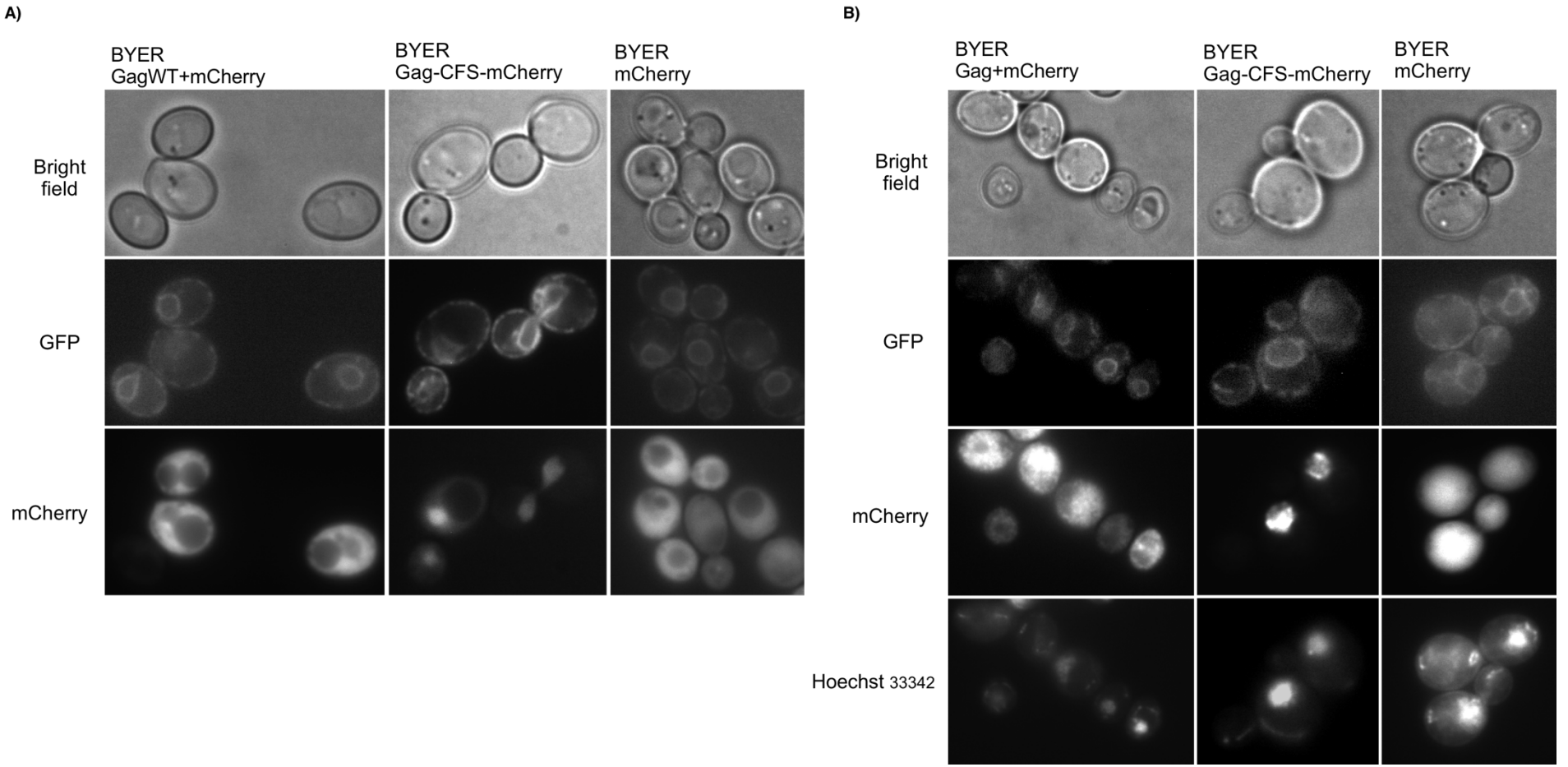

Fig. 2.14. Localization analysis of different Gag protein variants in the cell. A) Fluorescence analysis of living cells; B) Fluorescence analysis of fixed cells.

