

## Online Resources

### Applied Microbiology and Biotechnology

Robust signal peptides for protein secretion in *Yarrowia lipolytica*: identification and characterization of novel secretory tags.

Ewelina Celińska\*<sup>1</sup>, Monika Borkowska<sup>1</sup>, Wojciech Białas<sup>1</sup>, Paulina Korpys<sup>1</sup>, Jean-Marc Nicaud<sup>2</sup>

<sup>1</sup>Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, ul. Wojska Polskiego 48, 60-627 Poznań, Poland

<sup>2</sup>Micalis Institute, INRA-AgroParisTech, UMR1319, Team BIMLip: Integrative Metabolism of Microbial Lipids, Domaine de Vilvert, 78352 Jouy-en-Josas, France

\*Corresponding author: Ewelina Celińska, phone: +48618466007, e-mail:

[ewelina.celinska@up.poznan.pl](mailto:ewelina.celinska@up.poznan.pl), ORCID 0000-0001-8372-8459

## Online Resource ESM\_1. *E. coli* strains and bacterial plasmids used in this study

<i>E. coli</i> JM109	characteristics: <i>F'</i> ( <i>traD36, proAB+, lacIq, Δ(lacZ)M15</i> ), <i>endA1, recA1, hsdR17(rk-mk+)</i> , <i>mcrA, supE44, λ, gyrA96, relA1, Δ(lacproAB), thi-1</i>
----------------------	--

All the following strains are descendants of JM109 and bear the corresponding GGVs (TOPO backbone + insert)

<i>E. coli</i> strain No	Type and signature of the GG element	Backbone Vector	Resistance	Insert type
GGE 0065	GGV0065	TOPO	KanR	SSP 1
GGE 0066	GGV0066	TOPO	KanR	SSP 2
GGE 0063	GGV0063	TOPO	KanR	SSP 3
GGE 0067	GGV0067	TOPO	KanR	SSP 4
GGE 0068	GGV0068	TOPO	KanR	SSP 5
GGE 0069	GGV0069	TOPO	KanR	SSP 6
GGE 0070	GGV0070	TOPO	KanR	SSP 7
GGE 0071	GGV0071	TOPO	KanR	SSP 8
GGE 0072	GGV0072	TOPO	KanR	SSP 9
GGE 0062	GGV0062	TOPO	KanR	SSP 10
GGE 0073	GGV0073	TOPO	KanR	SoAMY_no ATG_Bsal_del
GGE 0064	GGV0064	TOPO	KanR	TIGAMY_no ATG
GGE 0003	GGV0003	TOPO	KanR	P1 TEF
GGE 0016	GGV0016	TOPO	KanR	T3 Lip2
GGE 0013	GGV0013	TOPO	KanR	M <i>URA3ex</i>
GGE 0037	GGV0037	TOPO	KanR	InsUP_zeta NotI
GGE 0038	GGV0038	TOPO	KanR	InsDOWN_zeta NotI
GGE 0029		pSB1A3-RFP	AmpR	pSB1A3-RFP (iGEM collection)

All the following strains are descendants of JM109 and bear the following GGVA's ( pSB1A3 backbone + GGA with the following constant elements: *InsUP-zeta NotI\_M\_URA3ex\_pTEF\_\**variable insert\*\_T3 tLip2\_InsDOWN\_zeta NotI)

<i>E. coli</i> strain No	Type and signature of GG element	Backbone Vector	Resistance	*variable insert* type
GGE 0076	GGVA0076	pSB1A3	AmpR	SSP1 TIGAMY
GGE 0078	GGVA0078	pSB1A3	AmpR	SSP2 TIGAMY
GGE 0079	GGVA0079	pSB1A3	AmpR	SSP3 TIGAMY
GGE 0080	GGVA0080	pSB1A3	AmpR	SSP4 TIGAMY
GGE 0081	GGVA0081	pSB1A3	AmpR	SSP5 TIGAMY
GGE 0082	GGVA0082	pSB1A3	AmpR	SSP6 TIGAMY
GGE 0083	GGVA0083	pSB1A3	AmpR	SSP7 TIGAMY
GGE 0084	GGVA0084	pSB1A3	AmpR	SSP8 TIGAMY
GGE 0085	GGVA0085	pSB1A3	AmpR	SSP9 TIGAMY
GGE 0086	GGVA0086	pSB1A3	AmpR	SSP10 TIGAMY
GGE 0087	GGVA0087	pSB1A3	AmpR	SSP1 SoAMY
GGE 0088	GGVA0088	pSB1A3	AmpR	SSP2 SoAMY
GGE 0089	GGVA0089	pSB1A3	AmpR	SSP3 SoAMY
GGE 0090	GGVA0090	pSB1A3	AmpR	SSP4 SoAMY
GGE 0091	GGVA0091	pSB1A3	AmpR	SSP5 SoAMY
GGE 0092	GGVA0092	pSB1A3	AmpR	SSP6 SoAMY
GGE 0093	GGVA0093	pSB1A3	AmpR	SSP7 SoAMY
GGE 0094	GGVA0094	pSB1A3	AmpR	SSP8 SoAMY
GGE 0095	GGVA0095	pSB1A3	AmpR	SSP9 SoAMY
GGE 0096	GGVA0096	pSB1A3	AmpR	SSP10 SoAMY

