Supplemental Figures and Tables legends

Figure S1. Related to Figure 1.

A- Conservation of the "rhoptry" DNA sequence motif (in red) in invasion-related gene promoters from *P*. *falciparum* (Pf), *P. berghei* (Pb), *P. vivax* (Pv) or *P. knowlesi* (Pk). See also figure 1B.

B- Strategy used to C-terminally tag PfAP2-I with GFP. The presence of a PCR product with primers 1 and 4 or 2 and 4 only in the transgenic line demonstrates that the locus has been correctly modified. The primer sequences are listed in Table S1. See also figures 1D, E.

Figure S2. Related to Figure 1.

Protein sequence alignment of the AP2-I sequence of *P. falciparum* (Pf), *P. berghei* (Pb), *P. vivax* (Pv) and *P. knowlesi* (Pk). Boxed in dark grey is the ACDC domain, in dark blue are the AP2 domains, in light blue is the AT-hook domain and in light grey is the NLS. In red are highlighted the conserved amino acid residues. See also figure 1C.

Figure S3. Related to Figures 2 and 3.

A- ChIP-qPCR from a parasite line expressing PfAP2-I-GFP indicates that PfAP2-I does not associate with the *SPE2* motifs bound by PfSIP2. 3D7 *wt* and PfSIP2N-HA parasites were used as negative and positive controls, respectively. The results are shown as fold enrichment of ChIP performed with anti-GFP antibody versus ChIP performed with IgG (n=3). Data are represented as mean \pm SD. See also figure 2D.

B- DNA motif analysis using DREME identified GTGCA as the most common DNA motif within all of the PfAP2-I ChIP-seq peaks (e-value=2E-25). The next five highest scoring motifs after GTGCA found in the ChIP-seq peaks are shown on the right. See Table S2 for a complete list of DREME results. Matching the top DREME motif against known DNA motifs bound by *Plasmodium* AP2 domains (Campbell *et al.*, 2010) using TOMTOM shows that GTGCA matches the DNA motif bound by the third AP2 domain of PfAP2-I and the AP2 domain of PfSIP2 (lowest E-value) and has reduced similarity to other known AP2-associated sequence motifs. See also figure 3A.

C- Plot representing the distance of the top DREME DNA motif site relative to the ChIP-seq peak summits across the three ChIP biological replicates. The mean motif distance from the peak summit are +3, +2 and +3 bp for replicates 1-3.

D- Integrative genomics viewer (IGV) plot (Thorvaldsdottir *et al.*, 2013) showing the genomic region encoding *msp2*, *msp4* and *msp5* and containing ChIP-seq trimmed peaks 3 and 4. The log2 enrichment ChIP/Input for each ChIP replicate is shown. A grey block represents the peak called by MACS2 (the number of the correspondent trimmed peak is indicated), arrows represent genes, and small black, red and blue lines represent the presence of a predicted DNA motif associated with PfAP2-I-D3 or –D2 or –D1, respectively.

E- Sanger sequencing verifying that the correct mutations were introduced in PfAP2-I-GFP::*msp5*MUT parasites, compared to the parental strain, PfAP2-I-GFP. Boxed in red are the mutated nucleotides. See also figure 3B.

F- ChIP-qPCR with PfAP2-I-GFP-*msp5*MUT parasites shows that PfAP2-I-GFP binding to *mps5* is diminished (p=0.07) (red box). PfAP2-I binding to other genes is not significantly altered (p>1) (AP2-I ChIP-seq targets). The results are shown as fold enrichment of ChIP performed with anti-GFP antibody versus ChIP performed with IgG (n=3). Data are represented as mean \pm SD. 3D7 *wt* and PfAP2-I-GFP parasites were used as negative and positive controls, respectively. See also figure 3C.

G- Time points collected for the microarray timecourse experiment performed with PfAP2-GFP and PfAP2-I-GFP-*msp5*MUT parasites. See also Table S3 and Figure 3D.

H- Plot of the average Rnits (Sangurdekar, 2014) p-value for all transcripts detected by microarray (minus antigenic variant genes) showing that the *msp5* gene is one of the most significantly altered transcripts (average p-value 4.65E-2) (see Table S3 for full list of Rnits p-values). Rnits compares multiple time-series expression data sets by summarizing probes into gene-level information. For all genes, it fits a series of B-splines with varying curvature and degrees of freedom. Under the null hypothesis H_0, a single model is fit for all data sets. P-values from the hypothesis test are then plotted and the least complex spline parameters that result in uniformly distributed null p-values are automatically chosen. Each gene is attributed a p-value from 0 to 1 until all p-values are uniformly distributed. While genes with p-values closer to 1 have less changes in transcription levels between expression data sets, genes with p-values closer to 0 have higher transcription levels differences

between time-series expression data sets. The transcription level of *msp5* drops significantly in the mutant parasites, and is therefore attributed a low Rnits p-value.

