

# **Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations**

Michiel Vandenbussche<sup>1</sup>, Günter Theissen<sup>2</sup>, Yves Van de Peer<sup>1</sup>  
and Tom Gerats<sup>1,3</sup>.

<sup>1</sup>Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium

<sup>2</sup>Lehrstuhl for Genetics, Friedrich Schiller University of Jena, Philosophenweg 12, D-07743 Jena, Germany

<sup>3</sup>present address: Department of Experimental Botany, University of Nijmegen, Toernooiveld 1, 6525E, Nijmegen, The Netherlands

## **Supporting Material**

### **Contents:**

**Page 2:** Methodology: Assembly of the MADS-box dataset

**Page 3:** Species Table

**Page 4-11:** C-terminal motifs classified per subfamily

## **Assembling the MADS-box sequence dataset**

To extract the MIKC type MADS-box sequence dataset from the public databases, we have used the following approach: In a first step, one representative protein sequence of each known MADS-box subfamily (according to recently published phylogenetic analyses, see for example reference 1) was chosen. The full-length protein sequence of each selected subfamily member was used to search the public databases for homologous sequences, using the BLAST program (<http://www.ncbi.nlm.nih.gov/BLAST/>). The protein sequence dataset was built using the blastp option to search the available protein databases, while the nucleotide dataset was assembled using tblastn to search the available nucleotide databases. To ensure that even distantly related MADS-box genes would be retrieved in the homology searches, the standard settings of some search parameters of the BLAST program were changed. The expectancy value was raised to 100, and the number of homologous sequences to be displayed was increased to 500. Homologies shown in the output BLAST page were inspected visually, and MADS-box sequences were retrieved by using the sequence retrieval option and saved in batches as text files in Genbank format. These text files were subsequently imported in the GCG program using the FromGenbank function (Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisc.). Because of the low stringency settings of the homology searches with the different subfamily representatives, a large proportion of the collected sequences was identified multiple times; such duplicates were removed automatically during import in GCG using the 'remove duplicates' option.

To yield a more comprehensive overview of the taxonomic distribution within subfamilies, the sequence identifiers were renamed manually as follows: they all start with a two- or three-letter code indicating the species name (see Species Table below for abbreviations), followed by the gene name as found in the database and terminating with the database Accession Number (according to either DNA, EST, Protein or Patent databases). Full-length sequences were aligned using the PILEUP function, followed by a manual alignment of the C-terminal regions using the SeqLab Editor of the GCG software package. To illustrate conservation of the C-terminal conserved motifs, we have displayed the alignments of the C-terminal motifs per subfamily. C-terminal motifs of sequences of which the sequence name terminates with 'COR' (corrected) have been identified in other reading frames than published in the database and/or beyond the proposed stopcodon. An asterisk behind a peptide motif indicates a stopcodon; motifs not terminating with an asterisk are followed by a variable number of less-conserved residues (not shown).