2.3 Discussion

The subject of this research is *Totiviridae* family viruses and their satellites, found in *Saccharomyces sensu stricto* clade. We show that spread and diversity of the *Totiviridae* family viruses among yeast of *Saccharomyces sensu stricto* clade correlates with host genome variability (2.1.1 and 2.1.2). Also, we have found that some of L-A virus genome sequence fragments are identical within all known viruses, however certain regions have quite high variability. The most variable regions of L-A virus Gag protein are located on the outer side of the virus capsid, while the amino acids forming the inner side are very conservative (2.1.3). We believe that the variability observed on the surface of the capsid is the result of virus adaptation to the host-cell protein variability and the conservative regions are critical to the virus by being involved in the virus replication cycle. We have shown that one of most conservative regions – ribosomal frameshift region – is necessary for ensuring of the replication of the dsRNA virus (2.2.1 and 2.2.2). We have also demonstrated, that this region may be related to the translocation of Gag protein into the nucleus (2.2.5). Furthermore, we have identified three host proteins that are potentially interacting with this region and are essential for virus replication and/or transcription (2.2.3 and 2.2.4). The molecular mechanisms on how C-terminus of Gag protein determines virus replication will be revealed in additional studies.

RNA viruses carry out their replication in viral replication complexes on subcellular membrane surfaces (Nagy and Pogany, 2012; de Castro *et al.*, 2013). During recent years, an important information has been obtained on the interaction of proteins of RNA viruses with nucleus structures (Rawlinson and Moseley, 2015). The dsRNA virus interaction with membrane structures was shown only in Reovirus (Fernández de Castro *et al.*, 2014; Fernández de Castro *et al.*, 2015). In this work, we discovered that part of the host cell proteins interacting with L-A virus capsid and necessary for replication of the virus are related to endoplasmic reticulum (2.2.3 and 2.2.4). Moreover, we found that the Gag protein of the L-A virus can be localized to the nucleus (2.2.5). Based on the obtained results, we assume that the L-A virus replication complex may be formed on endoplasmic reticulum membrane. This finding opens new horizons for research into replication cycle of L-A virus.

Two unique dsRNA satellites was identified in *S. paradoxus* (2.1.5). We confirmed that one of them encodes a killer toxin and is responsible for biocidal activity against other yeast; also, we suggested why the protein encoded by another dsRNA satellite does not exhibit biocidal activity against other yeast. We identified L-A virus and M satellite genome sequences in the *de novo* assembled transcriptomes from yeast. With the increase in the number of the publicly available yeast RNA-seq data with links to ecological niches, it will be possible to estimate spread of the L-A viruses and their M satellites in the different geographic areas. We have found that the most commonly used laboratory strains of the *S. cerevisiae* are infected exclusively with the ScV-LA-1 variant

of L-A virus (2.1.1). This discovery is particularly important, because transcriptomic or proteomic analysis data of a cell with L-A virus may be misinterpreted: observed changes of gene expression may not necessarily be related to the process of assay, but merely to the change in the amount and state of the virus in the cell.

We documented EVE formation of M satellite origin in *S. paradoxus* (EVE-M28), presumably enabled by the activity of the endogenous retrotransposons. Also, based on the comparison of *Saccharomyces sensu stricto* clade genomes, we have found that EVE-M28 is detected only in the European population of *S. paradoxus*. Similarity of EVE-M28 and M-28 satellite is observed not only in protein sequences but also at the nucleotide level, which points on possibility that EVE-M28 has evolved evolutionally recently. Meanwhile, other *Saccharomyces sensu stricto* clade homologues of M satellites of proteins detected in chromosomes have significant similarity only in protein level. Their relationship with transposons is also observed in genetic environment of these homologues, but are less obvious than EVE-M28 and requires additional studies for confirmation. Future studies will clarify the data on EVE elements in yeast and also can help answer question – is all the observed EVE cell components in yeasts related to transposon activity and what function they perform in the yeast cell.

CONCLUSIONS

1. Eight previously undescribed dsRNR viruses from *S. cerevisiae* and *S. paradoxus* was discovered by using developed universal strategy to obtain full length dsRNA virus genome sequences.
2. The variability of L-A dsRNA viruses in different strains of *Saccharomyces sensu stricto* clade significantly correlates with the host genome variability and indicates the possibility of virus and host coevolution.
3. Formation of the endogenous viral element EVE-M28 is likely related to the activity of transposons in the yeast cell.
4. The C-terminus of the ScV-LA-1 Gag protein is necessary for the assembly of the capsid capable replicate dsRNA genome.
5. Detected 15 host cell proteins modulating the level of replication of the dsRNA, among which 13 directly interacting with the proteins encoded by L-A virus and two involved in the a signal transduction in the cell.
6. The molecular determinant for Gag-CFS-mCherry fusion protein translocation to the nucleus is C-terminus of the Gag proteins.