I- RT-qPCR experiments show that while *msp5* transcript levels are reduced in PfAP2-I-GFP-*msp5*MUT parasites, the transcript levels of other genes are similar in *wt* and mutant parasites (PfAP2-I-GFP) at 42h post-invasion. The results are the average of 4 independent experiments. Data are represented as mean \pm SEM. See also figure 3D.

Figure S4. Related to Figure 4.

A- Integrative genomics viewer (IGV) plots for several PfAP2-I target genes. The log2 enrichment ChIP/Input for each ChIP replicate is shown. Grey blocks represent the peaks called by MACS2 (the peak number indicated above each peak), arrows represent genes, and small black, red and blue lines represent the presence of a predicted DNA motif associated with PfAP2-I-D3 or –D2 or –D1, respectively.

B- Within 250bp of the summit of each ChIP-seq peak, the DNA sequence motifs associated with PfAP2-I-D3 and –D2 are identified in the majority of peaks. However the motif associated with D1 is less commonly found. Black boxes represent the presence of the DNA motif within the peak and each row represents a PfAP2-I ChIP-seq peak. Although the D2 DNA motif can be seen in approximately 80% of the peaks containing a D3 DNA motif, the same trend is observed genome-wide. In a random test of 1000 genomic intervals similar in length to the trimmed peaks found by ChIP-seq, D2 and D3 motifs co-occur more frequently than individual occurrences (data not shown). Thus, the high percentage of motif co-occurrence seen in our data does not suggest biological significance. Black dots indicate genes tested by ChIP-qPCR. See also Figure 4.

Figure S5. Related to Figures 4 and 5.

A- Protein alignment of the three AP2 domains of PfAP2-I (PF3D7_1007700) and the AP2 domain of PfAP2-Sp (PF3D7_1466400). The conserved residues are highlighted in red. Highlighted with * are the mutated residues. See also figure 4A.

B- Western blots with anti-GFP indicate that PfAP2-I-GFP is expressed in the PfAP2-I-GFP-D1mut and PfAP2-I-GFP-D2mut strains. See also figures 4A-B.

C- Western blot probed with anti-GST antibody showing expression of GST-D1wt/mut, GST-D2wt/mut and GST-D3wt/mut. ni: non-induced, i: induced. See also figure 4C.

D- Sanger sequencing shows that the correct mutations were introduced into the coding sequence of D1 and D2 in PfAP2-I-GFP-D1mut or PfAP2-I-GFP-D2mut parasites (bottom), as compared to the parental strain, PfAP2-I-GFP (top). See also figures 4A-B.

E- Coomassie blue stained SDS-PAGE gels showing purified GST-fusion proteins used for PBM analysis: PfAP2-I AT-hook fused to the second AP2 domain (GST-AT-D2) (left), the purified AT-hook domain alone (GST-AT) (middle) and the second PfAP2-I AP2 domain (GST-D2) (right) - lysate (L), flow-through (FT), wash (W) and elutions (E). See also figure 4C.

F- PBM experiments performed using purified GST-fused proteins or bacterial lysate from cells expressing GST-AT-D2, GST-AT or GST-D2 (as control), showing that GST-AT-D2 does not bind the same DNA motif as GST-D2 but rather it binds the same motif as GST-AT. The assay was performed twice on different microarray versions (see experimental procedures) with independent protein purifications. See Table S6 for PBM position weight matrices.

G- Parasites expressing the NLS sequence of PfSIP2 (Flueck *et al.*, 2010) fused to GFP were used as one of the IP controls. WB using anti-GFP antibodies shows that the NLS targets the protein to the nucleus but not exclusively. Histone H3 was used as a nuclear marker. See also figure 5A.

Figure S6. Related to Figure 5.

A- ChIP-qPCR of the same genes as shown in figure 2C demonstrates that PfBDP1-HA and PfCHD1-GFP bind the same genes as PfAP2-I-GFP. The results are shown as fold enrichment of ChIP performed with anti-GFP antibody versus ChIP performed with IgG and are the average of 3 biological replicates. Data are represented as mean \pm SD. 3D7 *wt* parasites were used as a negative control (see also Figure 5). Figure 5C and S6B show some of the same data. *gexp02* (PF3D7_1102500) was included as an additional negative control for PfBDP1 binding due to the unexpected binding of the protein to *ama1*, a non PfBDP1 ChIP-seq gene target (PfBDP1 binding to this gene, as diagnosed by ChIP-qPCR, had been previously detected by Josling *et al.*, 2015).