## Species Table

Species	code	Species	code
<i>Agapanthus praecox</i>	Ap	<i>Lolium temulentum</i>	Lte
<i>Akebia quinata</i>	Aq	<i>Lycopersicon esculentum</i>	Le
<i>Anemone nemorosa</i>	An	<i>Lycopodium annotinum</i>	La
<i>Antirrhinum majus</i>	Am	<i>Magnolia praecocissima</i>	Mp
<i>Aquilegia alpina</i>	Aa	<i>Malus domestica</i>	Md
<i>Aquilegia caerulea</i>	Ac	<i>Medicago sativa</i>	Ms
<i>Arabidopsis lyrata</i>	Al	<i>Michelia figo</i>	Mf
<i>Arabidopsis thaliana</i>	At	<i>Momordica charantia</i>	Mc
<i>Aranda deborah</i>	Ad	<i>Nicotinia sylvestris</i>	Ns
<i>Asarum europaeum</i>	Ae	<i>Nicotinia tabacum</i>	Nt
<i>Berberis gilgiana</i>	Bg	<i>Oncidium cv.</i>	Oc
<i>Betula pendula</i>	Bp	<i>Orchis italica</i>	Oi
<i>Brassica napus</i>	Bn	<i>Oryza sativa</i>	Os
<i>Brassica oleracea</i>	Bo	<i>Pachysandra terminalis</i>	Pt
<i>Brassica oleracea var botrytis</i>	Bob	<i>Panax ginseng</i>	Pg
<i>Brassica rapa</i>	Br	<i>Papaver californicum</i>	Pc
<i>Calycanthus floridus</i>	Cf	<i>Papaver nudicaule</i>	Pn
<i>Canavalia lineata</i>	Cl	<i>Paulownia kawakamii</i>	Pk
<i>Capsicum annuum</i>	Ca	<i>Peperomia hirta</i>	Phir
<i>Ceratopteris richardii</i>	Cr	<i>Petunia hybrida</i>	Ph
<i>Chloranthus spicatus</i>	Cs	<i>Petunia inflata</i>	Pi
<i>Chrysanthemum x morifolium</i>	Cm	<i>Petunia integrifolia</i>	Pin
<i>Cimicifuga racemosa</i>	Cra	<i>Phalaenopsis equestris</i>	Pe
<i>Clematis chiisanensis</i>	Cc	<i>Physcomitrella patens</i>	Pp
<i>Clematis integrifolia</i>	Ci	<i>Picea abies</i>	Pa
<i>Corylus avellana</i>	Cav	<i>Picea marinea</i>	Pm
<i>Cryptomeria japonica</i>	Cj	<i>Pimpinella brachycarpa</i>	Pb
<i>Cucumis sativa</i>	Cus	<i>Pinus radiata</i>	Pr
<i>Cycas edentata</i>	Ce	<i>Pinus resinosa</i>	Pres
<i>Daucus carota</i>	Dc	<i>Piper magnificum</i>	Pmag
<i>Delphinium ajacis</i>	Da	<i>Pisum sativum</i>	Ps
<i>Dendrobium grex</i>	Dg	<i>Platanus occidentalis</i>	Po
<i>Dicentra eximia</i>	De	<i>Poa annua</i>	Pan
<i>Elaeis guineensis</i>	Eg	<i>Populus balsamifera</i>	Pb
<i>Eucalyptus globulus</i>	Egl	<i>Populus tomentosa</i>	Pto
<i>Eucalyptus grandis</i>	Eug	<i>Populus tremuloides</i>	Pt
<i>Fragaria x ananassa</i>	Fa	<i>Ranunculus bulbosus</i>	Rb
<i>Gerbera hybrida</i>	Gh	<i>Ranunculus ficaria</i>	Rf
<i>Glycine max</i>	Gm	<i>Rosa rugosa</i>	Rr
<i>Gnetum gnemon</i>	Gg	<i>Rosa x hybrida</i>	Rh
<i>Gnetum parvifolium</i>	Gp	<i>Rumex acetosa</i>	Ra
<i>Gossypium hirsutum</i>	Ghi	<i>Sagittaria montevidensis</i>	Sm
<i>Helianthus annuus</i>	Ha	<i>Sanguinaria canadensis</i>	Sc
<i>Helleborus orientalis</i>	Hor	<i>Saururus chinensis</i>	Sch
<i>Hemerocallis hybrid</i>	Hh	<i>Silene latifolia</i>	Sl
<i>Hieracium piloselloides</i>	Hp	<i>Sinapis alba</i>	Sa
<i>Hordeum vulgare</i>	Hv	<i>Solanum tuberosum</i>	St
<i>Hyacinthus orientalis</i>	Ho	<i>Sorghum bicolor</i>	Sb
<i>Hydrangea macrophylla</i>	Hm	<i>Syringa vulgaris</i>	Sv
<i>Ipomoea batatas</i>	Ib	<i>Tacca chantieri</i>	Tc
<i>Ipomoea nil</i>	In	<i>Thalictrum thalictroides</i>	Tt
<i>Juglans regia</i>	Jr	<i>Trautvetteria carolinensis</i>	Tca
<i>Lilium longiflorum</i>	Ll	<i>Triticum aestivum</i>	Ta
<i>Lilium regale</i>	Lr	<i>Trollius laxus</i>	Tl
<i>Liquidambar styraciflua</i>	Ls	<i>Vitis vinifera</i>	Vv
<i>Liriodendron tulipifera</i>	Lt	<i>Zea mays</i>	Zm
<i>Lolium perenne</i>	Lp		