<b>Online Resource ESM_2. <i>Y. lipolytica</i> strains and expression cassettes used in this study</b>		
<b><i>Y. lipolytica</i> Po1g</b>	Genotype: <i>Mata</i> , <i>leu2-270</i> , <i>ura3-302::URA3</i> , <i>xpr2-332</i> , <i>axp-2</i> <i>Po1g</i> is a derivative of the wild-type strain W29 (ATCC 20460; CLIB89); part of YLEX Expression Kit (Yeastern Biotech Co. Ltd., Taiwan)	
The following strains were transformed with pYLSC vector bearing respective gene of interest devoid of its native SP <i>pBR322</i> backbone, hybrid promoter ( <i>hp4d</i> ), <i>XPR2</i> pre-leader, <i>LEU2</i> —selection marker; <i>SoAMY</i> and <i>TIGAMY</i> genes were cloned in <i>SfiI/KpnI</i> sites <a href="http://www.yeastern.com/Products.php?pid=273">http://www.yeastern.com/Products.php?pid=273</a> (Yeastern Biotech Co. Ltd., Taiwan)		
<b><i>Y. lipolytica</i> 1.1.2</b>	<i>Po1g</i> derivative bearing pYLSC- <i>SoAMY</i> construct Used as a positive control in <i>microSIT</i> assay	
<b><i>Y. lipolytica</i> 2.1.2</b>	<i>Po1g</i> derivative bearing pYLSC- <i>TIGAMY</i> construct Used as a positive control in <i>microSNT</i> assay	
<b><i>Y. lipolytica</i> JMY2101</b>	Genotype: <i>MATa</i> , <i>ura3-302</i> , <i>xpr2-322+pXPR2-SUC2</i> <i>JMY2101</i> is a derivative of the wild-type strain W29 (ATCC 20460; CLIB89)	
The following strains were transformed with corresponding GGA (GGA released from <i>pSB1A3</i> backbone) and contain the following constant elements integrated with the genome: <i>InsUP-zeta_NotI_M_Ura3ex_pTEF_*variable insert*_T3 tLip2_InsDOWN_zeta_NotI</i>		
<b><i>Y. lipolytica</i> strain No</b>	<b>Type and signature of the GG element</b>	<b>*variable insert* type</b>
GGY01	GGA0076	SSP1 TIGAMY
GGY02		
GGY03		
GGY04		
GGY05		
GGY10	GGA0078	SSP2 TIGAMY
GGY11		
GGY12		
GGY13		
GGY14		
GGY24	GGA0079	SSP3 TIGAMY
GGY27		
GGY29		
GGY30		
GGY35		
GGY37	GGA0080	SSP4 TIGAMY
GGY39		
GGY40		
GGY41		
GGY42		
GGY45	GGA0081	SSP5 TIGAMY
GGY46		
GGY48		
GGY49		
GGY50		
GGY60	GGA0082	SSP6 TIGAMY
GGY61		
GGY68		
GGY65		
GGY67		
GGY70	GGA0083	SSP7 TIGAMY
GGY71		
GGY72		
GGY73		
GGY77		
GGY97	GGA0084	SSP8 TIGAMY
GGY98		
GGY101		
GGY102		
GGY103	GGA0085	SSP9 TIGAMY
GGY125		
GGY126		
GGY128		
GGY130		

GGY131		
GGY155		
GGY156		
GGY157	GGA0086	SSP10 TIGAMY
GGY159		
GGY160		
GGY172		
GGY173		
GGY175	GGA0087	SSP1 SoAMY
GGY183		
GGY195		
GGY207		
GGY208		
GGY212	GGA0088	SSP2 SoAMY
GGY220		
GGY226		
GGY255		
GGY257		
GGY262	GGA0089	SSP3 SoAMY
GGY264		
GGY270		
GGY304		
GGY306		
GGY310	GGA0090	SSP4 SoAMY
GGY314		
GGY316		
GGY331		
GGY332		
GGY337	GGA0091	SSP5 SoAMY
GGY338		
GGY339		
GGY345		
GGY346		
GGY347	GGA0092	SSP6 SoAMY
GGY342		
GGY343		
GGY352		
GGY353		
GGY356	GGA0093	SSP7 SoAMY
GGY359		
GGY364		
GGY370		
GGY373		
GGY380	GGA0094	SSP8 SoAMY
GGY383		
GGY389		
GGY394		
GGY395		
GGY397	GGA0095	SSP9 SoAMY
GGY400		
GGY403		
GGY412		
GGY413		
GGY414	GGA0096	SSP10 SoAMY
GGY417		
GGY418		