B- DNA motif discovery analysis using DREME shows that GTGCA is the most common DNA motif bound by PfBDP1-HA (1.5E-12) by ChIP-seq (top left). This DNA motif is identical to the one bound by PfAP2-I by ChIP-seq using DREME (bottom left). We also find a second DNA motif (C/GCACCT) bound by PfBDP1 (4.2E-5) (top right). This second DNA motif is similar to the one previously predicted using Gene Enrichment

Motif Searching (GEMS) to be enriched in micronemal gene promoters (Young *et al.*, 2008) (bottom right). See also figure 5B.

C- Alignment of the nucleotide sequences found in the promoters of several micronemal and rhoptry neck genes shows that the second PfBDP1-bound motif (TAACT) (red) identified in C is conserved. Black dots highlight PfBDP1-HA ChIP-seq targets. We note that some of these genes are not PfAP2-I-GFP ChIP-seq targets.

D- The M20.1 ACAACT "micronemal" DNA motif is also found in the promoters shown in D. These results suggest that a TF responsible for regulating micronemal genes may bind a TAACT or ACAACT DNA motif, as both motifs are present upstream of these genes. However, an interaction between the TF and PfBDP1 may not be required for all micronemal genes to be transcribed, since not all of the genes were identified in the Josling *et al.*, 2015).

Figure S7. Related to Figure 7.

A- ChIP-qPCR positive control demonstrating that PfBDP1 associates with a known target locus at the trophozoite stage. This gene was selected from the PfBDP1 ChIP-seq trophozoite target genes list from (Josling *et al.*, 2015). Data are represented as mean \pm SD and n=2. See also figure 7A.

B- Strategy used to C-terminally tag PfBDP1 with HA and the destabilization domain DD in the PfAP2-I-GFP expressing parasite. The presence of a PCR product for the reactions performed with primers 5 and 7 only in transgenic line demonstrates that the locus has been correctly modified. The primers sequences are listed in Table S1. See also figures 7B-D.

Table S1. List of primers used in this work. Related to the STAR methods.

Table S2. ChIP-seq data and GO-term searches. Related to Figures 2 and 3.

Table containing: 1)-6) MACS2 peak calling results for PfAP2-I-GFP ChIP-seq with anti-GFP/IgG antibodies for replicates 1, 2 and 3, 7) the list of peaks present in 2 out of 3 and 3 out of 3 biological replicates (merged peaks), 8) the list of merged peals trimmed to just include the overlapping genomic regions between biological replicates (trimmed peaks), the gene annotations corresponding to each trimmed peak and the peaks overlapping between PfAP2-I-GFP (this study) and PfBDP1-HA (Josling et al., 2015) ChIP-seq, 9)-11) GO-term searches

and 12)-14) a list of DNA motif occurrences bound by PfAP2-I AP2 domains and PfBDP1 as determined by FIMO and DREME.

Table S3. DNA microarray data. Related to Figure 3.

Results of the two DNA microarrays experiments performed with PfAP2-I-GFP and PfAP2-I-GFP::*msp5*MUT parasites (for a scheme of the collected time points see Figure S3G). Pearson correlation shows that the same time points were collected for PfAP2-I-GFP and PfAP2-I-GFP::*msp5*MUT parasites and that there is a high correlation between the two independent timecourses. Rnits (Sangurdekar, 2014), which attributes a p-value to each gene depending on the changes in transcript level, was used to compare the changes in gene expression on mutant parasites versus the parental line. The transcript that changed the most (lower p-value) on both timecourses on mutant parasites versus wt was HsDHFR, as expected, since PfAP2-I-GFP::*msp5*MUT express this gene (they were selected with WR) but PfAP2-I-GFP parasites do not. In order to determine which genes changed consistently between the two independent timecourses, Rnits was run on both timecourses and the data compared. Transcript levels for 49 genes had lower p-values than *msp5* but 61% of these (30/49) were genes encoding for either ribosomal RNA or antigenic variants or spliceosomal RNA, which are known to significantly and consistently change expression values between experiments (Rovira-Graells et al., 2012) (see also Figure S3H).

Table S4. IP proteomics results. Related to Figure 5.

IP proteomics results, including control IP experiments and SAINT results.

Table S5. ChIP-seq target genes microarray data. Related to Figure 2 and 6.

