

Table S1. List and characteristics of the detected pseudogenes in the 8 studied yeast genomes.

The table is in Lafontaine_SupTab.xls

Pseudogenes are named according to their position along the chromosome (from left to right end, see text for nomenclature). Chromosomes, coordinates and orientation of pseudogenes are given in columns 2-5. The closest functional homolog among the 8 species studied (bestmatch) and classification of its product into protein families are given in column 6-7. Gene nomenclature refers to (16), family classification refers to (58).

Columns 8-10: characteristics of pseudogene by comparison to its bestmatch.

R.L: relative length of the pseudogene with respect to its bestmatch length.

D.A: amino acid divergence between the pseudogene and its bestmatch (see METHODS).

D.N: nucleotide divergence between the pseudogene and its bestmatch (see METHODS).

Columns 11-12:

Stop: number of in-frame stop codons according to FASTY

FS: number of frameshifts according to FASTY

Column 13: pseudogene(s) of the same family, with the same pattern of disabling mutations. Pseudogenes with an asterisk are part of a larger duplicated region including other genes.

Note that a few pseudogenes were already mentioned in the original annotation of *C. glabrata*, *D. hansenii*, *K. lactis* and *Y. lipolytica* (13), and the result of the present work served for the recent updates of their annotation, as well as for the annotation of the newly sequenced genomes of *Z. rouxii*, *S. kluyveri* and *K. thermotolerans* (24).

Table S2. List of gene families (or singletons) with detected pseudogenes.

The table is in Lafontaine_SupTab.xls

The phyletic patterns of each family indicate the presence (s, c, r, l, t, k, d, y) or absence (-) of the pseudogene members (column 2) and of the active gene members (column 4) in each species. The one-letter abbreviations are as follows : 's' for *S. cerevisiae*, 'c' for *C. glabrata*, 'r' for *Z. rouxii*, 'l' for *K. lactis*, 't' for *K. thermotolerans*, 'k' for *S. kluyveri*, 'd' for *D. hansenii* and 'y' for *Y. lipolytica*.

The general annotation of the *S. cerevisiae* gene members is given in column 7, according to the annotation of the gene indicated in column 6.

The number of pseudogenes is not correlated to the number of their functional paralogs in the same family (from 1 to at least 20 active paralogs can be found per pseudogene). In addition, 61 pseudogenes have been identified by similarity with genes from another species, because they have no paralog and are unique in their genome.

Table S3. **Pseudogenes with evidence of transcription in *S. cerevisiae***

Name	stops	frameshifts	R.L	ref
SACE0Ap1	0	1	1	1
SACE0Ap2	1	0	0.9	1
SACE0Ap7	1	4	0.11*	1
SACE0Ap13	0	1	1	1
SACE0Bp1	5	15	1	1, 2, 3
SACE0Cp2	2	0	0.06*	1
SACE0Cp3	10	7	1	1
SACE0Dp6	1	0	0.82	1, 2
SACE0Dp7	5	11	0.57*	1
SACE0Hp1	1	8	0.86	2
SACE0Lp4	0	0	0.55*	3
SACE0Pp4	1	2	0.77	2

The number of disabling mutations within each pseudogene is given in columns 1 and 2. The relative length (R.L) of the pseudogene with respect to its closest functional homolog is indicated in column 3. The asterisk indicates that the pseudogene is truncated.

Last column indicates the data set where evidence of transcription is found (see text for details).

1: (46), 2: (37), 3: (43).

Figure S1. Alignments of some pseudogenes with their bestmatches

a)

ZYRO0G16588g 1 MSNFFRDSSLGFKPRSNIFSKLRVKDNEASNENSNTQEGDSSLEDFLNSNKEDIISPAT
 KLTH0Gp3 S-NFLRDGGTGKVRSNIFSKFRSK--ESVQESDFS AENDANSFESSMTQEKP--SGHLD

ZYRO0G16588g 61 TIINSDDGGKFGHSLQLRSSTPNPST-TKAEKPGRGLLEDDLEITEVRDVPISKEESGTS
 KLTH0Gp3 SRISKDDTSVTTTHSLTESTPKGKRFRSRDPKAFRSGKDEEELEITEVRQVLHQDLRKLIN

ZYRO0G16588g 120 SGT PNSNQINAKTVNLSLPVLPASQQRSERECIDASSNDVLLDAFTNTQKICSNLKQEQ
 KLTH0Gp3 S-----DNSVNEKLATLKT-----TDVSKDTSINDVLLDAFTNTQKICSSLRVE