# SQUAMOSA/API SUBFAMILY

## PaleoAPI Lineage

os-mads28-osa011675 LPPWMLRTSHT\*~  
 os-rap1b-ab041020 LPPWMLSHING\*~  
 os-fdrnads6-af139664 LPPWMLSHING\*~  
 os-mads14-af058697 LPPWMLSHING\*~  
 zm-m15-aj430632 LPPWMLSHLSS\*~  
 zm-m4-aj430641 LPPWMLSHLSC\*~  
 lte-mads1-af035378 LPPWMVSHLNGG\*~  
 lp-mads1-ay198326 LPPWMVSHLNGG\*~  
 hv-mads5-aj249144 LPLWMVSHING\*~  
 ta-tamads11-ab007504 LPLWMVSHING\*~  
 zm-mads3-af112150 LPPWMLSHLNA\*~  
 zm-zap1-l46400 LPPWMLSHLNA\*~  
 os-Osmads15-af058698 LPPWMLSHLNA\*~  
 lte-mads2-af035379 LPPWMLSHLNA\*~  
 lp-mads2-ay198327 LPPWMLSHLNA\*~  
 hv-mads8-aj249146 LPPWMLSHLNA\*~  
 sb-sbmads2-u32110-COR LPPWMLSHLNAR\*~  
 dg-domads2-af198175 LPPWMLSHVNGQ\*~  
 bp-mads4-x99654 MPPWMLSHING\*~  
 am-defh28-ay040247 IPPWLLQHVNQ\*~  
 ph-fbp29-af335245 MPPWMIRHVNNEG\*~  
 st-potm1-1-u23757 MPOWMLRHLNG\*~  
 sc-scm1-af002666 MPOWMLRHLNG\*~  
 le-tdr4-aam33098 MPOWMLRHLNN\*~  
 in-pnsah1-ab013105 MPOWMLSHLQG\*~  
 In-PnsAH2-AB013106 MPOWLLPHLSR\*~  
 nt-nap1-1-af009126 MPPWMLRHLNG\*~  
 ns-nsmads1-af068725 MPPWMLRHLNG\*~  
 nt-mads11-af385746 MPPWMLRHLNN\*~  
 ph-fbp26-af176783 MPPWMLRHLNG\*~  
 Ca-MADS6-AF130118 MPPWMLRHLNG\*~  
 ph-pfg-af176782 MPPWMLRHLNG\*~  
 bp-mads5-x99655 LPPWMLRHLNQ\*~  
 md-mads2-u78948 LPAWIVRHLNE\*~  
 sl-sl-m5-x80492 VPSWMLNHLAEQ\*~  
 mp-mpmads15-q948u1 MPPWMLRHVNE\*~  
 sa-samadsb-u25695 LPAWMLRPTTNE\*~  
 bob-fu1b-aj505842 LPAWMLRPTTKE\*~  
 bob-fu1d-aj505844 LPAWMLRPTTK\*~  
 bob-fu1c-aj505843 LPAWMLRPATNE\*~  
 at-fu1-u33473 LPAWMLRPTTTNE\*~  
 bob-fu1a-aj505841 LPAWMLRPTTNE\*~

## EuAPI Lineage

at-ap1-z16421 NCNLGCFAA\*~  
 pt-ap1-af034093 NCNLVRFCAA\*~  
 bo-boi1ap1-u67451 NCNLGCFAA\*~  
 bob-ap1c-aj505846 NCNLGCFAA\*~  
 bo-boi2ap1-u67452 NCNLGCFAA\*~  
 bob-ap1a-aj505845 NCNLGCFAA\*~  
 bo-ap1-z37968 NCNLGSFAA\*~  
 sa-madsc-2-af109403 NCNLGCFAA\*~  
 sa-ap1-x81480 NCNLGCFAA\*~  
 bo-boical-u67454 NCNLGYFAA\*~  
 bo-bocal-136926 NCNLGYFAA\*~  
 bob-bobcal-136927 NCNLGYFAA\*~  
 brp-aj251300 NCNLGYFAA\*~  
 at-cal-136925 N.YLGCYAA\*~  
 ha-ham75-af462152 SCHMRCFPS\*~  
 cm-cdm111-ay173054 SCHMRCFPS\*~  
 ha-ham92-ay173071 SHHLRCFPS\*~  
 nt-nap1-2-af009127 PCHMGCFAT\*~  
 nt-squa15-u63162 PCHMGCFAA\*~  
 ns-mads2-af068726 PCHMGCFAT\*~  
 ntmads5-af068724 PCHMGCFAT\*~  
 le-mads-mc-af448521 LYNMNKHL\*~  
 bp-mads3-x99653 SCHLGCFFAT\*~  
 ps-peam4-aj291298 TCHLGCFF\*~  
 md-mads5-aj000759 ECHLGCFAA\*~  
 pt-ap1-af034094 SCHLGCFFGT\*~  
 am-squamosa-x63701 SCHLGCFAA\*~  
 dc-mads1-aj271147 PCNLRCFA\*~

## TM3 SUBFAMILY

pa-DAL3-X79281-COR  
pr-prmads6-u90347  
pr-prmads8-aac27353  
pr-prmads4-u90345  
pr-prmads7-ab80810  
pr-prmads9-u90344  
pr-prmads5-u90346  
gg-ggm1-aj132207  
at-at5g51860  
at-at5g51870  
os-baa81886  
zm-mads1-af112148  
zm-ay104805  
os-fdrnads8-aad38369  
eg-opmads1-af207699  
at-at4g22950  
at-agl14-at4g11880  
egl-etl-aad16052  
at-agl20-at2g45660  
sa-madsa-t10422  
ph-fbp21-af335239  
nt-mads1-s46526  
ph-fbp20-af335238  
ph-fbp28-af335244  
pb-mads1-aac33475  
cm-cdm36-ay173065  
at-at5g62165-ay096509  
ph-fbp22-af335240  
mp-mpmads9-ab050651

