

## Supplementary material for Book Appendix Table IV

### The Evolutionary history of Proteins Involved in Prereplication Complex Assembly

Lakshminarayan M. Iyer and L. Aravind.

#### Multiple sequence alignment of ORC2 proteins

		Helix-1	Str-1	Hel-2	Str-2	Hel-3	Hel-4	
Secondary Structure		.HHHHHHH	.EEEE	.HHHHHHH	.EEEE	.HHHHHHHHHH	.HHHHHHHHH	
150.t00006_Ehis_67471552	57	NFKNIPPILVVHNLNLYGV	CKRDFFIKVNRFEKSR	---	I IKPVKPECGLL	SENECANVLFKIVNRAYNIFNQ	PDFSNKTSDDVQSLNLSLT	----
TA19765_Tann_65305665	266	MIFWKTWILNGFHLIFYCK	SNLNLLNAF	---	SKIALRFDTKTL	SKSEALKRTSRKVESLSNSY	-----	AKTINQ
ECU10_1130_Ecun_19069364	12	CKKELNVLARSFNILFYCK	SKTALLRKM	---	FPSATYLNCKIMS	SRREILAEIMEAVRHRSKL	-----	EAQEVSKAPTIREI
cgd4_1930_Cpar_66357218	56	HEKLLTWSLSGFSVLLYCF	SKRNFLDF	FVKKINGNY	---	VALTIRGYFKNIKFKTCLFELLKAMDHTED	FLKDKSTKSYNNFISNDCSIDSMIAKI	---
orc-2_Cele_17534571	148	FGKWKLYLAAGFNILLHGV	SKRDVLT	TEFENE	---	LYSDY---	TYMRVDARKDGLNVKVLVGA	INENMKLNCNV
AN7723_2_Anid_67901472	244	FPQWNFEFSQGFNICYG	SKRRLTQ	GFADWL	YRRHSSAPSVV	VVNGHTPNLS	IRSIFATIVTAVLGADIP	-----
orc2_Spom_4538672	247	FHQWYFELVNNFNLLFYCF	SKEHFLSS	FVEKLLPCF	---	PIFVVKGYFPQLQLKNVLSL	LEFLEVTPAA	-----
ORC2_Scer_6319534	301	FPQYWFELTQGFSLFYGV	SKRNFL	EEFAIDY	LSPKI	---	20-PCLILNGYNPSCNYR	VDVFKETD
DDB0190652_Ddis_66826059	85	FNKIYTDLKFVSVLVSCF	SKIQLI	ETVKE	FCTDG	---	PSLEHF	KGYLPNLSVRD
At2g37560_Atha_42571101	72	YSKWWFELRCGFLLMYCF	SKKALV	EDFASAS	LTDY	---	SVVINGYLPSSVNLK	QSSGSLSKGQETFPS
Orc2_Dmel_7299828	340	FPKWMCILNEGFNILLYGL	SKHQLL	QSFHRE	VLHKQ	---	TVLVVNGFPFSLTIK	DMLDSTSD
ORC2L_Hsap_5453830	298	FHKWMLQLHLGFNIVLYGL	SKDLLE	RFTTML	QDS	---	IHVING	FPFGISVKS
ORC2p_Xlae_7512175	279	FYKWMQLHLGFNILLFGL	SKQSL	LEKERT	SLLQDS	---	LHVIN	EFFPSSITAKS
Consensus/95%		.b.....sa.lh..G.Gsp..hh..h.....						
		Str-3	Hel-5	Str-4	Hel-6	Str-5	Hel-7	
150.t00006_Ehis_67471552		RPFI	FMLNDI	DPFPL	LNNQYPLHFLPRI	SSLPN	CRFILNT	IDG
TA19765_Tann_65305665		---	FYMI	IHGID	DLFLNE	---	GFNTL	KD
ECU10_1130_Ecun_19069364		EKY	KLVMANF	DFGMLE	-----	FSSLR	NFAIT	GTVE
cgd4_1930_Cpar_66357218		KNIF	LI	IHSID	SISIRQ	---	YLP	PAISQ
orc-2_Cele_17534571		QQL	LILIIDNIE	EAPD	WRSDQ	---	EAF	CEL
AN7723_2_Anid_67901472		DPIT	VFINS	IDAP	SLRRAT	---	NQ	ALLAR
orc2_Spom_4538672		GKIV	FLVHNID	GESL	IDER	---	FQ	SALAA
ORC2_Scer_6319534		IKLIL	VVHNLD	GPS	IRKNT	---	FQ	TMSF
DDB0190652_Ddis_66826059		DHY	VVVIHND	DG	YSLRNET	---	SQ	LTLAL
At2g37560_Atha_42571101		CFIC	VVHNID	GP	ALRDPE	---	SQ	TLAR
orc2_Dmel_7299828		THL	FLIVHNLD	GAM	LRNVK	---	AQ	AILSR
ORC2L_Hsap_5453830		LEL	FLLIHNLD	SQ	MRGEK	---	SQ	IIGQ
ORC2p_Xlae_7512175		LEL	YLLIHNLD	SQ	MRGEK	---	CQ	VLGQ
Consensus/95%		.h.hhh.sh.....b.....						
		Hel-8	Hel-9	Hel-10	Str-6	Str-7	Hel-11	
150.t00006_Ehis_67471552		IF	KKLST	FQ	KNICIT	ILV	DTG	TSES
TA19765_Tann_65305665		IL	EALSR	NHQ	KLPS	LI	AQ	I
ECU10_1130_Ecun_19069364		VVR	NVPR	NR	SRLV	KE	L	LS
cgd4_1930_Cpar_66357218		IL	NCLT	PS	HQV	VNT	I	A
orc-2_Cele_17534571		FW	KSLAV	NS	QKLR	FL	FF	Q
AN7723_2_Anid_67901472		VL	KSLP	EN	AQNLY	RV	LL	T
orc2_Spom_4538672		VL	YSLP	ANS	RGIF	K	FL	L
ORC2_Scer_6319534		VL	QSLT	V	NS	K	MY	K
DDB0190652_Ddis_66826059		VL	KSLTE	I	STD	IF	K	EL
At2g37560_Atha_42571101		VL	QSLT	P	NG	Q	N	V
Orc2_Dmel_7299828		VF	SSLT	NS	RG	I	Y	M
ORC2L_Hsap_5453830		VL	RSLT	P	N	A	R	G
ORC2p_Xlae_7512175		VL	RSLT	P	N	A	R	G
Consensus/95%		hh..ls.....hh..hh.....						