ZYRO0G16588g 180 LHVQTE NSK LKTVQSYQSDKERIMDRFAEYKLLNSFSEKSKLFEQKKKEDTQKKEF
 KLTH0Gp3 LORQKSENREQAQTIMSYKTELGTIEDRIARYROLLSVIEEKSNHLOKQKNDSEGLKEL

ZYRO0G16588g 240 KENYDKLLKVVESYKDDIHGLKNNLSQLRFLKKDSDELAKKVKELIYLLKRELDSCSGQL
 KLTH0Gp3 QRSVAELQSRIRDYGGKDCSTIKKMLEETRLMOKEFASVSKKAKEIYLLKKELESQGL

ZYRO0G16588g 300 SEFKLKNSSLVQEFGKIRGELMVNLTQNLISQGSQILEKIGILESNFIGAYHVDWNNFSE
 KLTH0Gp3 SEFKLKNRELQELRALRKQTESLAENAIKMLEDEVKKEKELG-----

ZYRO0G16588g 360 SSTSKLTECIGTRTNISSEFKTSFGENNRSLVDYYQKSFTEFMGRFNQIDGYVDKLSTS
 KLTH0Gp3 --INKLIKRDNQI-LSLSAEDKLELGSMLQTML-----MVEIK-----KYSPTQKH

ZYRO0G16588g 420 IKEGTELOS AKFDAHTS SLFSELRRGKDESWELIKKWSNENENRLOIVQDKINSNLNSF
 KLTH0Gp3 IPSFTKVGSTDTTLQLAIIYSENRLINDVSQVFEQGLSRVE---EISRDITIKSDSYTFT

ZYRO0G16588g 480 QLIKTTSE-EYTAISGDIKDRLOKVOSESFEILQKHLOKLPGFFDEVREGHIKSKKWMK
 KLTH0Gp3 QQMTILGELKDSALS--LSEKYQNEIDINSFN--EGVVQPIISGLGQAMKQVLS----ITS

ZYRO0G16588g 539 LAQELSSYKESLTNSREYEANISGLQSQISSLOLQKSQALSSIGTKEAQYEDLCNSLISK
 KLTH0Gp3 IAERLNQYKIQVDTGLKYEQKLEDIYEKLOILSMOKSEAVALLKSKLEIEELSDOVFTK

ZYRO0G16588g 599 ETELARFKEIEKELHGKIENISSEVEQOKNKWLRNLNENITLKANSENKLVVONEILKAF
 KLTH0Gp3 TNALNELTTKGDFAFNDSGRIRKALQAKTQELEKIKQELNLLITANSKSKLMAQAEIAKIV

ZYRO0G16588g 659 QSENNITLKORTAQLEDVRQQYEKENSTRLDKIQRI NEQLQKLNVEVVLKARELELEEN
 KLTH0Gp3 MQRDLSLKQDCKQLQASKFDIEEIECVNIKEKAKRATEQVQVNLNVEVVLKARELELEEN

ZYRO0G16588g 719 RNLKRLIEDNKMEFEDTTDDYRRLKQKRVIVLESKQDIVSEKLELQDQMEKMQAVVTGLK
 KLTH0Gp3 RKVNNVLEQLHFDKAKSSDQLREYKQHISVLENDKKNCASETLECDRIIFLEQQLENAQ

ZYRO0G16588g 779 QKSSLOKNEHDYSKQLEAQAEMRQIPKQPKRPENQNGRKEPAKRRIREHLAPMFKPO
 KLTH0Gp3 KETQGLKGRQNLHMKRKYSAV-----EGAQRR-----E

ZYRO0G16588g 839 APSLPPPLPSTQRQLKSNKDKDEEAEFDLSSSLNDDLELITNPSPIKPVIAKRGDTKIR
 KLTH0Gp3 SPSA-----NHSNDGFDLSSSLNDDLEMTNPSAPP-----

ZYRO0G16588g 899 PPMGSRKLLLLLDEQDSTPSKHRWKKRRI
 KLTH0Gp3 -----K

ZYRO0G16588g 1 MSNFFRDSLGFKPRSNIFSKLRV--KDNEAS-----NENSNTQEGDSSI--ED
SAKL0Gp7 R-----DNALTFKPRSNIFSKLRVSNKDNSLSSGALRVIHSDDNMNMNKTEDSGSSIMEN