Expression patterns of ChIP-Seq target genes as determined by previous microarray studies (Bozdech *et al.*, 2003; Hu *et al.*, 2010; Le Roch *et al.*, 2003).

Table S6. Position weight matrices for all AP2 domains tested in this study. Related to Figure 4.

Figure S1 A

В

A		pfrap2	TTTA	TA TGC	TATGAA	G TGCA	AATGA				
		pbrap2	CATT	CA TGC		G TGCA	AAGAT				
		pvrap2	GTGT	TA TGC		G TGCA	AAGAG				
		pkrap2	TAAG	TA TGC	TCTACGT	G TGCA	GTTGA				
		pfrhoph2	TTTA	TA TGC	TATATGA ⁻	TGTGTTT	TAGGT				
		pbrhoph2	AAA	A <mark>ATGC/</mark>		ATTATG	ATATG				
		pvrhoph2	GCC	т <mark>сТGC/</mark>		AGCGGTT	GCCCG				
		pkrhoph2	ATTT	CA TGC	Астсссст	GCGCCC	TCACG				
		pfclag9	AATA	TA TGC	AATATTA ⁻	TTAGTAT	AAAAT				
		pbclag9	AAAT	CA <mark>TGC</mark>		TATGTAT	GTTCA				
		pvclag9	AGTG	A <mark>ATGC/</mark>	ATACCCAA	AAAAGC <mark>A</mark>	AAAA				
		pkclag9	TAAT	CT TGC	A GTAAAAT	GAAAATT	TGGTT				
		pfasp	TACTA	AA <mark>TGC</mark>	TTTAAAC	TGCA	TTAT				
		pbasp	TGAA	TA TGC	CTTATT	TGCA	TAAA				
		pvasp	GTTG	TA TGC	GCG-GAA	A TGCA	IGTAT				
		pfmsp4	TGAAC	CA <mark>TGCA</mark>	AAAN ₄₇ TC	A TGCA T	ГСТТТ				
		pbmsp4	GATTAC	GT TGCA	AAAN ₃₄ TG	T TGCA	ATATT				
		pvmsp4	CTTTG	тG TGC /	GAN ₂₉ CTC	TGCA T	ATTT				
		pkmsp4	CATCA	CTTGC/	ATTTAGAA	ATGCA	GAAT				
		pfgap45	TATT	тG TGC	AATAA-AAG	GTGCA	AGAA				
		pbgap45	GTCTA	TA TGC	ATGAA	GTGCA	CCACA				
		pvgap45	GTGTA	TA TGC	AATAAAC	TGCA	ADDDC				
		pkgap45	GCCTT	TC TGC	AATAAAG	TGCA	TCAGA				
		pfgap50	AATA	TA TGC	ATATAAAT.	ATTACAC	TTAAA				
		pbgap50	CATTA	TATGC	ATAATTAC	ACACAC	AGTATT				
		pvgap50	GTCCA	CATGC	A	- TGCA T	rggga				
		pkgap50	TCCCC	TATGC		GTATGC	CACCC				
B				<u>1</u> ►	<u>2</u> →			endogen	ous locus		
	5'pfap2-l	pfa	p2-I		3'	pfap2-I	_				
	<u> </u>				3						
					. 2 ∖	_	_	plasmid			
		—— 5'hsp86	BSD	3'PbDT	pfap2-I G	FP 3'hrp2	2				
				1	2 3	4		modified	locus		
	5'pfap2-l	pfa	o2-l		G	FP 3'hrp2	25'hsp	86 BSD 3	'PbDT —		
		. ,	4 4 5		< <u>√</u> <	4	Kb <u>1+4</u>	4 2+4 2+3 1+4 2+4 2+3			
1+4: 2.4 Kl						-					
			—2+4: 0. —2+2 : 0.	9KD			d.8=				
			-2+3: 1.	/KD				3D7 PfAP2	2-I-GFP		