EVNAQLVIRPP  
EVETQLVMRPP  
EVETQLVMRPP  
EVETQLVMRPP  
EVETQLVMRPP  
EVETQLVIRPP  
EVQTQLVMRPP  
DVETQLNIGPP  
DVETDLFIGFL  
EVETDLFIGLP  
DVETELFIGLP  
DVETELYIGLP  
DVETELYIGLP  
DVETDLYIGLP  
EVETELYIGWP  
EVETGLFIGPP  
EVVTDLFIGPP  
DVETELFIGPP  
EVETQLFIGLP  
EVETQLFIGLP  
DVETELFIGPP  
DVETELFIGPP  
DVETELFIGLP  
DVETELFIGLP  
EVETDLFIGLP  
EVETDLFIGLR  
EVETELFIGRP

## STMADS11 SUBFAMILY

st-mads11-t06996  
ph-fbp25-af335243  
hv-mads1-cab97349  
hv-mads1-2-cab97350  
os-cac29335  
zm-m20-a430634  
os-baa81880  
zm-m26-a430693  
zm-m19-a430633  
zm-m21-a430635  
bob-svpa-cad48304  
at-svp  
ib-aak27150  
le-jointless-q9fuy6  
pk-aaf22455  
mp-mpmads1-ab050643  
ph-fbp13-af335237  
st-mads16-t06995  
ib-aak27151  
at-agl24-af005158  
cl-af144623  
gg-ggm12-AJ132218

SDTSLKLCCLA~  
SDTSLKLGLP~  
SDTSLRLGLS~  
SDTSLRLGLS~  
SDTSLKLGLP~  
SDTSLRLGLP~  
SDVSLKLGLP~  
SDVSLKLGLP~  
SDVSLKLGLP~  
SDISLKLSP\*~  
SDISLRLGLP~  
SDTSLRLGLP~  
SDTSLKLGLP~  
SDTSLKGLA~  
SDTSLKLGLP~  
SDTSLKLGVP~  
SDTFLKLGLP~  
SITSLKLGLP~  
SDTSLKLGLP~  
SDTSLKLGLP~  
SDTSLKLGLP~  
SDTSLHLGLP~

## AGAMOUS SUBFAMILY

rr-baa90744	DQISLQLV*	Nt-AG-T03592	DQPSLQLV*
rr-baa90745	DQISLQLV*	Bn-AG-A43484	DQTALQLV*
md-mads-cac80858	DQISLQLV*	At-AG-P17839	DQTALQLV*
Cav-mads1-aad03486	DQMALQLV*	pres-aad01266	EQTTLQLG*
jr-cac38764	DQMALQLV*	pa-da12-t14847	EQTTLQLG*
rh-aad00025	DQISLQLV*	pe-mads1-af234617	QQTALQLG*
fa-stag1-af168468	DQVSLQLV*	pm-sag1a-aac97157	EQTTLQLG*
cus-cum1-aac08528	DNMALQLV*	ce-cyag-af492455	DQAAQLG*
pb-ptag1-aac06237	DQMALQLV*	Gg-GGM3-AJ132209	EQTALHLG~
Ra-s57586	NQTPLQLV*	mp-mpmads11-bab70746	DQTALHLG*
gh-gaga1-caa08800	DQTPLQLV*	mp-mpmads2-bab70737	EQTALQLG*
gh-gaga2-caa08801	DQTPLQLV*	ho-hag1-aad19360	QQTALQLG*
Ha-HAM45-aao18228	DQTPLQLV*	Os-bab32985	QPTTLQLG~
Ha-HAM59-aao18229	DQTPLQLV*	Os-MADS3-s59480	QPTTLQLG~
Cm-aao22984	DQTPLQLV*	zm-ucsd78a-aab81103	QPTTLQLG*
s1-s1m1-caa56655	DQTTLQLN*	hv-hvag1-af486648	QPTALQLG~
pin-pag11-aaa68001	DQTALQLV*	Le-TAGL11-aam33102	DHKR~~~~*
dc-mads4-cac81071	QHVPQLV*	Ph-fbp7-caa57311	DKKSLDLE*
Le-TAGL1-AY098735	DQTPLQLV*	Ph-Fbp11-caa57445	DKKSLQLE*
Bn-SHP1-aak00646	DQPPLQLV*	Mc-aao20104	DKKMLHLG*
At-ag11-SHP1-P29381	DQPPLQLV*	cus-cum10-aac08529	DKKMLHLG*
At-ag15-SHP2-P29385	DQPPLQLV*	Ghi-GHMADS2-aan15183	DKKILHLG*
ph-pmads3-q40885	DQPPLQLV*	Vv-MADS5-af373604	DKKVLHLG*
le-tag1-aaa34197	DQPPIQLV*	Md-MADS10-caa04324	DKKNLHLG*
ls-aad38119	DQTPLQLV*	AT-AGL11-Q38836	DKKILHLG*
Am-FAR-cab42988	DQLPLQLV*	zm-zag1-jq2289	DRKDFNDQ~
vv-MADS1-af265562	DQTALQLV*	hv-hvag2-af486649	DRKTLNSV~
Am-plena-A44343	DQTALQLV*	zm-zag2-caa56504	ATELNLGY~
Ph-fbp6-x68675	DQTALQLV*	OsMADS13-AF151693	PTELNLGY~
pg-gag2-caa86585	DQTALQLV*	os-agamous-bab90168	QTALHLGY~
pb-ptaq2-aac06238	DOLFS~~~*	Ho-MADS1-aaf08830	QTALHLGY~