ZYRO0G16588g 46 LFLNSNKEDITSPATTIINSDGGKFGHSLQLRSTPNPSTTKAEK-PGRGLL--EEDDL
SAKL0Gp7 VFLHSGNSGKHSSSTLFDDDTTSDDHKPNHINESTPKALKNYCIKGPNGGIHDEEDEL

ZYRO0G16588g 103 ITEVRDVPISKEESGTS SGT P NS N I N A K T V N L S L P V L P A S Q Q R S E R E C D I D A S S N D V L
SAKL0Gp7 ITEVREI-----TQDDTSFDMELADQ---KASTLIESIKPPOHSNT-----SSNDVL

ZYRO0G16588g 163 LEAFTNTQKICSNLQELHKVQTEENSKLKTQVQSYQSDKEKIMDRFABYKKLNSFSEKS
SAKL0Gp7 LEAFTNTQKICSNLQELQKQNKQNAQKDKILTYQSEIAKINELLSYKKFLSKIIEKS

ZYRO0G16588g 223 KLLFEQKKKEDTQLKEFKENYDKLKKVESYKDDIHGLKNNLSQLRFLKDKSDTEAKKV
SAKL0Gp7 RWLLQKKSDDTKIHDLKVNYDGIIEKLEKEYLKE SMELKVTIEQLRDLKQSTDAEISKKD

ZYRO0G16588g 283 KEIEYLKRELDSCSGQLSEEKLNKNSLIVQEFGKIRGELMVNLTQNLQOGSQITILEKIGIL
SAKL0Gp7 KEIEYLKRELNNCSGQLSEEKLNKNSDMLQELNNTREESMTKMELLFSTQESKIREGFESI

ZYRO0G16588g 343 ESNFVIGAYHVDWNNFSESSISKLETCIGTLRTNISSEFKTSFGENNRSLVD--YYQKSFTE
SAKL0Gp7 IKGYSSILKTDTPALVETEIMKLS-----TVFKETSITYLEKYFFCNLIP

ZYRO0G16588g 402 FMGRFNQIDGYIDKLSSTSIKEGTELOSAKFAHTSSIFSELRRGKDESWELIKKWSNENE
SAKL0Gp7 LCVYFLHFVHFIGKLTHTSTRS--TESLAQIIDTHEKNVLSNGKEIKLHSDHKLEKVAKE--

ZYRO0G16588g 462 NRQIVQDKINSSNLNSFOLIKTISEEYTAISGDIKDRLOKYOSESF-EILOKHLQKLPG
SAKL0Gp7 --IILIKDQ-----OLGLKSAIEY-----LOKVTSSIBENVOTHSEVSMN

ZYRO0G16588g 521 FFDEVREGHIKSKKWEKLAQELSSYKESLNSREYBANISGLOSQISSIQLOKSOAASS
SAKL0Gp7 LMSKLETS-----MSTISKELIQYKSELDONKEYESKINDFORQVHSIALQKSEVVAL

ZYRO0G16588g 581 LGTKEAQYEDLQNSLISKETELAREKEIEKELHGTENISSEVEQQKQKWLRLNEENITL
SAKL0Gp7 MKTKDAEIEDLNNQLFNKTKLTEETEAATEALKDEITSLKEDARSKDLLNNVQNELKAI

ZYRO0G16588g 641 KANSNKLVVQNELLKAFQSENNTLKORTAQLEDVROQYKENSSTRIDKIQORINEQLOKL
SAKL0Gp7 KNNFDSKFSAQNELVKMVTQERDELKKLRSRETDSKHRVEKEGRVYVDKFORVNEQLOKL

ZYRO0G16588g 701 NVENVOLKAHELELEENRNLKRLIEDNKMEFEDTDDYKRLKQRVIVLESKDQDIVSEK
SAKL0Gp7 NVEIVOLKARELELEENRKLNETEQFDQAFRANDEFQEVKEKILSLESEKQRFITDR

ZYRO0G16588g 761 LELQDQMEKMQAVVTGLKQKVSSQKNEHDYSKOLEAQAAEMRQIPKQPKRPENQGNCRK
SAKL0Gp7 LECODKIEELELO---LKEATKNIQ-----HLRDOA-----HK