Figur	0	C	2												
i igui		0	10	20	30	40	50	60	70	80		DC ₁₀₀	110	120	130
PfAP2-I PbAP2-I PkAP2-I PvAP2-I	I HETL HETL HEAL	VNE VNE ATD	RDINIKNE NNINTKSE KNITSKNE	+ NSKDNNKEMM Dgnehsekla Ernengnell Frnengtdi	IGMSHMNHDD Eysnnakgge Tglstlkehe Tglaslnehe	QDLIYEKKDI Neliyekkdi Gevtyekkei Gevoydkkei	DIESLRLHYI IYMNYI DRE-ISMHMN DRE-TSMHMN	DQEKLDVPSLV QDKLDVPSLV QDKLDVPSLV QDKLDVPSLV	EICKQQLIY EICKQQLIY EICKQQLIY	ILKDMCSDCST TLKDMCTDSNN TLKDMSADSSS TLKDMCADSNS	TDEKTSFLY SDEKTSFLY SDEKTSFMY	HLNRLRSAVT HLNRLRNALT HLNRLRNAVT HLNRLRNAVT	VYDLHNYIAVI VYDLHNYIAVI VYDLHNYIAVI VYDLHNYIAVI	GPCLSYNKL GPCLSYNKL GPCLSYNKL	PSTHNI PSTHNI PSTHNI PSTHNI
Consensus	HELL	vnŧ	.#iniKnE	#.n#ne\$.	ig.sn	.#liY#KK#	ie\$hy.	Q#KLDVPSLV	EICKQQLIY	ILKOMc.Dsn.	sDEKt.SF\$Y	HLNRLRnAvT	VVDLHNYIAVI	GPCLSYNKL	PSTANI
	131		140	150	160	170	180	190	200	210	220	230	240	250	260
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	SVCI SVCI SVSI SVCI SVCI)YLK)YLK)YLK)YLK)YLK	QQLNILRA QQLNILRA QQLNILRA QQLNILRA QQLNILRA	ADSQQ YANNN ADSQQ YSNNY ADSQQ YSGNY ADSQQ SSSNY ADSQQ ISSNY	NNNNNNNNNS MNYYDLHNEY VSYLELHNDY VSYLELHNDY Log.#11N*y	NNNNNNNNN EEAIINKKL EDILNEKKG EDIIHDKKG ##.in#kk.	NNNNNNNNNN YYNNSNFDN TIYTGSNGIC ATTTASNSMC	INNNNNYLNYY INNKYG Qgntns Qgntns Ignmns		MSDKKFLGINN NG NN NN Nr	IFTHYNLSNN IANDTNYNNN IINSHLSIKG ILNSQLSMKG	ANGNYNNITN FNLANYMG SSIHINSANT SSIHMNSANS .nnn	HANANANPTT SSKDNTNLSII TSNYSGNYNGI TSNYSGNATGI .sn.n.N.t.I	SHSNIKGNSH IGKYINQNNF IANGHISYNTI IASGHISINA 1.s.inn.	lsnann Qnsggn Eggnlg Eggnlg
	261		270	280	290	300	310	320	330	340	350	360	370	380	390
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	AMNY TNNY SFTL CFTL	/LYS /C-D .GDH .GDH	DQDNYYDD QNENHYDD Edsniyed Edsniyed ##.N.y#D	YYDCLMRTKL YYDCLMRTKL YYDCLMRTKL YYDCLMRTNL YYDCLMRTKL	TEETPNSLKY PNESYNNLKY SDDTPNCLKY SDDSPNCLKY .##spN.LKY	MINLKNETLS MMNSKQDNPI MMNSKTDSLS MMNSKNDSLS MMNSKNDSLS	SNAEFESVLSY NAADVDNVLSY INAEVENVISY SNTEVENVISY SNTEVENVISY SNa#v#nVISY	/LKKYENKSSK /LKKYELKSNK /LKKYEVKSNK /LKKYEVKANK /LKKYE•KSNK	GMNKNHLED	DNGMYKYNNNN	INNNNSNNNN NNT	NNNNNISNNI NKNDDVSKY- NYCSLEEELI NYCSLEEELI	SNNKDCDEYD' -NNKESD-YD' KYNRDSD-FE' KYSKDSD-FE' .nnk#sD.%#'	(QEEYLKDKN (QEEYLKDKN (QEDYLKDKT (QEDYLKDKT (QE#YLKDK.)	LYDSDH LYDSDH LYDSDL LYDSDL LYDSDSD LYDSD\$
	391		400	410	420	430	440	450	460	470	480	490	500	510	520