**Figure Legend:** Proteins are denoted by their gene names, species abbreviations and gi numbers separated by underscores. Residues are colored according to their conservation profile at 95% consensus. The coloring scheme and consensus abbreviations are as follows: h, hydrophobic (h: ACFILMVWY) l, aliphatic (l: LIV) and a, aromatic (a: FWY) residues shaded yellow; b, big (LIYERFQKMW) residues shaded grey, s, small (AGSVCDN) residues colored green; and p, polar (STEDKRNQHC) residues colored blue; and c, charged (HRKDE) and -, acidic (DE) residues colored magenta. The predicted secondary structure is shown above the alignment, where E represents a strand and H, helix. Poorly conserved inserts are replaced by a number that denotes the size of the insert. Note the conserved glycine rich loop between strand-1 and helix-2. Species abbreviations are as follows: Anid : *Aspergillus nidulans*; Atha : *Arabidopsis thaliana*; Cele : *Caenorhabditis elegans*; Cpar : *Cryptosporidium parvum*; Ddis : *Dictyostelium discoideum*; Dmel : *Drosophila melanogaster*; Ecun : *Encephalitozoon cuniculi*; Ehis : *Entamoeba histolytica*; Hsap : *Homo sapiens*; Scer : *Saccharomyces cerevisiae*; Spom : *Schizosaccharomyces pombe*; Tann : *Theileria annulata*; Xlae : *Xenopus laevis*.