REFERENCES

1. Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., Goecks, J. (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*. **44**(W1), W3–W10.

2. Babiano, R., Gamalinda, M., Woolford, J.L., de la Cruz, J. (2012) Saccharomyces cerevisiae Ribosomal Protein L26 Is Not Essential for Ribosome Assembly and Function. *Molecular and Cellular Biology*. **32**(16), 3228–3241.

3. Barros, M.H., da Cunha, F.M., Oliveira, G.A., Tahara, E.B., Kowaltowski, A.J. (2010) Yeast as a model to study mitochondrial mechanisms in ageing. *Mechanisms of Ageing and Development*. **131**(7–8), 494–502.

4. Bolger, A.M., Lohse, M., Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*. **30**(15), 2114–20.

5. Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., Boeke, J.D. (1998) Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)*. **14**(2), 115–32.

6. Breinig, F., Sendzik, T., Eisfeld, K., Schmitt, M.J. (2006) Dissecting toxin immunity in virus-infected killer yeast uncovers an intrinsic strategy of self-protection. *Proceedings of the National Academy of Sciences*. **103**(10), 3810–3815.

7. Bruenn, J.A., Warner, B.E., Yerramsetty, P. (2015) Widespread mitovirus sequences in plant genomes. *PeerJ*. **3**, e876.

8. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. (2009) BLAST+: architecture and applications. *BMC bioinformatics*. **10**, 421.

9. de Castro, I.F., Volonté, L., Risco, C. (2013) Virus factories: biogenesis and structural design. *Cellular microbiology*. **15**(1), 24–34.

10. Chang, M., Kang, H.J., Baek, I.J., Kang, C.M., Park, Y.S., Yun, C.W. (2013) Kdx1 regulates RCK1 gene expression by interacting with Rlm1 in Saccharomyces cerevisiae. *Biochemical and Biophysical Research Communications*. **435**(3), 350–355.

11. Chong, J.L., Chuang, R.Y., Tung, L., Chang, T.H. (2004) Ded1p, a conserved DExD/H-box trabslation factor can promote yeast L-A virus negative-strand RNA sunthesis in vitro. *Nucleic Acids Research*. **32**(6), 2031–2038.

12. Dihanich, M., Van Tuinen, E., Lambris, J.D., Biocenter, B.M. (1989) Accumulation of Viruslike Particles in a Yeast Mutant Lacking a Mitochondrial Pore Protein. *Molecular and Cellular Biology*. **9**(3), 1100–1108.

13. Dinman, J.D., Icho, T., Wickner, R.B. (1991) A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proceedings of the National Academy of Sciences of the United States of America*. **88**(1), 174–178.

14. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. **32**(5), 1792–7.

15. Fernández de Castro, I., Zamora, P.F., Ooms, L., Fernández, J.J., Lai, C.M.-H., Mainou, B.A., Dermody, T.S., Risco, C. (2015) Correction for Fernández de Castro et al., Reovirus Forms Neo-Organelles for Progeny Particle Assembly within Reorganized Cell Membranes. *mBio*. **6**(1), 2529.

16. Fernández de Castro, I., Zamora, P.F., Ooms, L., Fernández, J.J., Lai, C.M.-H., Mainou, B.A., Dermody, T.S., Risco, C. (2014) Reovirus forms neo-organelles for progeny particle assembly within reorganized cell membranes. *mBio*. **5**(1).

17. Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*. **9**(5), 397–405.

18. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic acids research*. **44**(D1), D279-85.

19. Fruhmann, G., Seynnaeve, D., Zheng, J., Ven, K., Molenberghs, S., Wilms, T., Liu, B., Winderickx, J., Franssens, V. (2017) Yeast buddies helping to unravel the complexity of neurodegenerative disorders. *Mechanisms of ageing and development*. **161**(Pt B), 288–305.

20. Fujimura, T., Ribas, J.C., Makhov, A.M., Wickner, R.B. (1992) Pol of gag-pol fusion protein required for encapsidation of viral RNA of yeast L-A virus. *Nature*. **359**(6397), 746–749.