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	DEN DEN DDN DDN DHN	INQL	HNNEHHTN	0HHANYHHHK	HQNQHLKQLI			/LINNMDNNNI NMMDQNFh NLFDSNLI NLFDSNLI N.mD.N.i	DANN <mark>N</mark> NNNN DGTY <mark>N</mark> NGYN SSNRNMNDS SSSRNMNDS d.n.Nnn.n	NNNNNNINNLS GRGTRDYDNNG SSSMVNLNLSS GSSMMNMNLSJ 8#*n.s	HLI NNL NNNTHTNNN PNSG	HNNMNNNSTL Igmnsnntni Nnnnnsnnl Sannnsnnl • Dodo ⁿ d • Di	INRNHLIKSEI NNKNYNIKLEI NSKSYNHKYD SGKNYNHKYD NNKNYNIK,#	RIKKGDYG-T KIKKNDTM GLKKNNSDYM GMKKNNSDYM ,iKKn#M	TTHTKS NSHNKP NTHARA NTHTRA ntHtk.
	521		530	540	550	560	570	580	590	600	610	620	630	640	650
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	TTEO NTEO CTEO CTEO	ihpe ihpe ihpe ihpe ihpe ihpe	LPRIPGV LPRIPGV LPRIPGV LPRIPGV LPRIPGV	RFNPKKQQHL RFNPKKQQHL RFNPKKQQHL RFNPKKQQHL RFNPKKQQHL	AAHNDNTREI AAHNDNTREI AAHNDNTREI AAHNDNTREI AAHNDNTREI AAHNDNTREI	RRYFSVKQY(RKYFSVKQY(RRYFSVKQY(RRYFSVKQY(RryFSVKQY(R~YFSVKQY(GFEQARILAYK GFEQARILAYK GFEQARILAYK GFEQARILAYK GFEQARILAYK	KARQEAEKAGA KARQEAEKAGA KARQEAEKAGA KARQEAEKAGA KARQEAEKAGA	RC KPMFHVH RC KPMFHVH RC KPMFHVH RC KPMFHVH RC KPMFHVH	GSRKAVDAAI G-RKAIDG-IS G-RKTMENA-1 G-RKTVDSA-f G_RKTVDSA-f G_RKay#.ait	NDLLRSEH- NELIKYSLI NDYHK NEAMK N#1.k	EENFNNHHI NTQHSSNEHDI NNNDHHI NNNDHGI	MNHNNNNNNN PNYNNGAPNG MEDGTFGNNN MEDGTMGNNN MEDGTMGNNN	INNQHHHQNN SNNAIMGNNC Iggn	NSNIHH NMIMAN
	651		660	670	680	690	700	710	720	AT-hool	(NL	5 750	760	770	780
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus		IMHN GAL	T 	HLGHNMSHNM NMGGGNMNTI HYGGSNYGGN -MGGSGYGGS hm <mark>G</mark> g.n	NHNYNNNNN NGMPNIGGMN MGMMNNAGSN MGMMSGSGNN Ngm.nn.gn	HYINNNINN TLNGINQI AAHIGTI ASHAGTI	NNNNNNNSS NGMNMGNMNG NAAMNSGGMST	SNNNNNNNNNG Sennnamlgma Ggthehin Ggthehmn	NDVI LVTAPNDIL	HRI ETYKYEI TKI DLLKSET -KI ELLKNEN -KI DHLKNEN .KI #11K.E.	KGYKRGRGR KGIKRGRGR KGYKRGRGR KGIKRGRGR KGIKRGRGR	PPKRK PPKRK -SEDSI PPKRK -SENSI PPKRK -SEDSI PPKRK -SE#S	QMLLDDMEQTI NLSLDETEQS QLSLDDMEQTI QLSLDDMEQTI #\$slD#mEQt;	_CRNN [RRGYITGNN] _CRN _CRN LcRn	 NGSFND GE GE n#
	781		790	800	810	820 820	D2 830	840	850	860	870	880	890	900	910
PfAP2-I PbAP2-I PkAP2-I	NHEL NGEL NAEL	LEC LEC MEA	HERYDKDC Hdpydkdc Hdsfdksc	RPHKGVSYN RPHKGVSYN RPHKGVSYN	DRKGSHLAYH DRKGSHLAYH DRKGSHLAYH	SIGKNFQHR SIGKIFQHR SIGKNFQHR	RFPIKKLGFEN RFPIKKLGFEN RFPIKKLGFEN	KAKELAIQCRL KAKELAIQCRL KAKELAIQCRL	EAEQAGATT EAEQAGATT EAEQAGATT	ENRTKRMRNL Enrnkrgrnl Enrtkrirnl	LTLNSENAL LNNNSNNNL LNNNSENAV	EMMIDONSIDI DSMLDNSSMLI Egmmemnsldi	NDDSNIHQHMI DHDDNMNDNTI PDDNLI	(EKGGGMNGN Igankgmnrn Idtnknlngn	LRHSRH HLLPHH HIMPQH