ZYRO0G16588g 821 EPAKRRIREHLAPMFKPQAPSLPPPLPSTORQLKSNKDKDEEADEFDLSSSLNDDLELT
SAKL0Gp7 RPSKK-----PPPVETSVKSPAS----KTGNVDEEHLSSSLNDDLELT

ZYRO0G16588g 881 NPSPV--KPVIAKRGDTKIRPPMGSRKKLLLLDEQDSTPS---KHRWKKRRV
SAKL0Gp7 NPSPVQIKPLRAKRGSSVVRTSMATKKLLLLSDDDDITNKSrvKNTKRRV

b)

CAGL0J03168g	308	MFTEVRKEYDGDKKFDV LNNLDWVFNE LLVVKDYVSKYSP PYWNIFDKYYQFYDELHIL
CAGL0Dp2		LFTEVRKEDDGDKYFEL LNNLDWVINK LLVVEDYVSKYSSPYWRIFEKYYLFYCVFHV
CAGL0J03168g	368	INELVDAEPETIIITLDIIDFDKTFONTLVKDFGFKRKETKTVIGDTQKETLFKDYLNLIIV
CAGL0Dp2		VNELVDGEAETIIITLDTIDLEKTFONTLVKDCGFKWKETXTVI-DPQKRILFKGYLNLIIV
CAGL0J03168g	428	IKMTEWLCNLOKTEFKTFKDRSIPPOSDAENLLLLLEGTTCFQMFSSQOVEVAAGSSQAKI
CAGL0Dp2		NKKITTRFGKLNMEF!KFKDRSILPOANGKNLLLLXGTKNCSRMSXQVKVAAYSSQTKI
CAGL0J03168g	488	LVGVIEKFSSELTENRLSNWISVDEEVRRMLKYNELYDLDPNAT--APENEVPGGLLEYV
CAGL0Dp2		LLVFFR!FSELTENRLSNWISVIDEXVRRMLVYIEM!NLKALILMLSPENKVPGGIPDYV
CAGL0J03168g	546	IALANDQ
CAGL0Dp2		IVLANDK
YOR028C	207	RRVSVSLSEQVNEGERYNNDGQLIGKGTGKPLRNTKRAAQNRSAQKAFRQRREKYIKNLE
CAGL0Hp3		RRSSIFSDMPVYNNGERYNANGELIGKGTGKPLRNTKRAAQNRSAQKAFRQRREKYIKALE
YOR028C	267	EKSKLFDGLMKENSELKKMIESLKSK
CAGL0Hp3		EKAKQYDSLKKNMELKNLVEVLKNQ
SAKL0F08734g	14	HKPKSGTTHKIREQLNFSDDKKWKFSSRRLELIDKFGLSERKASEQDDNIRQIATILR
CAGL0Mp5		NKHSKNGTTHKIREQLGFIDEKKWKRFSRRLELIDKFOLSQFKASEQDLNIRQIANILR
SAKL0F08734g	74	TEFGYSVSSSAEFEKLVTAAVQSVRRNRKRSTK
CAGL0Mp5		TEFGYPVICSMEFEKLVTAALQSVRRNRKRSKK

In-frame stop-codon (*) and frameshift mutation (!) are boxed. Identical residues are shaded in black and similar residues are shaded in grey. a) Alignments of “full-size” pseudogenes *KLTH0Bp3* and *SAKL0Gp13* with their bestmatch *ZYROOG16588g* (943 aa). Both pseudogenes have only one conserved frameshift mutation (it lies within a poorly aligned region of 30 bp). b) Alignments of the truncated pseudogenes *CAGL0Dp2* with *CAGL0J03168g* (827 aa), *CAGL0Hp3* with *YOR028C* (300 aa), *CAGL0Mp5* with *SAKL0F08734g* (638 aa). *CAGL0Dp2* and *CAGL0Hp3* are truncated from their 5'-end. *CAGL0Mp5* is truncated from its 3'-end. *CAGL0Hp3* and *CAGL0Mp6* have intact coding sequences.