21. García, R., Bermejo, C., Grau, C., Pérez, R., Rodríguez-Peña, J.M., Francois, J., Nombela, C., Arroyo, J. (2004) The global transcriptional response to transient cell wall damage in Saccharomyces cerevisiae and its regulation by the cell integrity signaling pathway. *The Journal of biological chemistry*. **279**(15), 15183–95.

22. Geuking, M.B., Weber, J., Dewannieux, M., Gorelik, E., Heidmann, T., Hengartner, H., Zinkernagel, R.M., Hangartner, L. (2009) Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science*. **323**(5912), 393–396.

23. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. **29**(7), 644–652.

24. Käll, L., Krogh, A., Sonnhammer, E.L.L. (2007) Advantages of combined

transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Research*. **35**(SUPPL.2), 429–432.

25. Katoh, K., Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*. **30**(4), 772–780.

26. Konovalovas, A., Servienė, E., Serva, S. (2016) Genome sequence of Saccharomyces cerevisiae double-stranded RNA virus L-A-28. *Genome Announcements*. **4**(3), 3–4.

27. Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., Tsai, I.J., Bergman, C.M., Bensasson, D., O'Kelly, M.J.T., van Oudenaarden, A., Barton, D.B.H., Bailes, E., Nguyen, A.N., Jones, M., Quail, M.A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., Louis, E.J. (2009) Population genomics of domestic and wild yeasts. *Nature*. **458**(7236), 337–41.

28. Louis, E.J. (2016) Historical Evolution of Laboratory Strains of Saccharomyces cerevisiae. *Cold Spring Harbor Protocols*. **2016**(7), pdb.top077750.

29. Lukša, J., Ravoitytė, B., Konovalovas, A., Aitmanaitė, L., Butenko, A., Yurchenko, V., Serva, S., Servienė, E. (2017) Different Metabolic Pathways Are Involved in Response of Saccharomyces cerevisiae to L-A and M Viruses. *Toxins*. **9**(8), 233.

30. McBride, R.C., Boucher, N., Park, D.S., Turner, P.E., Townsend, J.P. (2013) Yeast response to LA virus indicates coadapted global gene expression during mycoviral infection. *FEMS Yeast Research*. **13**(2), 162–179.

31. Nagy, P.D., Pogany, J. (2012) The dependence of viral RNA replication on co-opted host factors. *Nature Reviews Microbiology*. **10**(2), 137–149.

32. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*. **32**(1), 268–274.

33. Peter, J., De Chiara, M., Friedrich, A., Yue, J., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J. (2018) Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature*. **556**(7701), 339–344.

34. Petranovic, D., Tyo, K., Vemuri, G.N., Nielsen, J. (2010) Prospects of yeast systems biology for human health: integrating lipid, protein and energy metabolism. *FEMS yeast research*. **10**(8), 1046–59.

35. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*. **25**(13), 1605–12.

36. Pfeiffer, P., Radler, F. (1984) Comparison of the killer toxin of several yeasts and

the purification of a toxin of type K2. *Archives of microbiology*. **137**(4), 357–61.

37. Potgieter, A.C., Page, N.A., Liebenberg, J., Wright, I.M., Landt, O., van Dijk, A.A. (2009) Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. *Journal of General Virology*. **90**(6), 1423–1432.

38. Rawlinson, S.M., Moseley, G.W. (2015) The nucleolar interface of RNA viruses. *Cellular microbiology*. **17**(8), 1108–20.

39. Ribas, J.C., Wickner, R.B. (1998) The Gag domain of the Gag-Pol fusion protein directs incorporation into the L-A double-stranded RNA viral particles in Saccharomyces cerevisiae. *Journal of Biological Chemistry*. **273**(15), 9306–9311.

40. Rinaldi, T., Dallabona, C., Ferrero, I., Frontali, L., Bolotin-Fukuhara, M. (2010) Mitochondrial diseases and the role of the yeast models. *FEMS yeast research*. **10**(8), 1006–22.

41. Rodríguez-Cousiño, N., Esteban, R. (2017) Relationships and evolution of doublestranded RNA totiviruses of yeasts inferred from analysis of L-A-2 and L-BC variants in wine yeast strain populations. *Applied and Environmental Microbiology*. **83**(4).