## AGL2 SUBFAMILY PART I

sa-samadsd-Y08626	YMLGWL P	~
at-agl9-SEP3-at1g24260	YMLGWL P	~
ph-fbp2-m91666	YMAAGWL P	*~
ns-mads3-af068722	YMAAGWL P	*~
le-tdr5-x60480-COR	YMAAGWL P	*~
vv-mads4-af373603	YMPGWL P	*~
am-defh200-s71757	YISGWL P	*~
bp-mads1-cab95648	YMSGWL P	*~
eug-egm1-af029975-COR	FMPGWF P	*~
cm-cdm44-ay173057	YMPGWYQ	*~
ps-t06543	YMGGLWLP	*~
am-defh72-s71756	YNMTGWL P	*~
ad-x69107	YMPPGWLG	~
at-agl3-aa136250	FFPGWMV	*~
bob-agl3a-cad48302	FFPGWMV	*~
AT-agl2-SEP1-m55552	YIIPGWML	*~
bob-sep1a-cad48303	YIIPDWML	*~
at-agl4-SEP2-at3g02310	YIIPGWML	*~
cus-cagl2-af135962	FLIPGWML	*~
vv-mads2-af373601	FIPGWML	*~
pt-magl2-af185574	FIPGWML	*~
pt-magl4-af185574	FIPGWML	*~
pt-af034095	FIPGWML	*~
md-mads1-t17023	FIPGWML	*~
md-mdmads8-caa04919	FIPGWML	*~
md-mdmads9-caa04920	FIPGWML	*~
nt-mads4-af068723	FIPGWML	*~
ph-fbp9-af335236	FIPGWML	*~
ca-mads2-af129875	FIHGWML	*~
ph-fbp23-af335241	FIHGWML	*~
eug-egm3-aac78284	FIPGWML	*~
md-mads7-aj00076	YIIPGWML	*~
md-mads3-u78949	YIIPGWML	*~
md-mads6-aj000760	YIIPGWML	*~
dc-cmb1-q39685	FAQGWML	*~
md-mads4-u78950	FFPGWML	*~
mp-mpmads13-bab70747	YMPGWL V	*~
fa-RIN-af484683	FIPGWML	*~
Le-Lemadsrin-AF448522	VVPGWML	*~
ph-fbp4-af335234	VLPGWML	*~
dc-mads5-cac81072	VIIPGWML	*~
le-tagl2-aam33104	MIIPGWML	*~
le-tm29-cac83066	MIIPGWML	*~
ph-pmads12	MVPGWML	*~
ph-fbp5-af335235	IIIPGWML	*~
am-defh49-s78015	LVIPGWML	*~
ha-ham137-ay173072	QM QGWPA	*~
gh-grcd1-aj400623	QM QGWPA	*~
cm-cdm77-ay173058	QM QGWPA	*~