<b>Online Resource ESM_3. Synthetic Golden Gate Primers (GGPs) and Golden Gate Fragments (GGFs) used in this study</b>		
<b>GGPs</b>		<b>Comments</b>
<b>Amplification of constant elements of GGAs</b>		
GGP_InsertUP_zeta_A_F_Not	GGTCTCTGCCTGCGGCCGctgtcgggaaccgc	Amplification of GGFs
GGP_InsertUP_zeta_B_R	GGTCTTACCTctagcaaagtcttctgtgc	
GGP_M_Ura3ex_B_F	GGTCTTAGGTataactctgtatagcatacattatacgaag	
GGP_M_Ura3ex_C_R	GGTCTTAGCTtcgcttcggataactcc	
GGP_P1_TEF_C_F	GGTCTTAGCTgggttgccggcg	
GGP_P1_TEF_D_R	GGTCTCATTcttcgggtgtgagttgac	
GGP_T3_LIP2_E_F	GGTCTGTACgttctgtgtatctaaagctattatc	
GGP_T3_LIP2_L_R	GGTCTTACTCcgatttcttagaggaacgc	
GGP_InsertDOWN_zeta_L_F	GGTCTGAGTcatgtgtaaacactcgctctg	
GGP_InsertDOWN_zeta_M_R_Not	GGTCTCTCGACGGCCGCactgaaggccttctgtgag	
<b>Amplification of a gene in G1 position</b>		
GGP_G_SoAMY_noSP_x_F	GGGGGTCTctGCCAGAAGACCCCCAC	
GGP_G_SoAMY_noSP_K_R	GGGGGTCTctGTACCTAGTGGTGGTGGTGG	Amplification of variable insert R
GGP_G_TIGAMY_noSP_x_F	GGGGGTCTctGCCAGACCAGAAACAACCGTTC	
GGP_G_TIGAMY_noSP_K_R	GGGGGTCTctGTACTCAATGATGATGATGATGATGTCTC	Amplification of variable insert R
<b>BsaI deletion</b>		
SoAMY BsaI del noSP 1 F	CAGAAGGACCCCACTTCTAGATGGACGAAAC	Deletion of BsaI site from SoAMY gene sequence
SoAMY BsaI del 1 R	CCAGTTGGTCAGCTCAGTTCCCGtGtCACGTTTCG	
SoAMY BsaI del 2 F	CATGGTTGAGTTTCGAAACGTGcaCaCGGGAAC	
SoAMY BsaI del 2 R	CTAGTGGTGGTGGTGGTGCAGCTTAGCG	
<b>Amplification of variable insert</b>		
GGP_SP_B03564g_D_F	ATGAAGTTCACATTTGCTGCCGTTACCGCCGCTGGCCTCGTCCGCCATTGCC	Amplification of variable insert F
GGP_SP_D20680g_D_F	ATGAAGTTCTCCACCGCCCTTCTGGCTCTGGCCGCGTCCGCACTGCC	
GGP_SP_E22374g_D_F	ATGAAATCTCTATTGCTGTGCTGTGGCGGTCCCGGCCACTGCC	
GGP_SP_D06039g_D_F	ATGAAGTCTCAGCGGTCTCAATCGCTGCTGCCCTGGCCTCGTGGTGGCTGCC	
GGP_SP_D06149g_D_F	ATGAAGCTCTCCATCGTTCTCGTGGCTCTGGCAGCCGTCTCCTCTGCC	
GGP_SP_Lip2NATIVE_D_F	ATGAAGCTTCCACCATCTTTTACAGCCTGCGCTACCCTGGCTGCTGCC	
GGP_SP_Lip2pre 3xXA_D_F	ATGAAGCTGTCTACCATTCTGTTTACCGCTTGCTACTCTGGCTCTCGCTTGGCC	
GGP_SP_SoAMY NATIVE_D_F	ATGAAGGTGCTCGCCCTGCTGGTTACTGTCTGCTTTTCCGTTGCCCTGCTGCC	
GGP_SP_TIGAMY NATIVE_D_F	ATGTTATTCCAACCGACTTTGTGCGCGCCCTTGGACTCGCCGCTTGATCGTCCAAGGCGGAGATGCC	
GGP_SP_XPR2pre_D_F	ATGAAGCTCGCTACCGCCTTACTATTCTCACGGCCGTTCTTGGCC	
<b>GGFs</b>		