Figure S1. Alignments of some pseudogenes with their bestmatches

a)

ZYRO0G16588g 1 MSNFFRDSSLGFKPRSNIFSKLRVKDNEASNENSNTQEGDSSLEDFLNSNKEDIISPAT
 KLTH0Gp3 S-NFLRDGGTGKVRSNIFSKFRSK--ESVQESDFSAENDANSFESSMTQKEP--SGHLD

ZYRO0G16588g 61 TIINSDDGGKFGHSLQLRSSTPNPST-TKAEKPGRGLLEDDLEITEVRDVPISKEESGTS
 KLTH0Gp3 SRISKDDTSVTTTHSLTESTPKGKRF SRDPKAFRSGKDEEELEITEVRQVLHQDLRKLIN

ZYRO0G16588g 120 SGT PNSNQINAKTVNLSLPVLPASQQRSERECIDASSNDVLLDAFTNTQKICSNLKQEQ
 KLTH0Gp3 S-----DNSVNEKLATLKT-----TDVSKDTSINDVLLDAFTNTQKICSSLRVE

ZYRO0G16588g 180 LHVQTE NSK LKTVQSYQSDKERIMDRFAEYKLLNSFSEKSKLFEQKKKEDTQKKEF
 KLTH0Gp3 LORQKSENREQAQTIMSYKTELGTIEDRIARYROLLSVIEEKSNHLOKQKNDSEGLKEL

ZYRO0G16588g 240 KENYDKLLKLVESYKDDIHGLKNNLSQLRFLKKDSSTELAKKVKELIYLLKRELDSCSGQL
 KLTH0Gp3 QRSVAELQSRIRDYGGKDCSTIKKMLEETRLMOKEFASVSKKAKEIYLLKRELNESGQL

ZYRO0G16588g 300 SEFKLKNSSLVQEFGKIRGELMVNLTQNLISQGSQILEKIGILESNFIGAYHVDWNNFSE
 KLTH0Gp3 SEFKLKNRELQELRALRKQTESLAENAIKMLEDEVKKEKELV-----

ZYRO0G16588g 360 SSTSKLTECIGTRTNISSEFKTSFGENNRSLVDYYQKSFTEFMGRFNQIDGYVDKLSTS
 KLTH0Gp3 --INKLIKRDNQI-LSLSAEDKLELGSMLQTML-----MVEIK-----KYSPTQKH

ZYRO0G16588g 420 IKEGTELOS AKFDAHTS SLFSELRRGKDESWELIKKWSNENENRLOIVQDKINSNLNSF
 KLTH0Gp3 IPSFTKVGSTDTTLQLAIIYSENRLINDVSQVFEQGLSRVE---EISRDITIKSDSYTFT

ZYRO0G16588g 480 QLIKTTSE-EYTAISGDIKDRLOKVOSESFEILQKHLOKLPGFDFEVREGHIKSKKWMK
 KLTH0Gp3 QQMTILGELKDSALS--LSEKYQNEIDINSFN--EGVVQPIISGLGQAMKQVLS----ITS

ZYRO0G16588g 539 LAQELSSYKESLTNSREYEANISGLQSQISSLOLQKSQALSSIGTKEAQYEDLCNSLISK
 KLTH0Gp3 IAERLNQYKIQVDTGLKYEQKLEDIYEKLOILSMOKSEAVALLKSKLEIEELSDOVFTK

ZYRO0G16588g 599 ETELARFKEIEKELHGKIENISSEVEQOKNKWLRINENITLKANSENKLVVONEILKAF
 KLTH0Gp3 TNALNELTTKGDFAFNDSGRIRKALQAKTQELEKIKQELNLLITANSKSKLMAQAEIAKIV

ZYRO0G16588g 659 QSENNITLKORTAQLEDVRQQYEKENSTRLDKIQRI NEQLQKLNVEVVLKARELELEEN
 KLTH0Gp3 MQRDLSLKQDCKQLQASKFDIEEIECVNIKEKAKRATEQVQVNLNVEVVLKARELELEEN

ZYRO0G16588g 719 RNLKRLIEDNKMEFEDTTDDYRRLKQKRVIVLESKQDIVSEKLELQDQMEKMQAVVTGLK
 KLTH0Gp3 RKVNNVLEQLHFDKESDQLREYKQHISVLENDKKNCASETLECDRIIFLEQQLENAQ

ZYRO0G16588g 779 QKSSLOKNEHDYSKQLEAQAEMRQIPKQPKRPENQNGRKEPAKRRIREHLAPMFKPO
 KLTH0Gp3 KETQGLKGRQNLHMKRKYSAV-----EGAQRR-----E

ZYRO0G16588g 839 APSLPPPLPSTQRQLKSNKDKDEEADEFDLSSSNDDLELNTNPSPIKPVIAKRGDTKIR
 KLTH0Gp3 SPSA-----NHSNDGFDLSSSNDDLEMTNPSAPP-----