## AGL2 SUBFAMILY PART II (monocotyledons)

Zm-m24-AJ430638	P A W M S *	
Zm-m31-AJ43060	P A W M A *	
os-baa81882	P E W M A *	
ta-bj219318	P A W M A *	
ta-bj276769	Q A Y V D Q P N N K	S A G W I *
hv-hvmads7-aj249145	Q A Y M D Q L N N R	S A G W I *
ta-bj218990	Q A Y M D Q L N S R	S A G W I *
zm-zmm8-y09303	Q A Y M D Q L . N E	S . G W I *
zm-zmm14-cab85962	Q A Y M D Q L N N E	S . G W I *
os-osmads1-lhs-s53306	Q A Y M D H L S N E	S . G W I *
ta-bj313906	H A Y L D H L N K E *	
ta-bj265532	H A Y L D H L N K E *	
ta-bj211160	N . F L D Q L N K E *	
zm-zmm3-y09301	Q A Y M D H L N N D *	
os-osmads5-u78890	Q A Y M D H L N . Q *	
zm-bg837363	F M P T W L P *	
zm-zmm7-caa70485	F M P A W L P *	
os-osmads8-u78892	F M P T W L P *	
sb-sbmads1-u49734-cor	F M P T W L P *	
ta-bq902720	F M P T W L P *	
ta-bj274042	F M P P W L P *	
ta-bj275311	F M P P W L P *	
hv-mads9-cab97355	F M P P W L P *	
os-fdrnads1-AF141966	Y M P P W L P *	

## AGL6 SUBFAMILY

zm-zag3-t03398	F M L G W V L *
zm-zag5-t03408	F M L G W V L *
pa-mads1-af372840	F M L G W V L *
ta-tamads12-baa33458	F M L G W V L *
os-osmads6-t04167	F M L G W V L *
lp-mads4-ay198329	F M L G W V L ~
bob-agl6a-cad48306	F V Q D W F L *
at-agl13-at3g61120	F V Q * W V S ~
at-agl6-at2g45650	F V Q G W V L *
vv-mads3-af373602	F I Q G W V L *
md-mads11-caa04325	F I Q G W V L *
ph-pmads4-baa94287	I M Q G W G L *
mp-mpmads3-ab050645	F M H G W I L *
mp-mpmads4-ab050646-COR	F I Q G W V L *
ap-apmads3-ab079261	F M L G W V L *
gg-ggm9-cab44455	Y I . . W W V *
gp-gpmads3-baa85630	Y I . . W W V *
gg-ggm11-aj132217-COR	Y I Q G W V V *
pr-mads2-t09571	Y M Q G W M V *
pres-t10486	Y M Q G W M V *
pr-prmads3-t09603	Y M Q G W W V *
pa-dal1-t14846	Y M Q G W W V *



