

## Additional File 14. Mating pheromones

A potential candidate for MF $\alpha$  was identified in *A. adenivorans* as the product of ARAD1C26928g, which contains eleven repeats of a 13 amino-acid peptide (IPKWEQM/LPGQPID) matching the sequence of other yeast MF $\alpha$  (Fig. S13A). Consensus processing sites for the signal peptidase and for Kex2p are present, but not for Ste13 since there are no X-Ala/X-Pro repeats. This suggests that MF $\alpha$  processing in *A. adenivorans* is atypical: either there is only Kex2-like processing and 19 amino acid pheromones would be produced, partially divergent at their N-terminus, or another proteolytic machinery is involved in N-terminal trimming (Fig. S13B). We note that genes encoding Kex2, Kex1 and Ste13 genes exist in the sequenced strain.

The genes encoding the precursors of the other mating pheromone MF $\alpha$  are very short sized and poorly conserved, making them difficult to identify in genomes distant from *S. cerevisiae* without synteny data <sup>22</sup>. A possible candidate encoded by ARAD1D33583g was detected after a search for short proteins ending in CAAX required for prenylation and carboxylmethylation of Mfap (Fig. S13C). The 44 amino acid gene product is poorly conserved outside from its C-terminal CVIA motif, thus preventing prediction of the N-terminal processing sites <sup>23</sup>. All the machinery necessary for maturation and export of MF $\alpha$  is well conserved.

### Figure S14A Precursors of MF $\alpha$ in hemiascomycetous yeasts.

Over lining indicates: green= signal peptide; grey= prodomain, blue= dibasic aminoacids (Kex2 site); yellow= Xala/Xpro repeats (Ste13 processing); red= mating pheromone alpha. Notice absence of Xala/Xpro repeats in *A. adenivorans* pheromone precursor.

### Figure S14B Alignment of alpha pheromones of representative species of hemiascomycetes.

Alignments were generated using Multalin and manually adjusted.

### Figure S14C Alignment of *A. adenivorans* putative precursor of pheromone $\alpha$ (Mfap).

Sequences are derived from: *S. cerevisiae* YDR461W (Sace\_MFA1), *S. cerevisiae* YNL145W (Sace\_MFA2), *C. glabrata* CAGLOC01919g (Cagl\_MFA2), *Naumovozyma (Saccharomyces) castellii* Scas564.10d1 (Nacas\_MFA), *Eremothecium (Ashbya) gossypii* ABL196Cp (Ergo\_MFA2), *C. albicans* orf19.2164.1 (Caal\_MFA1), *Y. lipolytica* YALI0C17028g (Yali\_MFA2), and *A. adenivorans* ARAD1D33583g. Note that loss of the conserved DN landmark prevents definition of the N-terminal processing site. Sequences were aligned using ClustalX2.

>YPL187wp *S. cerevisiae*  
 MRFPISFTAVLFAASSALAAPVNTTEDETAQIPAEAVIGYLDLEGDFDVALPFSNSTNGLLFINTSIAAKEEGVSLDKREAEAWH  
 WLQLKPGQPMYKREAEAEAWHWLQLKPGQPMYKREADAEAWHWLQLKPGQPMYKREADAEAWHWLQLKPGQPMY

>DEHA2F19580p *D. hansenii*  
 MKFSIFALTTLVSTISLAAAAPTDPDSTETGKYVVDYLVPDEAINNKVEITDDQQLVVEESGKKYVLIVNATLAESVISKAGIDIE  
 GLEAAFAKSDDTASVSKRDANADAEAKFHWMTYRFFQPnlRKREANADAKFHWMTYRFFQPnlKKREANADAEAKFHWMTYR  
 FFQPnlKKREANAANADAKFHWMTYRFFQPnlKKREANADAEAKFHWMTYRFFQPnl

>PAS\_chr2-1\_0883p *K. pastoris*  
 MKSLILNIISVTLAITSTAASAPVESIFANQPDSSTLDTNDGVGVMSTIKEEDFGKHFVENQILDEAVIMSLKLRKGVNLFLLDDI  
 GLATELIGNKIAQIEAIDLSERLAQSWTNIKNRFLGKREAEAEAEAEAEFRWRNNEKNQPFgKREAEAEAEAEAEAEAEAEAEFRW  
 RNNEKNQPFgKREAEAEAEAEAEAEAEAEAEFRWRNNEKNQPFgKREAEAEAEAEAEAEAEAEAEFRWRNNEKNQPFgKREADAEAEAEAEAEFR  
 WRNNEKNQPFgKREAEAEAEAEAEAEAEAEAEFRWRNNEKNQPFgKREAEAEAEAEAEAEAEAEAEFRWRNNEKNQPFgKREAEAEAEAEAEFRWRN  
 NEKNQPFgKREADAEAEAEAEAEAEAEAEAEFRWRNNEKNQPFgKREASIDTGTDDGAYWSWRKNSVLERQ