ZYRO0G16588g 899 PPMGSRKLLLLLDEQDSTPSKHRWKKRRI
 KLTH0Gp3 -----K

ZYRO0G16588g 1 MSNFFRDSLGFKPRSNIFSKLRV--KDNEAS-----NENSNTQEGDSSI--ED
SAKL0Gp7 R-----DNALTFKPRSNIFSKLRVSNKDNSLSSGALRVIHSDDNMNMNKTEDSGSSIMEN

ZYRO0G16588g 46 LFLNSNKEDITSPATTIINSDGGKFGHSLQLRSTPNPSTTKAEK-PGRGLL--EEDDL
SAKL0Gp7 VFLHSGNSGKHSSSTLFDDTTGSDDKPNHINESTPKALKNYCIKGPNGGIHDEEDEL

ZYRO0G16588g 103 ITEVRDVPISKEESGTS SGT P NS N Q I N A K T V N L S L P V L P A S Q Q R S E R E C D I D A S S N D V L
SAKL0Gp7 ITEVREI-----TQDDTSFDMELADQ---KASTLIESIKPPOHSNT-----SSNDVL

ZYRO0G16588g 163 LEAFTNTQKICSNLQELHKVQTEENSKLKTQVQSYQSDKEKIMDRFABYKKLNSFSEKS
SAKL0Gp7 LEAFTNTQKICSNLQELQKQNKQNAQKDKILTYQSEIAKINEKLLSYKKFLSKIEKS

ZYRO0G16588g 223 KLLFEQKKKEDTQLKEFKENYDKLKKVESYKDDIHGLKNNLSQLRFLKDKSDTEAKKV
SAKL0Gp7 RWLLQOKKSDDTKIHDLKVNYDGIIEKLEKEYLKE SMELKVTIEQLRDLKQSTDAEISKD

ZYRO0G16588g 283 KEIEYLKRELDSCSGQLSEEKLNKNSLIVQEFGKIRGELMVNLTQNLQOGSQITILEKIGIL
SAKL0Gp7 KEIEYLKRELNNCSGQLSEEKLNKNSDMLQELNNTREESMTKMELLFSTQESKIREGFESL

ZYRO0G16588g 343 ESNFVIGAYHVDWNNFSESSISKLETCIGTLRTNISSEFKTSFGENNRSLVD--YYQKSFTE
SAKL0Gp7 IKGYSSILKTDTPALVETEIMKLS-----TVFKETSKTYLEKYFFCNLIP

ZYRO0G16588g 402 FMGRFNQIDGYIDKLSSTSIKEGTELOSAKFAHTSSIFSELRRGKDESWELIKKWSNENE
SAKL0Gp7 LCVYFLHFVHF!GKLTHTSTRS--TESLAQIIDTHEKNVLSNGKEIKLHSDHKLEKVAKE--

ZYRO0G16588g 462 NRQIVQDKINSSNLNSFOLIKTISEEYTAISGDIKDRLOKYOSESF-EILOKHLQKLPG
SAKL0Gp7 --IILIKDQ-----OLGLKSAIEY-----LOKVTSSIBENVOTHSEVSMN

ZYRO0G16588g 521 FFDEVREGHIKSKKWEKLAQELSSYKESLNSREYBANISGLOSQISSILOKSOAASS
SAKL0Gp7 LMSKLETS-----MSTISKELIQYKSELDONKEYESKINDFORQVHSIALQKSEVVAL

ZYRO0G16588g 581 LGTKEAQYEDLQNSLISKETELAREKEIEKELHGTENISSEVEQQKQKWLRLNEENITL
SAKL0Gp7 MKTKDAEIEDLNNQLFNKTKLTEETEAATEALKDEITSLKEDARSKDLLNNVQNELKAI

ZYRO0G16588g 641 KANSNKLVVQNELLKAFQSENNTLKORTAQLEDVROOYEKENSTRIDKIQORINEQLOKL
SAKL0Gp7 KNNFDSKFSAQNELVKMVTQERDELKPKLSRETDSKHRVEKEGRVYVDKFORVNEQLOKL

ZYRO0G16588g 701 NVENVOLKAHELELEENRNLKRLIEDNKMEFEDTDDYKRLKQRVIVLESKDQIVSEK
SAKL0Gp7 NVEIVOLKARELELEENRKLNETEQFDQAFRANDEFQEVKEKILSLESEKQRFITDR