## GLO/PI SUBFAMILY

rb-rbpi-2-ac42575	PFTFLVQPIHPNFQ	t1-pi-2type2-aao26547	PFSLQIQTIQPNLQ
tca-pi-2-aao26554	PFTFRVQPTHPNLQ	t1-pi-3-type2-aao26549	PFSLQIQPIHPNLQ
phir-phpi-1-aac42580	PIAFHVQPLHPNLQ	ho-pi1-af134114	PMALRVQPVQPNLQ
Pmag-pmpi-1-aac42581	PIAFHVQPLHPNLQ	ho-hpi2-af134115	PMALRVQPVQPNLQ
rf-rfpi-1-aac42573	PFTFQLHPSQPNLQ	lr-lrgloa-ab071379	PMAFRVQPIQPNLH
rf-pi-1b-aao26532	PFTFRLHPSQPNLQ	ap-pi-bac66962	PMAFRVQPIHPNLQ
rb-rbpi-1-aac42574	PFTFRVQPAQPNLQ	tc-pi-af230713	PLAFRVQPLQPNLQ
tca-pi-1-aao26553	PFIYRVQPTQPNLQ	oi-bac22579	PMTFRVQPFQPNLH
at-PI-d30807	.FGYRVQPIQPNLQ	cs-pi-af230710	PFIYRVQPIQPNLQ
a1-pi-aaf25591	.FGYRVQPIQPNLQ	aq-pi type2-aao26484	PFAFRAQPIQPNLQ
nt-glo-x67959	PFAFRVQPMQPNLQ	aq-pi type1-aao26485	PFAFRAQPIQPNLQ
ph-fbp1-m91190	PFAFRVQPMQPNLQ	bg-pi-1-aao26508	PFAFRGQPIQPNLQ
sv-svpi-1-aac42576	PFAFRVQPMQPNLQ	bg-pi-2-aao26509	PFAFCVQPIQPNLQ
am-glo-q03378	PFAFRVQPMQPNLQ	da-dapi-1-aac42577	PFTFRAQPMQPNLQ
gh-gglo1-aj009726	PFSFRVQPMQPNLH	mp-mpmads8-ab050650	PFAFRVQPIQPNLH
cm-cdm86-aao22986	PFSFRVQPMQPNLH	mf-mfpi-1-af052863	PFTFRVQPIQPNLH
ha-ham31-aao18230	PFSFRVQPMQPNLH	ltpi1-af052864	PFTFRLOPIQPNLH
md-pi-aj291490	PFAFRVQPIQPNLQ	cf-pi-1-af230708	PFAFRIQPIQPNLH
rr-bp-ab038462	PFALRVQPNQPNLH	cf-pi-2-af230709	QLAFRVQPLQPNLQ
cus-cum26-af043255	PFAFRVQPIQPNLQ	de-depi-1-af052857	PFAFRVQPIQPNLH
ph-pmads2-x69947	PFALRVQPMQPNLH	aa-pi-aao26500	PFTFRVQPIQPNLQ
bp-mads2-cad32764	PFAFRVQPIQPNLQ	tt-pi-aao26537	PFTFRVQPIQPNLQ
hm-pi-af230711	PFAFRVQPIQPNLQ	hor-pi-1-aao26526	PFAFHIQPMQPNLQ
s1-s1m2-x80489	PYGFVRVQPMQPNLQ	hor-pi-2-aao26527	PYNFHVQQMQPNLQ
Eug-egm2-af029976	PSTYHVQPIQPNLQ	cra-pi-3-type1-aao26515	PFAFH.....NLQ
ms-ng19-af335473	PFSFRLQPMQ..LH	cra-pi-3type2-aao26516	PLAFH.....NLQ
dc-mads2-cac81069	PFAFRVQPNQPNLH	cra-pi-1-aao26513	PFSFRVQPIQPNLQ
zm-m29-cac33850	PFTFRVQPNHPNLQ	cra-pi-2-aao26514	PFSFRVQPIQPNLQ
zm-m18-cac33849	PFTFRVQPNHPNLQ	ae-pi-af230707	PFAFCVQPMQPNLH
os-mads2-t03894	PITFRVQPSHPNLQ	hor-pi-3-aao26528	PFSFRVQPMQPNLH
zm-m16-cac33848	PITFRVQPSHPNLQ	pn-pnpi-1-aac42570	PFGFQVPPMQPNLT
os-mads4-137527	PFTFRVQPSHPNLQ	sc-scpi-af130871	PFAFCVQAIQPNVH
sm-pi-aaf73941	PFGFRVQPMQPNLQ	pmag-pmpi-2-af052867	PFAFRVQPIQPNLQ
an-pi-2type1-aao26495	PFTFRLHSTKPNLQ	pa-dal11-1-af158539	DPELRLQPNQPNLK
an-pi-2-type2-aao26495	PFTFLVHSTKPNLQ	pa-dal11-2-af158540	DPELRLQPNQPNLK
ci-pi-2-aao26521	PFSFCVHHPAKPDLQ	pa-dal13-1-af158543	APLLRLQPNQPNLQ
an-pi-1type1-aao26493	PFTFRVQPIQPNLQ	pa-dal13-2-af158544	APLLRLQPNQPNLQ
an-pi-1type2-aao26494	PFTFRVQPIQPNLQ	Pr-prdgl-af120097	APLLRLQPNQPNLQ
ci-pi-1-aao26520	PFTFRVQPIQPNLQ	cj-mads1-aal05440	PPAFRVQPSQPNLQ
t1-pi-1type1-aao26544	PFNFRVQPIQPNLQ	gg-ggm15-cac13991	LDDVCYQP.QPNLQ
t1-pi-1-type2-aao2654	...FRVQPIQPNLQ		
t1-pi-2type1-aao26546	PFSLQIQTIHPNLQ		

## Bs SUBFAMILY

ae-ap3-2-af23069  
 gg-ggm13-cab4445  
 zm-m17-cac81053  
 at-abs-cac85664  
 ph-fbp24-af33524  
 am-defh21-cac85225

P . FRLQPAQPNLQD  
 PAFRLQPTQPNLQE  
 . FRLQPTQPNLQD  
 PTYNLQLAQPNLQN  
 P . YRLQPSHPNLQD  
 . SYRLQPTQPNLQD

LQL\*  
 LQL\*  
 LQL\*

## DEF/AP3 SUBFAMILY

### EuAP3 Lineage

bn-ap3-af124814  
 bo-boi2ap3-u67455  
 bo-boi1ap3-u67453  
 at-ap3-ay087369  
 am-deficiens-x52023  
 Sv-svap3-af052869  
 Hm-hmap3-af230702  
 Hp-hpdef2-af180365  
 Hp-hpdef1-af180364  
 gh-gdef2-aj009725  
 cm-cdm115-aa022985  
 cm-cdm19-ay173064  
 jr-ap3-j313089  
 ra-d1-u28482  
 sl-sl3m3-x80490  
 ph-gp-x69946  
 Nt-ntdef-x96428  
 Te-ap3-af052868  
 st-stdef4-x67508  
 dc-mads3-aj271149  
 ms-nmh7-141727  
 gm-ax478039