SoAMY nt no SP	<p>CAGAAGGACCCCACTTCTAGATGGACGAAACACTATCGTTACCTTTCGAGTGGAAAGTGGGCCGACATTGCTCCGAGTGTGAGAACTTCTGTGGTGAAGAACCTTTGCC GGTGTTCAGGTGTCTCCCTGCTGAGTCCGTGGTCTGTAGGGCCGACCTGGTGGGAGAAGTACCAGCTGTCTCTACACCTCAACAACCGAGGCGGAGACGAGGCCG TTTCTCCGATATGTTTTCCGATGTAACAACGTGGGAATCCGAATTTACGTGGACCTGGTCCCAACACCATGGTACTTTCGAACGGCCAGGGATCTGCCGGTAACACTGCGAC CCCTTCCAAAGTCTTACCCTGCTGTCTCGTACACTCTGAGAACTCCATACCTCCTGTGACATCGATTACACTGACTGTCTTCCATTGAAACTGCGAGCTGACCGGCTCAAG GACCTGGATCAGTCCAGGACTACGTGCGAGGAAAGATCGAGGAGTACATGAAACCACTATTCTCTGGGAGTGGCCGGTTTTCGAGTGGACGCCGTAAGCATATGTGGCC CGCCGACCTCCAGGCTATCTCGGCTCCCTCAACGACCTGTCTACCGACCAGGATTTGCTTGGGTGCCGAGCTTTTCATCTTTCAGGAAGTGATTGACACTTCTACCGATCTG TCAGAACAACCTGAGTACACCGTTTTCGGCAAGGTGTGCGAGTTCCTCTTGGCAACGACCTGGGACCCGCTTTCGAGGAGAGAACCCTCTCCATTACCTGAAGAAGTGGGGTA CTGAGTGGGGCTGCTCGACGGTGGCGATACCGTTTTCTGTTCTGGACAACCACGATAACGAGCGAGACTCTCAGATTTCTGCTTACCAACGATAAGCCCTACAAGGCTG CCATGGCCTTCATGCTCGTCACTCTACGACACCACTACCGAGTCTGTCTTACAAGTTTGACTCTCGGATCAGGGTCCCCCTTCAACGGCGACGATATTCGTCTCCCG AGTTCCGCTCCGACGGAGCCTGTACCAACGGCTGGTTTTGCCAGCATCGATGGTCCCCTGTCTTCAACATGGTTGAGTTTTGAAACGTGGTCTCGGAACTGAGCTGACCACT GGTGGGACAACGGTTCGAGCAGATTGCCTTCTCGAGGTGACAAGGGCTTTACGCCGCTACTGTCAACGAGGATATCGTACCTTACTACTACCGGACTGCCGACGGTTC CTACTGTGATGTCATGCTGGCTCTGTGGTAAACGGCGCTGCACCGAAAGACTCTGACCGTGTCTGGCGGTGAGTCTACGTTGAGCTGGTGGTGTGAGCTGGAGGCTGC TGTCGCTATTCAGTCAACGCTAAGCTGACCAACCACCACCACCTAG</p>	codon optimized sequence encoding SoAMY truncated at 5' end to eliminate its native SP	
TIGAMY nt no SP	<p>AGACCAGAAACAACCGTTCTCACGCAACCGGTTCTTTGGATGATTTCTGGCTGCACAATCACCAATCGCTTCCAAGGATTTTGAACAACATTGGTCCATCTGGTGCATATTC AGAAGGTGTTAATCCAGGTGTTGTTATTGCTTCCACATTAAGCAAGTCCAGATTACTTCTACACTTGGGTTAGAGATGCTGCATTGACAGTTCAATATTTGGTTGAAGAATTGG TTGCTGGTAATGCTTCTTTGCAATCTTGTATCCAAGATTACATCTTTCACAAGCTAGATTGCAAACTGTTGAAAATCCATCAGGTTCTTTGCTTCAGGTGGTTTAGGTGAACCAA AGTTCCATGTTGATGAAACTGCTTTACTGATTCTTGGGGTAGACCACAAAGAGATGGTCCACCATTGAGAGCTATTGCAATGATCTTTCGCTAACTATTTGATCGATAACGGT CATCAATCAACAGTTGAAGATATCATCTGGCCAAATCGGTAGAAACGATTTGTACATGCTTCTCAACATTGGAACGAAACTACATTCGATATCTGGGAAGAAGTTCATTTTCATC TTTCTTACTACAGCAGTTCAATATAGAGCTTTAGTCAAGGTTCTGCTTGGCATCAAAGTTGGGTACACTTGTGATAACTGTGGTTCTCAAGCACCAAAATTTGTGTTTCT GCAATCTTACTGGACTGGTTCACATATTTGGCTAATACTGGTGGTGGTAGATCTGGTAAAGATGTTTCAACTATTTAGGTGTTATTGGTCTTTGATCCAAATGCTGATTGTG ATGATGTTACATTTCAACATGTTTCAGCTAGAGCATTAGCTAACCATAAAGCAAGTTGTTGATCTTTAGATCTATGCAATTAAGTCTGGTATTCCATCAGGTTCTGCAGTTG CTGTTGGTAGATATCCAGAAGATGTTTACCAAGGTGGTCACTCATGGTACTTACTACAGCTGCAGCTGCAGAACAAATTGACGATGCAATCTATCAATGGAACCATGTTGGTCA TATCGATATCAACGCTGTTAATTTGGATTTCTTAAATCTATCTATCCATCAGCTGCAGAAGGTACATACACTTCTGATTCTACTTTCCAAGATATCATCTCAGCAGTTAGAACA TATGCTGATGTTTCTGTCTGTTATTGAAAAGTACACTCCACCAGATAATTTGTTGGCTGAACAATCCATAGAGAAACAGGTATCCATTGTCAGCTGCATCTTAACTTGGTCT TATGCTGATTGAATACAGCTGCACAAAGAGAGCTTATTGTTCCATCACCATGGAACCTCAAACCTCTACAGATTTGCTGATAAGTGTTCAGCAACTTCTGCTACAGGTCATA CGCAACCAACTAATACAGCTTGGCCAACCTACAACCTCAACCACCAGAAAGACCAGCTTACTCCACCATCTGAAGTTACTTTGACTTTTAAATGCAATGGTTGATACAGCTTTCCG GTCAAAACATATATTTGGTGGTCTATTCCAGAATTAGGTTTCATGGGACCCAGCAAACTGCTTTGTTAATGTCAGCAAACTCTGGACTCAGGTAATCCAGTTGGACATTGTCA ATTTCTTACCAGCTGGTACTTCTTTCGAATACAAGTTATTAGAAAGGATGATGGTTCATCTGATGTTGTTGGGAATCTGATCCAATAGATCATACAATGTTCCAAAAGATTGC GGTCTAACACAGCTACAGTAACTCTTGGTGGAGACATCATCATCATCATTGA</p>	codon optimized sequence encoding TIGAMY truncated at 5' end to eliminate its native SP	
GGF_SP_B03564g_D_F	SSP_1	GGGGTCTCtAATGAAGTTCACATTTGCTGCCGTTACCGCCGCTGGCCTCGTCCGCCATTGCCaGAGACCCCC	F elements to be aligned with their complementary strand R – not for amplification
GGF_SP_D20680g_D_F	SSP_2	GGGGTCTCtAATGAAGTTCACACCGCCTTCTGGCTGTGGCCGCTGCCACTGCCaGAGACCCCC	
GGF_SP_E22374g_D_F	SSP_3	GGGGTCTCtAATGAAATCTCTATTGCTGCTGCTGGCCGCTCCCGGCCACTGCCaGAGACCCCC	
GGF_SP_D06039g_D_F	SSP_4	GGGGTCTCtAATGAAGTTCACGCGGTCTCAATCGCTGCTGCCCTGGCCTCGTGGTGGCTGCCaGAGACCCCC	
GGF_SP_D06149g_D_F	SSP_5	GGGGTCTCtAATGAAGTTCATCGTTCTCGTGGCTGTGGCAGCGTCTCTGCTGCCaGAGACCCCC	
GGF_SP_Lip2NATIVE_D_F	SSP_6	GGGGTCTCtAATGAAGTTCACCATCTTTTACAGCCTGCGTACCCTGGCTGCTGCCaGAGACCCCC	
GGF_SP_Lip2pre 3xXA_D_F	SSP_7	GGGGTCTCtAATGAAGTGTCTACCATCTGTTTACCCTTGTGCTACTCTGGCTCTGCTTGGCcaGAGACCCCC	
GGF_SP_SoAMY NATIVE_D_F	SSP_8	GGGGTCTCtAATGAAGTGTCTGCCTGCTGGTACTGTCTGCTTTCCGTTGCCCTGCCaGAGACCCCC	
GGF_SP_TIGAMY NATIVE_D_F	SSP_9	GGGGTCTCtAATGTTATTCCAACCGACTTTGTGCGCGCCCTTGGACTCGCCGCTTGTCTGCTCAAGGCGGAGATGCCaGAGACCCCC	
GGF_SP_XPR2pre_D_F	SSP_10	GGGGTCTCtAATGAAGTGTCTACCGCCTTACTATTCTACGCGCTTCTTGGCcaGAGACCCCC	
GGF_SP_B03564g_x_R	SSP_11	GGGGTCTCtGGCAATGGCGACGAGGCCAGCGCGCGGTAACGGCAGCAAATGTGAATTCATTaGAGACCCCC	R elements to be aligned with their complementary
GGF_SP_D20680g_x_R	SSP_12	GGGGTCTCtGGCAGTGGCAGCGCGGCCAGCCAGAGGCGGTTGAGAACTTCATTaGAGACCCCC	
GGF_SP_E22374g_x_R	SSP_13	GGGGTCTCtGGCAGTGGCCGGGACCGCCAGCAGCAGCAATAGAGATTTCATTaGAGACCCCC	
GGF_SP_D06039g_x_R	SSP_14	GGGGTCTCtGGCAGCCACCAGCGAGGCCAGGCGAGCAGGATTGAGACCGTGTGAGAACTTCATTaGAGACCCCC	