PvAP2-I Consensus	N. #L	\$Ec	n#.%DKdC	RPHKGVSTN	DRKGSHLAYH	ISIGKNFQMRI	REPIKKLGEEN	KAKELAIQCRL	EAEQAGATT		LINNSENIN Ln.NS#Nal	*.H.*.nS.d.	יבטאנו dD.nח.ו	101MKNANGN 1nkg\$NgN	п1п5ųп 8.\$p.М
	911 		920	930	940	950	960	970	980	990	1000	1010 2-1	1020	1030	1040
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	QNYF HNFF HY-1 HHH1 hn	1-GA INGQ (NAH (NAH NAH	GI GLGSNVTN GM GI Gi	SGQSYHQTNK Sganpyqhhg NNSN-Mhpnk NNNNMhpnk Sg.nq.nk	LLNS MLGAIGGGYH M M \$1	INKIPHSDGQU INKIPHSDGQU INKIPHSDGQU INKIPHSDGQU INKIPHSDGQU INKIPHSdgQU	DSENEFSPTKF DSENDFSPTKF SENDYSPTKF SENDYSPTKF #SEN#%SPTKF	RTRAPRGRRME TRAPRGRRME TRAPRGRRME TRAPRGRRME TRAPRGRRME	SL TARASIL SL TARANIL SL TARASIL SL TARASIL SL TARASIL	TPVEGVRFDPY TPVEGVRFDPY TPVEGVRFDPY TPVEGVRFDPY TPVEGVRFDPY	SYSHFAKYL SYSHFAKYL SYSHFAKYL SYSHFAKYL SYSHFAKYL	ENENSKEPKI ENENSKEPKI ENENSKEPKI ENENSKEPKI ENENSKEPKI	SKYLLKKHGFI GKYLLKKHGFI SKYLLKKHGFI SKYLLKKHGFI SKYLLKKHGFI	ikahslayht ikahslayht ikahslayht ikahslayht ikahslayht	VKCAYK VKCAYK VKCAYK VKCAYK VKCAYK
	1041	1	050 +	1060	1070	1080	1090	1100	1110	1120	1130	1140	1150	1160	1170
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	AVPE AVPE AVPE AVPE AVPE	T DE T DE T DE T DE T DE	ELYNIFNY ELLNIFNI ELINIFNY ELINIFNY EL.NIFN!	DYNNLINNQN DAKNMMNNGN DSKNLGNNQN DSKNLMNNQN D•kN\$nNNqN	NHINLGFHDG NQ SMMNLPFKDS SLMNLAFKDT nm.nl.f.d.	FGTNGLMKT NGMINI YGNNSGANG YGNNAGANG •g•Ng••n•	ULNNQHTNQHT FKDNFAN AYND-NNYGG AYNA-NNYGF	INQHTNQHTNQ GAVENNAPHF MGYTNSANGA MGYSSSANGA	MINQHINQH ASNNFGN LIGNMGY LMGNMGY n#mgn		INTNQYNNQH NNLM MNGNI NNGAH	TNQHTNQHNN NNMYMINGI GGSYGMDH' GGAYGMDH' .n.vm#.	QMKNQYNSHII ALGNNPN YYNGMGII YYGGMPIGGY gn.pni	ISQYAGHANA MDM IGYGGGNFAQ GYSGGNFAQ 1g.h.l	NHFNGN NHHAGN NHGAYK SHGAYK nH _• agn
	1171	L 1	180	1190	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	NSNO SGN- GGVI GGVI	ITNH ITHG IAHG	LGNMNMLN ISNITNIS YMNNMGNN IMNSMGNS i.Nn	NMYNSNMDNN NPINNNFPGN NTEKNYDMSN GGEKNADMGN nnnngN	YNNNYSNNYP IGAFKNYTTP MNEEESTIGN MNDEETTIGN .nsp	GSGNNGLNNI GYGNNGMYYY GSSNNA-NNI GSSNNA-NNI GSSNNA-NNI	1TSCVHNNNGG /PPGM——NNLF .PISMDNANTL 1LSGNDGANGP 1PSgn.nnNg.	GEIDMDDNDNI AelnaI Lunydggnni Alnsyegannh .e.n.dni	ENNDHENMM NGLDPNGVI NSANKIDIM NAASKMGIM #dg.m	TLEKEDSINNE NAPNDNKIMDF NNCTYNYNGGN NGCAYNYSGGS n#.i	SIITTTAT S NAINEG SAYNDSAGG s.i	NINSCNNNSN NINGKDANLEI TA AYAGAGAATYI ninganI	NANSLLNRLYI NDNTGMDGRN: GTGSQNNYHHI GAAGQNNFHHI nans #	ISNGS <mark>N</mark> GNTKI [GNSMNNN]GLLTNNGII]GLLPNSGIV gnNnn	LNIINN INQYNK YGASLG YGAGLG .nn.