SDIITFHLL E\*  
 SDIITFHLL E\*  
 SDIITFHLL E\*  
 SDIITFHLL E\*  
 SDLTTFALLE\*  
 SDLTTFALLE\*  
 SDLTITYALLD\*  
 SDLTITYALLG\*  
 SDLTITYALLG\*  
 SDLTITYALLG\*  
 SDLTITYSLLG\*  
 SDLTITYGLFG\*  
 SDLTITYTLL E\*  
 SCLTITYTYL E\*  
 SCVTITYALL . \*  
 SDITTFALLE\*  
 SDITTFAL . A\*  
 SDITTFAL . G\*  
 SDITTFAL . G\*  
 SDL . ITFA . . \*  
 SDLTITYPLLF\*  
 SDLTITYPLLF\*

# DEF/AP3 SUBFAMILY

## PaleoAP3 Lineage

pc-pcap3-aac42587	YNQH YV*	mp-mpmads7-ab050649	HDLRLA*
pn-pnap3-1-aac42588	YSQH YA*	lt-ltap3-af052878	HDLRLA*
ha-ham91-aao18231	HGLRLD*	cf-ap3-1-af230699	NDLRLA*
oc-aao45824	RLAH CL*	cf-ap3-2-af230700	HDLRLA*
sm-ap3-aaf73934	HELRLA*	ae-ap3-1-af230697	HDLRLA*
rb-rbap3-2-aad31697	YSLRLA*	bg-ap3-2type1-aao26506	YDFHLA*
rf-rfap3-2-af130870	YGLSLA*	bg-ap3-2type2-aao26507	YDFHLD*
an-ap3-3type2-aao26491	YGFQLA~	an-ap3-2-aao26490	YGLTLA*
an-ap3-3type2-aao26492	YGFQLA~	ci-ap3-2-aao26519	YGLTLA*
hor-ap3-3a-aao26524	YNLQLA*	aa-ap3-2-aao26498	YGLSLA*
hor-ap3-3b-aao26525	SSLQLA*	tt-ap3-2b-aao26536	YGLSLV*
t1-ap3-3type1-aao26542	YNLRLA*	t1-ap3-2type1-aao26540	YGLSLA*
t1-ap3-3type2-aao26543	YNLRLA*	t1-ap3-2type2-aao26541	YGLSLA*
aa-ap3-3-aao26499	HNLRLA*	cra-ap3-2-aao26511	YGLRLA*
ac-ap3-3-aao26503	HNLRLA*	hor-ap3-2-aao26523	YLSLSA*
cra-ap3-3-aao26512	YNLRLG*	pt-ptap3-1-af052870	HNLHLA~
rf-ap3-3-aao26531	HNLRLA*	pt-ptap3-2-af052871	HNLHLA*
bg-ap3-1-aao26505	YFGVMH*	pb-ptd-aac13695	HELRLP*
os-ab003323	HDLRLG*	pto-ptap3-aao49713	HELRLP*
Os-osmads16-af077760	HDLRLG*	le-tdr6-x60759	RDLRLS*
Ta-tamads51-ab007506	HDLRLG*	ph-tm6-af230704	RDLRLA*
zm-silky1-af181479	HDLRLG*	hm-tm6-af230703	HDLRLA*
hh-mads1-af209729	HDLRLA*	rr-ab055966	HDLRLA*
l1-mads1-af503913	HDLRLA*	rb-ap3-1-af052876	HDLRLV*
Lr-lrdef-ab071378	HDLRLA*	tca-ap3-aao26552	HGLRLA*
tc-ap3-af230706	HDLRLA*	rf-rfap3-1-af052854	HDLRLA*
aq-ap3-1type1-aao26483	HDLRLA*	an-ap3-1-aao26489	HQLRLA*
aq-ap3-1type2-aao26488	HDLRLA*	ci-ap3-1-aao26518	HQLRLA*
po-ap3-1-aao26529	HDLRLA*	t1-ap3-1type1-aao26538	HDLRLG*
po-ap3-2-aao26530	RDLRLA*	t1-ap3-1type2-aao26539	HDLRLG*
aq-ap3-2type1-aao26487	NDLRLA*	aa-ap3-1-aao26497	EDLRLG*
aq-ap3-2_type_1-ay162839	NDLRLA*	tt-ap3-1-aao26534	EDLRLG*
de-deap3-1-af052875	HDLRLA*	cra-ap3-1-aao26510	HDLRLG*
Sc-scap3-af130868	NDLRLA*	hor-ap3-1-aao26522	SDLRS*
pn-pnap3-2-af052874	HDLRLA*	Phi r-phap3-af052879	YDLRLA*
cs-ap3-af230701	H*LRLG*	pa-da112-af158541	LDLKL*
mf-mfap3-af052877	HDLRLA*		