42. Rodríguez-Cousiño, N., Gómez, P., Esteban, R. (2013) L-A-lus, a new variant of the L-A totivirus found in wine yeasts with klus killer toxin-encoding mlus double-stranded RNA: Possible role of killer toxin-encoding satellite RNAs in the evolution of their helper viruses. *Applied and Environmental Microbiology*. **79**(15), 4661–4674.

43. Rodríguez-Cousiño, N., Gómez, P., Esteban, R. (2017) Variation and distribution of L-A helper totiviruses in Saccharomyces sensu stricto yeasts producing different killer toxins. *Toxins*. **9**(10), 9–11.

44. Rowley, P.A., Ho, B., Bushong, S., Johnson, A., Sawyer, S.L. (2016) XRN1 Is a Species-Specific Virus Restriction Factor in Yeasts. *PLoS Pathogens*. **12**(10), 1–27.

45. Scannell, D.R., Zill, O.A., Rokas, A., Payen, C., Dunham, M.J., Eisen, M.B., Rine, J., Johnston, M., Hittinger, C.T. (2011) The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the *Saccharomyces sensu stricto* Genus. *G3: Genes|Genomes|Genetics*. **1**(1), 11–25.

46. Schmitt, M.J., Breinig, F. (2006) Yeast viral killer toxins: lethality and self-protection. *Nature reviews. Microbiology*. **4**(3), 212–221.

47. Schneider, C.A., Rasband, W.S., Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature methods*. **9**(7), 671–5.

48. Tercero, J.C., Wickner, R.B. (1992) MAK3 encodes an N-acetyltransferase whose modification of the L-A gag NH2terminus is necessary for virus particle assembly. *Journal of Biological Chemistry*. **267**(28), 20277–20281.

49. Wickner, R.B., Kryndushkin, D., Shewmaker, F., McGlinchey, R., Edskes, H.K. (2018) Study of Amyloids Using Yeast. *Methods in molecular biology (Clifton, N.J.)*. **1779**, 313–339.

50. Yue, J.X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M.C., Fischer, G., Durbin, R., Liti, G. (2017) Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*. **49**(6), 913–924.

LIST OF PUBLICATIONS

The thesis is based on the following original publications

1. **Konovalovas, A**., Servienė, E., Serva, S. (2016) Genome sequence of *Saccharomyces cerevisiae* double-stranded RNA virus L-A-28. *Genome Announcements*. **4**(3), 3–4.
2. Lukša, J., Ravoitytė, B., **Konovalovas, A.**, Aitmanaitė, L., Butenko, A., Yurchenko, V., Serva, S., Servienė, E. (2017) Different Metabolic Pathways Are Involved in Response of *Saccharomyces cerevisiae* to L-A and M Viruses. *Toxins*. **9**(8), 233.
3. Grybchuk, D., Akopyants, N.S., Kostygov, A.Y., **Konovalovas, A.**, Lye, L.F., Dobson, D.E., Zangger, H., Fasel, N., Butenko, A., Frolov, A.O., Votýpka, J., d'Avila-Levy, C.M., Kulich, P., Moravcová, J., Plevka, P., Rogozin, I.B., Serva, S., Lukeš, J., Beverley, S.M., Yurchenko, V. (2018) Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite *Leishmania*. *Proceedings of the National Academy of Sciences*. **115**(3), E506–E515.
4. **Konovalovas, A.**, Miglė, S., Servienė, E., Serva, S. (2016) Does yeast virus possess specificity towards replication of satellite virus? *FEBS Journal*. **283** (Special Issue), 146.
5. Vepštaitė-Monstavičė, I., Lukša, J., **Konovalovas, A.**, Ežerskytė, D., Stanevičienė, R., Strazdaitė-Žielienė, Ž., Serva, S., Servienė, E. Crossing the helper-satellite virus specificity border evidenced by the *Saccharomyces paradoxus* K66 killer system. *Viruses* (manuscript submitted to journal).