GGF_SP_D06149g_x_R	SSP_15	GGGGTCTCTGGCAGAGGAGACGGCTGCCAGAGCCACGAGAACGATGGAGAGCTTCATTaGAGACCCC	strand F – not for amplification
GGF_SP_Lip2NATIVE_x_R	SSP_16	GGGGTCTCTGGCAGCAGCCAGGGTAGCGCAGGCTGTGAAAAGGATGGTGGAAAAGCTTCATTaGAGACCCC	
GGF_SP_Lip2pre 3xXA_x_R	SSP_17	GGGGTCTCTGGCAAGAGCGAGAGCCAGAGTAGCACAAAGCGGTAACAGAATGGTAGACAGCTTCATTaGAGACCCC	
GGF_SP_SoAMY NATIVE_x_R	SSP_18	GGGGTCTCTGGCAGAGGCAACGGAAAAGCAGACAGTAACCAGCAGGGCGAGCACCTTCATTaGAGACCCC	
GGF_SP_TIGAMY NATIVE_x_R	SSP_19	GGGGTCTCTGGCATCTCCGCCTTGGACGATCAAGGGCGGAGTCCAAGGGCCGCGCACAAAGTCGGTTGGAATAACATTaGAGACCCC	
GGF_SP_XPR2pre_x_R	SSP_20	GGGGTCTCTGGCAAGAACGGCCGTGAGAATAGTAAAGCGGTAGCGAGCTTCATTaGAGACCCC	

**Online Resource ESM\_4. (A)** A list of the 54 polypeptides identified through *Yarrowia lipolytica* CLIB122 genomic sequence-mining bearing the motif: MK x{7-15} (x,A/P){2-10} x{10-120} KR. Blue – pre-leader, purple - XA/XP dipeptide motif, black – pro-leader, red – KR dipeptide; the names of the 16 polypeptides eliminated from further analysis (alignment and consensus sequence determination) are crossed out. **(B)** The output data from SignalP and **(C)** Multiple Sequence Alignment (ClustalW) run with the remaining 38 polypeptides.