ZYRO0G16588g 761 LELQDQMEKMQAVVTGLKQKVSSQKNEHDYSKOLEAQAAEMRQIPKQPKRPENQGNCRK
SAKL0Gp7 LECODKIEELELO---LKEATKNIQ-----HLRDOA-----HK

ZYRO0G16588g 821 EPAKRRIREHLAPMFKPQAPSLPPPEPSTORQLKSNKDKDEEADEFDLSSSLNDDLELT
SAKL0Gp7 RPSKK-----PPPVETSVKSPAS----KTGNVDEEHLSSSLNDDLELT

ZYRO0G16588g 881 NPSPV--KPVIAKRGDTKIRPPMGSRKLLLLDEQDSTPS---KHRWKKRRV
SAKL0Gp7 NPSPVQIKPLRAKRGSSVVRTSMATKKLLLLSDDDDITNKSrvKNTKRRV

b)

CAGL0J03168g	308	MFTLEVRKEYDGDKKFDVLLNNLDWVFNELLLVVKDYVSKYSPPYWNIQFDKYYQFYDELHIL
CAGL0Dp2		LFLEVRKEDDGDKYFELLLNNLDWVINKLLVVEDYVSKYSSPYWRIFEKYYLFYCVFHV
CAGL0J03168g	368	INELVDAEPETIIITLDIITDFDKTFONTLVKDFGFKRKETKTVIGDTQKETLFKDYLNLIIV
CAGL0Dp2		VNELVDGEAETIIITLDTIDLEKTFONTLVKDCGFKWKETXTVI-DPQKRILFKGYLNLIIV
CAGL0J03168g	428	IKMTEWLCNLOKTEFKTFKDRSIPPOSDAENLLLLLEGTTCFQMFSSQOVEVAAGSSQAKI
CAGL0Dp2		NKKITTRFGKLNMEF!KFKDRSILPOANGKNLLLLXGTKNCSRMSXQVKVAAYSSQTKI
CAGL0J03168g	488	LVGVIEKFSSELLENRLSNWISVDEEVRRLMKYNELYDLDPNAT--APENEVPGGLLEYV
CAGL0Dp2		LLVFFR!FSELLENRLSNWISVIDEXVRLMVIEM!NLKALILMLSPENKVPGGIPDYV
CAGL0J03168g	546	IALLANDQ
CAGL0Dp2		IVLANDK
YOR028C	207	RRVSVSLSEQVNEGERYNNDGQLIGKGTGKPLRNTKRAAQNRSAQKAFRQRREKYIKNLE
CAGL0Hp3		RRSSIFSDMPVYNNGERYNANGELIGKGTGKPLRNTKRAAQNRSAQKAFRQRREKYIKALE
YOR028C	267	EKSKLFDGLMKENSELKKMIESLKSK
CAGL0Hp3		EKAKQYDSLKKNMELKNLVEVLKNQ
SAKL0F08734g	14	HKPKSGTTHKIREQLNFSDDKKWKFSSRRLELIDKFGLSERKASEQDDNIRQIATILR
CAGL0Mp5		NKHSKNGTTHKIREQLGFIDEKKWKRFSRRLELIDKFQLSQFKASEQDLNIRQIANILR
SAKL0F08734g	74	TEFGYSVSSSAEFEKLVTAAVQSVRRNRKRSTK
CAGL0Mp5		TEFGYPVICSMEFEKLVTAALQSVRRNRKRSKK

In-frame stop-codon (*) and frameshift mutation (!) are boxed. Identical residues are shaded in black and similar residues are shaded in grey. a) Alignments of “full-size” pseudogenes *KLTH0Bp3* and *SAKL0Gp13* with their bestmatch *ZYROOG16588g* (943 aa). Both pseudogenes have only one conserved frameshift mutation (it lies within a poorly aligned region of 30 bp). b) Alignments of the truncated pseudogenes *CAGL0Dp2* with *CAGL0J03168g* (827 aa), *CAGL0Hp3* with *YOR028C* (300 aa), *CAGL0Mp5* with *SAKL0F08734g* (638 aa). *CAGL0Dp2* and *CAGL0Hp3* are truncated from their 5'-end. *CAGL0Mp5* is truncated from its 3'-end. *CAGL0Hp3* and *CAGL0Mp6* have intact coding sequences.