Results discussed in the PhD thesis were presented in these conferences

1. **A. Konovalovas**, M. Survilaitė, E. Servienė, S. Serva. Evolutionary relationship of ubiquitous Saccharomyces cerevisiae dsRNA viruses. *XIV International Lithuanian Biochemical Society International Conference*. 2016 Birželio 28-30, Druskininkai, Lithuania.
2. L. Aitmanaitė, **A. Konovalovas**, E. Servienė, S. Serva. Healing the yeast from L-A virus(es). *XIV International Lithuanian Biochemical Society International Conference*. 2016 Birželio 28-30, Druskininkai, Lithuania.
3. J. Lukša, I. Vepštaitė-Monstavičė, R. Stanevičienė, Ž. Strazdaitė-Žielienė, A. Žilakauskis, **A. Konovalovas**, S. Serva, E. Servienė. Persistence of killer viruses in natural environment, *7th EMBO Meeting: Advancing the life sciences*, 2016 m. Rugsėjo 10-13, Manheim, Germany.
4. **A. Konovalovas**, A. Žilakauskis, I.Vepštaitė-Monstavičė, E. Servienė, S. Serva. Evolutionary relationship of ubiquitous Saccharomyces cerevisiae dsRNA viruses. *8th EMBO Conference: From Functional Genomics to Systems Biology*. 2016 m. Lapkričio 12-15, Heidelberg, Germany.
5. **A. Konovalovas**, L. Aitmanaitė, I. Vepštaitė-Monstavičė, R. Stanevičienė, E. Servienė, S.Serva. New Double-stranded RNA Viruses from *Saccharomyces sensu*

stricto. *28th International Conference on Yeast Genetics and Molecular Biology*. 2017 rugpjūčio 27 d. – rugsėjo 1 d., Prague, Czech Republic

6. L. Aitmanaitė, **A. Konovalovas**, P. Medvedevas, E. Servienė, S. Serva. Complex relationships between yeast totiviruses. *XV International Lithuanian Biochemical Society International Conference*. 2017 June 26-29, Dubingiai, Lithuania.

FINANCIAL SUPPORT

The work was supported by the Research Council of Lithuania grants MIP-035/2014 and SIT-7/2015.

CURRICULUM VITAE

Date and place of birth	13th of October, 1988, Vilnius
Contact e-mail:	aleksandras.konovalovas@gf.vu.lt
Work Address:	Department of Biochemistry and Molecular Biology, Institute of Biosiences, Life Sciences Center, Vilnius University Saulėtekio al. 7, LT-10223, Vilnius, Lithuania

Education:

2013-2018 PhD studies in Biochemistry, Institute of Biosiences, Life Sciences Center, Vilnius University

2011-2013 Vilnius University, Faculty of Natural Science, Master degree in Genetics

2007-2011 Vilnius University, Faculty of Mathematics and Informatics, Bachelor degree in Informatics (Bioinformatics)

Professional Experience:

2010-2011 – Trainee, Institute of Biotechnology, Vilnius University

2011-2013 – Junior Researcher, Thermo Fisher Scientific

2013-2017 – Ph.D student, Institute of Biosiences, Life Sciences Center, Vilnius University

2017-to date – Junior Researcher, Institute of Biosiences, Life Sciences Center, Vilnius University

Research Interests: Virology, Microbiology, Evolution, Bioinformatics, Virus–Host Interactions

Honours/Scholarships:

"Promotion of Students" Scientific Activities" project scholarship provided by Lithuania Research Council (2010)

"Thermo Fisher Scientific" scholarship (2011)

M.Sc. *Cum Laude* (2013)

Best poster award, 14th International Lithuanian Biochemical Society Conference (2016)

ACKNOWLEDGEMENTS

I am grateful to my supervisor prof. Dr. Saulius Serva for the proposed topic of dissertation, trust, invaluable discussions and help in preparing this dissertation. I am grateful to all the colleagues from the lab for discussions and cooperation.

I would like to thank to dr. Elena Servienė and all members of the Genetic laboratory from the Nature Research Centre.

I am also thank to prof. dr. Edita Sužiedelienė and to all my colleagues in the Department of Biochemistry and Molecular Biology for advices and help with experiments.

Many thanks to dr. M. Valius and A. Kaupinis for performed proteomics analysis, dr. Vyacheslav Yurchenko and dr. Danyil Grybchuk for discussions, advises and great cooperation, prof. dr. Kęstutis Sužiedelis for mCherry plasmid.

I also thank to dr. Laura Kalinienė and dr. Eglė Lastauskienė for the critical reading and comments on this doctoral thesis.

My special thanks go to my wife whose patience and understanding helped to prepare this dissertation.