**(A)**

```

>YALIOA00154g
MKLSNIFALATVVALAAPQLITPPPTAVEKRDDQAVLGADVWIQTMPNGHEKR
>YALIOA08360g
MKPSVIIALAVACMAAPMSPQGAYKDALDIINGQASKFNDAGLKKR
>YALIOA09262g
MKPSVILSLAVACLAAPMAPQGAYKDALDIINGQSSKFSDSSLRKR
>YALIOA10208g
MKLLSILLWISIGLCAAPQLENDYAKLLSMEHLEKR
>YALIOA11198g
MKFSAIAIAAASIVVADPIFLGKPAAGVDPNTFWSQSTPTPAPPPPAPTTQAPPPPAPTSSQAAQETCIAPASTNEKR
->YALIOA17534g
MKIPGTVPLVESTEILSPTSSDDPDDPLNWSPRRKLSMFCMVLVCAFCVPSASIFSVFIPISQTGLPLATLNQTGYMFLLFLGCIVHPLSVKYGKR
>YALIOA20350g
MKLSTILTFTACATLAAALPSPITPSEAAVLQKR
->YALIOA20790g
MKFSLKSTTQTPDSPPEEEETLAEEPTDPAIRTSFGDVVLTDVSCKDELKEVHMGYLNDQCYVIDNFAPTLLSQYVHIMAGRIKFTISEMTRMADRAFLVLTELEPPKKR
->YALIOB03564g
MKFTFAAVTAALSAMALGGLVDLGVKR
>YALIOB05104g
MKLSTIAISAILSVALAAPAQEELRSTAMLIQSEEFQRLNYDEGLSLDRDELRLVETAPGVRTWMTEAEKLVLKR
->YALIOB07293g
MKKQNLNALYPHPYFYVDVVPIHNPISVYFLCRWIFAGNTSPEYTAEYDVGVVRAVGDAREGLWKMGFGGLSRSEPSWDIRTKRLGLHVEEGVYAPEVTQRFRRQRQHFAERAKAEEEELKR
>YALIOB08778g
MKVAHIQLAAMVATALAAVLAIDYGQEYTKAALSPGINFEIVLTQDSKRKQPSAIGFKGKADSKFLERVYGSPAVLMEPRFPSDVVLYHKR
->YALIOB09713g
MKRKMEEDIPTEPPSAKQPRQSMASVSACVCRKRKSKCDQKFPSCSRCLAAGVVCMGLDAATKR
>YALIOB15510g
MKLSTLVLAALTAVAVADDYKSICDVDNHCPEDKPCCSYGECGTGAYCLGGCDPKSSEFELKSCMPMPVCKKR
>YALIOB16500g
MKFNPVISAVAMAALAQALLPQNVDAEAFNTLLTDNEANSPNHHAKAFEKIVIANEDESEIAPLVMNDKEQVPDSYIVVFKEDVDASTASHELWVEQVHTEQVASLAKR
>YALIOB22880g
MKFTLLLAGLSLVAAPADASASASYNLVNIINKHQGHHAKR
->YALIOC14564g
MKILTIDVFGLFVPRPSVPAQYLRIVQQHEKCSATVAQVQAGFHKAFKR
>YALIOC15851g
MKSAIGYMAVAVAAASCQFTVDSKNYDLSAISGPKSVEYTIETPPSKR
>YALIOD00759g
MKTSHTVALLAVATAAAIRVDSGNAVTLDIVKR
>YALIOD02024g
MKLTTTTLLSLASASFRPFSGHEIAKAARENVPSLGEQAILPKSPFKAASPLQNKASSGGTSVATSSVDCKRFVEIQDGGDVSYVNTQLGSSQAEYRMLVDTGSYMWVYSENCTSTACMKR
>YALIOD02981g
MKLATSLTLFTAVLAPLAAPAPDVAPAVAEGSASASASAILSVVSKQSKFSALQKR
>YALIOD04851g
MKFSTLSVAALAASASASALDPEVEKGNAFVGDKR
>YALIOD06039g
MKFSASIAALASLVAAADVPSIDIVGNKFFSNNSQFYIRGVAYQDSANETAKETFSDPLSDGDACKR
>YALIOD06149g
MKLSIVLVAAAVSAQPIVGVYDRDSAETVYQHFADAHVETRDLTKR
>YALIOD08206g
MKFSATILTALASAVANYAPIKVECPKDVNSRPDGLSHQEADYINERHKKTSKALREYLGLDIKARGNTFDTDSFLNETSPKLAIAAGGGFRAMLVGAGVIAALDKR
>YALIOD20680g