>ARAD1C26928p *A. adenivorans*  
 MHLKASFLLLIATAIVSTAAPISGDQNVTEPDVDHQIPECQPRSPPIDRRDEKWEPIPKWEQMPGQPIDKRDEKWKPIPK  
 WEQLPGQPIDRRDVKWEPNIPKWEQMPGQPIDKRDEKWPDIKWEQMPGQPIDKRDVKWKPIPKWEQMPGQPIDRRDV  
 KWTPDIKWEQMPGQPIDRRDVKWTPDIKWEQMPGQPIDKRDVKWEPNIPKWEQMPGQPIDKRDVKWKPIPKWEQMP  
 GQPIDKRDVWRWKPIPKWEQMPGQPIDK

>YALI0E16533p *Y. lipolytica*  
 MKFSTIALAAVACLVAAPAAPVGTGSHGPQSIPEEAIVGGLQGTENEIVFFNDDESgKQGIADAKKAQEAQFMDPQPDSEV  
 AAGNAKREASPEAWRWFwLPGYGEpNWKRdAMPADMDKEKREANPEAWRWFwLPGYGEpNWKRdAMPADMDKEKREAN  
 PEAWRWFwLPGYGEpNWKRdAMPADMDKEKREANPEAWRWFwLPGYGEpNW

\*

Sace	:	-WH	LQLKP-GQP-MY	:	13
Klla	:	-WS	ITLRP-GQP-IF	:	13
Klth	:	-WR	LSLSR-GQP-MY	:	13
Sakl_1	:	-WH	LSFSK-GEP-MY	:	13
Zyro	:	--H	IELDP-GQP-MF	:	12
Arad	:	IPK	EQMPG--QP-ID	:	13
Pipa	:	-FR	RNNEK-NQP-FG	:	13
Sakl_2	:	-WH	LSFSF-GEP-LY	:	13
Cagl	:	-WH	VKIRK-GQG-LF	:	13
Piso_1	:	KFH	FKYNN-FDP-IT	:	14
Piso_2	:	KFH	FTYSN-FDP-IT	:	14
Deha	:	KFH	MTYRF-FQPnlR	:	15
Yali	:	-WR	FWLPGYGEp-NW	:	14

Figure S14A

Figure S14B

# Figure S14C

		*		20		*		40		*		60		*					
Sace_MFA1	:	-----MQP	--	STATAAPKEKTS	SEKKDN	----	YIIKG	--	VFWD	DP	-----	-----	-----	AC	VIA	: 36			
Sace_MFA2	:	-----MQP	IT	ASTQATQKDKS	SEKKDN	----	YIIKG	--	LFWD	DP	-----	-----	-----	AC	VIA	: 38			
Cagl_MFA2	:	-----MQP	---	TIEATQKDN	TQEKRDN	----	YIVKG	--	FFW	SP	-----	-----	-----	DC	VIA	: 34			
Naca_MFA	:	-----MQP	---	TTQATHKDN	SAEKQDN	----	YIVKG	--	LFWD	DP	-----	-----	-----	AC	VIA	: 34			
Ergo_MFA2	:	-----MQL	----	TNNTNKDE	STENKDN	----	WIAKG	--	YMW	TP	-----	-----	-----	QC	VIV	: 33			
Caal_MFA1	:	-----MAA	QQ	SKKTGSGQTKDK	DAAKNN	AVRS	VSTG	--	NCC	ST	-----	-----	-----	CS	VV	: 42			
ARAD1D33583g	:	-----MYE	EP	VAIEFAVSR	PQSGVRV	PPR	---	VVM	IAEM	NLQ	GP	-----	-----	SI	SC	VIA	: 44		
Yali_MFA2	:	MTSWFDILQL	AK	QHKSHQSLAVD	NNKNDK	TKAP	LLPL	QDN	MAHYD	PS	WT	HE	QKVQ	EV	KRR	QSQS	FRL	CVIA	: 72
<b>Consensus</b>			<b>qp</b>		<b>kd</b>		<b>dn</b>		<b>g</b>		<b>w P</b>							<b>CVI a</b>	