	1301	L 1	310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	YNNN INGY DNN\ ENN\ .Nn.	INNN (NFN /NGP /NGP .N.n	NSSSGGGV NLHTGGIG SSIKDNTT SSNKGNAT nsgg	MYNINNYHNN NIIPNN ISNGYNMYHN IANGYNYYMN n.nny.nN	DNMNTYNNNN NNYNGHSDNI Eqrglntnyn Eqmglganys ##md#dd	INNILNDGNS INNYAQNI IGKSFSTGEVI IGKSFPSGEVI INN8#	/VGNNESMIHk IDGHNISGEKI _VAPGG <mark>S</mark> STML _VAPGT <mark>S</mark> SAMI .vg.n. <mark>S</mark> r	KKNYYNCINNO INSYTNFLIDO EDENSKMLNN IDDENPKYLNK 1v.n#d	VEDDLGNMN NKNSLEHGG AHLMDNSNS AHLMESSNG	NNHNNNNNN INHYANNNNS TGNNSSGSYII AGNNAGGSYLC	NNNNNNNN NNYNSGNYI NTSYDNYIM NSSYDNYLY Nnsnanna	SNNNNNNNN NSEFLDSINRI GSDCVLNNNN VSECVLNNNN •s#•••nnNn	LNYFMNEHNNI KNNSNTNNNYI RKMSAPHYGNI RKYGGAHYGNI .nysni	PTNNRGIYGS (TPNDGIRNS) PTIGGGPSP- (TIGGAAAAA PT.n.gis	NESINN KDTNNE FSHH AAFGHH n.
	1431 	1	440 +	1450 +	1460 +	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	NGL1 N H N	1DEG	RLNFIHTA MDGPA MEDPA	GKHYDKNNHQ DKIN GYGYYRHN YYGYYRLN 8vdk.N	SGSTT¥¥N¥L NNL EMH EMH 	NNDENYIKSI NNGINYYPNI IMNGENGINTI IMNAENDINAI NNgeNy!	INYTNMDNMK INNSYYO IPANA IQSNT I.n.t	KNNDNPIDDDN GQMDNELTSHN QYGITSTGSN QYGIASTGSN ,#.dntN	PNNIYEMNL MNTISPSNH MNINSEPYH MNMNAEPYH MN,ise,n\$	RVEKNILNNNN TNDr 	IDEHIINNTE INRNYLNQTY NKDMMNSAS SKAMYNSAS nN.t.	DHMSTNNKEN GIQNNSNSHGI GSYNLG GSYNLG gnn.g	CTLNGQSTILI NHKSDHINNLI TNNNGNNSSI TNKD-STGSI kng1	1NSNEENDETI 1DYINNNDNN LEGRDYIDNTI 1EGRDYLDNSI \$##.nD#t	ALDTII TNITEL RYYEHP RGAEHP t
	1561 	1	570 +	1580 +	1590 +	1600	1610	1620	1630	1640 16	47 I				
PfAP2-I PbAP2-I PkAP2-I PvAP2-I Consensus	QPP] QT LEQF LEQF 9•••	EKR -RTN RKFY RKFY	NYYNMMIN NYIRDRTN NIMNDAGN NIMNESGN NndN	EQGTADNNSN ITNNTDYGNN IDSGYNS IDSGYSGYSG idsn	VDSSNNSYSN Megennn-Tn Enganntyan Dsganssyan ••g• ^N nsy•N	INV-YTKNLNS INS-YDQNGGI INSPYGPNAGI INNPYGPNAGI INs.YN.g.	SNYNGFNNENS RNFNGYGNEDO INFNSYEGENF NFNSYDGENF .N%Ng%.nE#.	SNTTI SNGNI GTYNNIGSNA GGAANNIGSNA .nt	NEQNDSSLY NDPSIY NELNESSMY NELNESSMY Ne.n#sS.Y	FMNVNSEIQTE YMNSKSEIKTE YMNVKPEIKTE YMSVKPEIKTE %MnvksEIkTE	HL Q Q				







В





- *mtrap* ATAATAATAAAAAACTATTCTTTTATTTAT
- *myoA* TTTTTTTTTTTTTTA<mark>ACAACT</mark>AAATTAGAAGTGTG
- gap40 TGGATACATGAACAACACTTTTACTTTTCATT
- *imc1a* AAATATAATTAATTAATTAATTCATG
- *rh2a* TGTAAATAATA**ACAACT**ATTCGGAATTAAAC
- rh4 CACTATGTTATTACAACCTCATCTTTATGGAAT
- ron3 CATCTGAATTGAACAACGTTGTCTATATGTAAT