```



MKFSTALLAALAAVATAQPIVGHHAHHQHKR  
>YALIO022616g  
MKHLKSVAIIGGPPSAAPCIKALLAENYFGKITAYEQQPAPGAGVWYNYGLTRPTNLPCTDRMPAEPHKKDGSIIYNPMYRDLNTPHMLMAYKDFPFPEGVDLFPKR  
>YALIO025102g  
MKDLPTTIVTLQSQYPPGAQIQLRVHIGDHSKPLRQQCLEKALAIQRQSTRQVEAYERILKR  
>YALIO026554g  
MKFSVLLLAGAVLALPTATVVNTEAIDYSTITDLPGSEPTDVSLEKRADPFKR  
>YALIOE00110g  
MKIDIALVLLATAVSAAPPTFSKIQSPDAIASAAATAKSNHWKDEGYETVKGQVFDKYYQIWLENTNYDKAFGQADLKALTDGILLTNYWSLTHPSQPNYIAAVSGDYFGDMTDSFKR  
>YALIO04785g  
MKHPYHIVSLSPDGKTMVAAAARQVDRICLETGECKSTVVEESATRTTKTGDKPAAAKRTKYLPTSVSHVRSLSLQSGGKYLATTDETKSLVLLDPTTLEVIKKVQYPKR  
>YALIOE05423g  
MKLLNIFTIATVASAMSLRRQEPISHVEPADTIFHIDPDLQPHSSFSLSVKVPIPEGGAHYNRLQQRWYVHADPNARRDAESFFKKR  
>YALIOE05819g  
MKFSIVLPLAGALLAAGPIQHPMEGCVPRSVFDLGDCLDCNDFLVSDDPEQCLQCLTVYELDNKNTVLSTLEEQLPMQPEVPCVGSAPMITPGAIPVLAKR  
>YALIOE09757g  
MKITRPNLHNLHNSPNPQSCGGAALGRSLNSQSSLPSTMDQTIHTPHGYLSRNNISSTVATSEDEGMYSDLSNPLTSISNQFWPEHDDAFLMHYIKVVDNPTIAPFVGHTPPSGIIHRVAKDCAKR  
>YALIOE16533g  
MKFSTIALAAVACLVSAAAPAVGTGSHGQPISEIAIVGGLQGTENEIVFFNDDESQKGIADAKKAQEAQMDPQDPDEVAAGNAKREASPEAWRWFVLPYGEPNWKRADAMPADMKEKR  
>YALIOE21229g  
MKLYYIGILRPNKPEVQLASASDLSYSWFERSTVGOAFTFLAETVIPRTEPGQRQSEVENDYVGHAYLRSDGLGGVIITDQEYPPRVAYTLLNKVLEEYTKHPKAEWENAKEASSKLAMPELEDYVVKR  
>YALIOE22374g  
MKSLLLSLLAVPATAQLLLDLQGHDTTDRHQSQNHLLTKR  
>YALIOE26719g  
MKLATAFTILTAVLAAPLAAPAEAFLLLAGFAAKISASMTNETSDRPLVHFTPNKGMNDPNGLWYDEKDAKWHLYFYQNPNDTVWGTPLFWGHATSDDLTNWEDQPIAIAPKR  
>YALIOE28875g  
MKFSAIVSLFAVAVAAQKSIEQIVSEHQERFLAKR  
>YALIOE30415g  
MKFSTIALSIMAAVAIAPIAQEPAEVPVAPEAPAAPNMDDLKQLQGLQVQLNVLFTNDNEQFIVEDDGTSLRLLVDTTLIGSEGLAKREANPEAWRWFVLPYGEPNWKRADAMPADVDKRKR  
>YALIOE31064g  
MKFNHSLQFNAPFEWTSEYIAYSQKLIYSLERTALGVPNYEDENSTLIREADSNPEDFIKR  
>YALIOE32703g  
MKFSKTLALLLAVAGALAKGEDEICRVEKNSGKEICYPKVFVPTTEWQVVPDQVIPAGLHVSRMDYENGVEKAKINDPNEEVEGVAVAVGEEVPEGEVIEDLTHEENGDEGISANEKQVRAIEKAIKVKR  
>YALIOE33539g  
MKFSTLALAAMAAGFAAPATPDADCNEEKEVRDFHAHHQHKKR  
>YALIOE34133g  
MKAWHWSVTLVVIYLAIPVILYLLTRKDDRKPLSDIRKRRK  
>YALIOE34852g  
MKFLVCLLVLALASVVSANVLLLPWK  
>YALIOF05390g  
MKLTKLVALAGAAASPIQLVPREGSFLGFNYGSEKVGVLGGWVFLVLEPITPSLFEAFGNNDANVPVDEYHYTAWLGKEEAER  
>YALIOF05786g  
MKFSTVLLTAAVAAAELITLDWAETTQVISIPDATAGEQATSAVLSAAYDAIHDHLAAQTTPAPDAEKAEALDLTIQAPFFTMGLWIPDLSDFGIGTKNETISEWLHRIWEEKQGVAPNRR  
>YALIOF11803g  
MKTAAILSCLGLVASAIPDFTTSSQFPTKR  
>YALIOF12067g  
MKFSNVLLISALSAASVAAPVAPEPKNVVWPKVDPNVLPGPEKRTLEKR  
>YALIOF12969g  
MKIPNVSTPPAVPSLTKPLTVTKTKPKAKKEEPQEPQDMSLHLQLMLSTDNVQNNFSFDEFLQVPIVKDAIKTKR  
>YALIOF19030g  
MKFSATLLAAAVSVAVADPILGKGPDPNTFWFSQSPGAPAPSNPPAAPAPKPNPGGVDNPTFWFSQSSTPTAPPAPAPSTQAPAPPASSPAPPASSAAPAPQPEETCIAPAGAQEKR  
>YALIOF19646g  
MKLSLLLLSTLAVALPIAQQPASEYNQLDLINNQDLEKR  
>YALIOF20966g  
MKTSLQLRRNLPGAAAVWPPDFGKVFYGLSVAIGAATKAVRVSTFTAFSKFMADEDNRRTAPRVVVTAGLLTGITESFIVVPFESIKTMINRVQVAAGKPVAVPPEIKLEKPEPETAAGKPVETAPVTIKDKKR  
>YALIOF31251g  
MKLPILASLSGLAGLFTAHAFGAGNIPSDSSIEGKFRHGDIEDVLAGIVISHRGGKARKFDGLDIKR

**(B)**

SignalP output data

##gff-version 2

##sequence-name source feature start end score N/A ?

##	-----	-----	-----	-----	-----	-----	-----	-----	-----
YALIOA00154g	SignalP-4.1	SIGNAL	1	15	0.639	.	.	YES	
YALIOA08360g	SignalP-4.1	SIGNAL	1	15	0.513	.	.	YES	
YALIOA09262g	SignalP-4.1	SIGNAL	1	15	0.617	.	.	YES	
YALIOA10208g	SignalP-4.1	SIGNAL	1	17	0.780	.	.	YES	
YALIOA11198g	SignalP-4.1	SIGNAL	1	17	0.586	.	.	YES	
YALIOA20350g	SignalP-4.1	SIGNAL	1	17	0.571	.	.	YES	
YALIOB05104g	SignalP-4.1	SIGNAL	1	18	0.685	.	.	YES	
YALIOB08778g	SignalP-4.1	SIGNAL	1	18	0.618	.	.	YES	
YALIOB15510g	SignalP-4.1	SIGNAL	1	17	0.770	.	.	YES	
YALIOB16500g	SignalP-4.1	SIGNAL	1	27	0.572	.	.	YES	
YALIOB22880g	SignalP-4.1	SIGNAL	1	15	0.767	.	.	YES	
YALIOC15851g	SignalP-4.1	SIGNAL	1	15	0.451	.	.	YES	
YALIOD00759g	SignalP-4.1	SIGNAL	1	19	0.539	.	.	YES	
YALIOD02024g	SignalP-4.1	SIGNAL	1	17	0.702	.	.	YES	
YALIOD02981g	SignalP-4.1	SIGNAL	1	15	0.818	.	.	YES	
YALIOD04851g	SignalP-4.1	SIGNAL	1	18	0.555	.	.	YES	
YALIOD06039g	SignalP-4.1	SIGNAL	1	18	0.820	.	.	YES	
YALIOD06149g	SignalP-4.1	SIGNAL	1	16	0.822	.	.	YES	
YALIOD08206g	SignalP-4.1	SIGNAL	1	17	0.582	.	.	YES	
YALIOD20680g	SignalP-4.1	SIGNAL	1	16	0.600	.	.	YES	
YALIOD26554g	SignalP-4.1	SIGNAL	1	23	0.623	.	.	YES	
YALIOE00110g	SignalP-4.1	SIGNAL	1	17	0.811	.	.	YES	
YALIOE05423g	SignalP-4.1	SIGNAL	1	15	0.490	.	.	YES	
YALIOE05819g	SignalP-4.1	SIGNAL	1	17	0.689	.	.	YES	
YALIOE16533g	SignalP-4.1	SIGNAL	1	20	0.771	.	.	YES	
YALIOE22374g	SignalP-4.1	SIGNAL	1	15	0.755	.	.	YES	
YALIOE26719g	SignalP-4.1	SIGNAL	1	24	0.588	.	.	YES	
YALIOE28875g	SignalP-4.1	SIGNAL	1	16	0.540	.	.	YES	
YALIOE30415g	SignalP-4.1	SIGNAL	1	17	0.736	.	.	YES	
YALIOE32703g	SignalP-4.1	SIGNAL	1	17	0.759	.	.	YES	
YALIOE33539g	SignalP-4.1	SIGNAL	1	19	0.622	.	.	YES	
YALIOF05390g	SignalP-4.1	SIGNAL	1	15	0.570	.	.	YES	
YALIOF05786g	SignalP-4.1	SIGNAL	1	23	0.623	.	.	YES	
YALIOF11803g	SignalP-4.1	SIGNAL	1	15	0.556	.	.	YES	
YALIOF12067g	SignalP-4.1	SIGNAL	1	18	0.697	.	.	YES	
YALIOF19030g	SignalP-4.1	SIGNAL	1	17	0.748	.	.	YES	
YALIOF19646g	SignalP-4.1	SIGNAL	1	14	0.674	.	.	YES	
YALIOF31251g	SignalP-4.1	SIGNAL	1	21	0.641	.	.	YES	

(c)

Multiple Sequence Alignment ClustalW (MEGA 7.0 Package)

Gene Name	MSA
YALI0A00154g	--MKLSNIFALATVALAAPQP-----
YALI0A08360g	--MKPSVIALAVACMAAP-----
YALI0A09262g	--MKPSVILSLAVACLAAPMA-----
YALI0A10208g	--MKLLSILLWISIGLCAAIP-----
YALI0A11198g	--MKFSATAIA-AAASIVVADP-----
YALI0A20350g	--MKLSTILFTACATLAAALPSP-----
YALI0B05104g	--MKLSTIAISAILGSVALAAP-----
YALI0B08778g	--MKVAHIQLAAMVATALAAVLA----
YALI0B15510g	--MKLSTLVLA-ALTAVAVA-----
YALI0B16500g	--MKFNPVISAVAMAALQA-----
YALI0B22880g	--MKFTLLLAGLSLVAAAPADADASA--
YALI0C15851g	--MKSATIGYMAVAVAAA-----
YALI0D00759g	--MKTSHITVA-LLAAVATA-----
YALI0D02024g	--MKLTT-TLLALSSLSA-----
YALI0D02981g	--MKLATSLLFTAVLAAPLAAPAP---
YALI0D04851g	--MKFSTLSVAAALAASASA-----
YALI0D06039g	--MKFSAVSIAAALASLVAAA-----
YALI0D06149g	--MKLSIVLVALAAVSSAQP-----
YALI0D08206g	--MKFSATILT-ALAASAVA-----
YALI0D20680g	--MKFSTALLALAAVATA-----
YALI0D26554g	--MKFSVLLLAGAVLALPTAT-----
YALI0E00110g	MKIDIALVLLATAVSAAPP-----
YALI0E05423g	--MKLLNIFTIATVASA-----
YALI0E05819g	--MKFSIVLPLAGALLAAA-----
YALI0E16533g	--MKFSTIALA-AVACLVSAAAPAAP---
YALI0E22374g	--MK-SLLS----LLAVPATA-----
YALI0E26719g	--MKLATAFTILTAVLAAPLAAPAEA--
YALI0E28875g	--MKFSIVSL-FAVAVA-----
YALI0E30415g	--MKFSTIALS-IMAAVAIAPIA----
YALI0E32703g	--MKFSKTLAL-ALVAGALA-----
YALI0E33539g	--MKFSTLALA-AMAAIGFAAPATPDA-
YALI0F05390g	--MKLTKLVALAG-AALASP-----
YALI0F05786g	--MKFSTVLLTAAVAAA-----
YALI0F11803g	--MKTAAILSC-LGLVVASAIP-----
YALI0F12067g	--MKFSNLLISALSAASVAAPVAEPEP
YALI0F19030g	--MKFSATLLA-AAVSVAVADP-----
YALI0F19646g	--MKLSLLLLS---TLAVALPIA----
YALI0F31251g	--MKLPILASLSGLAGLFTTAHA-----