Research article

Classification and evolutionary history of the single-strand annealing proteins, RecT, Red β , ERF and RAD52

Lakshminarayan M Iyer, Eugene V Koonin and L Aravind*

Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

E-mail: Lakshminarayan M Iyer - lakshmin@ncbi.nlm.nih.gov; Eugene V Koonin - koonin@ncbi.nlm.nih.gov;

L Aravind* - aravind@ncbi.nlm.nih.gov

*Corresponding author

Published: 21 March 2002

BMC Genomics 2002, 3:8

Received: 7 January 2002 Accepted: 21 March 2002

This article is available from: http://www.biomedcentral.com/1471-2164/3/8

© 2002 lyer et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The DNA single-strand annealing proteins (SSAPs), such as RecT, Red β , ERF and Rad52, function in RecA-dependent and RecA-independent DNA recombination pathways. Recently, they have been shown to form similar helical quaternary superstructures. However, despite the functional similarities between these diverse SSAPs, their actual evolutionary affinities are poorly understood.

Results: Using sensitive computational sequence analysis, we show that the RecT and Red β proteins, along with several other bacterial proteins, form a distinct superfamily. The ERF and Rad52 families show no direct evolutionary relationship to these proteins and define novel superfamilies of their own. We identify several previously unknown members of each of these superfamilies and also report, for the first time, bacterial and viral homologs of Rad52. Additionally, we predict the presence of aberrant HhH modules in RAD52 that are likely to be involved in DNA-binding. Using the contextual information obtained from the analysis of gene neighborhoods, we provide evidence of the interaction of the bacterial members of each of these SSAP superfamilies with a similar set of DNA repair/recombination protein. These include different nucleases or Holliday junction resolvases, the ABC ATPase SbcC and the single-strand-binding protein. We also present evidence of independent assembly of some of the predicted operons encoding SSAPs and *in situ* displacement of functionally similar genes.

Conclusions: There are three evolutionarily distinct superfamilies of SSAPs, namely the RecT/ Red β , ERF, and RAD52, that have different sequence conservation patterns and predicted folds. All these SSAPs appear to be primarily of bacteriophage origin and have been acquired by numerous phylogenetically distant cellular genomes. They generally occur in predicted operons encoding one or more of a set of conserved DNA recombination proteins that appear to be the principal functional partners of the SSAPs.

Introduction

Homologous DNA recombination is a fundamental process in the biochemistry of DNA repair and replication, which contributes to the generation of the genetic diversity critical for natural selection. An important step in the recombination process is the pairing of homologous double-stranded DNAs followed by the exchange of DNA strands between the paired molecules. Experimental studies have shown that members of the archetypal RecA family of recombinases are central to this reaction in all extant forms of life [1,2].

Studies in Escherichia coli have shown that, although RecA is the principle protein involved in pairing and strand exchange, unrelated proteins, that have a much more restrictive phyletic distribution, can also promote similar reactions in a RecA dependent or RecA-independent manner [3]. These alternative or additional mediators of homologous recombination include the well-characterized prophage RecT, phage λ Red β and phage P22 ERF proteins [4,5]. Similarly, in yeast and vertebrates, the RAD52 protein is involved in the pairing and strand exchange reaction and can promote recombination in a RAD51 (the eukaryotic RecA homolog)-dependent or independent manner [6]. The RecT protein works in conjunction with the RecE-nuclease [7] and was initially described in genetic studies on the complemention of mutations in the RecBCD pathway of DNA repair [8–10]. Biochemically, RecT has been shown to bind single-stranded (ss) DNA 3' overhang regions generated by the RecE nuclease, and promote strand exchange between homologous DNA partners by assisting the pairing of complementary singlestranded regions [4,10]. The reaction catalyzed by the RecT/RecE system is similar to that described for the phage λ exonuclease (exo/Red β) and the single-strand annealing protein Red β . The similarity between these two systems is further extended by the observation that the RecT/E system can complement mutations in the $\lambda \exp(\text{Red}\beta)$ system [10,12]. In eukaryotes, RAD52 protein has been shown to exhibit properties similar to those of RecT and Redß proteins: it binds ssDNA and promotes strand exchange via the pairing of complementary single strands [6,13]. In vitro studies on quaternary structures have shown that the single strand annealing proteins (SSAPs), RecT, Red β , ERF and RAD52, form similar helical super-structures [14-17]. This has led to the proposal that RecT, Red β , ERF and the eukaryotic RAD52 function in an analogous fashion, and even are "structural homologs" [14].

However, no sequence or secondary structural similarities have been noticed between different SSAPs and current understanding of their evolutionary history and phyletic range remains poor. Here, we describe the results of an indepth sequence analysis of these proteins and delineate their evolutionary relationships and phyletic horizon in available genomes. We show that, in spite of the functional similarities, and the similar quaternary structures, there are three distinct superfamilies of SSAPs, namely the RecT/ Red β , RAD52 and ERF, that appear to be evolutionarily unrelated to each other. These superfamilies show a wide distribution in viral and cellular genomes, but appear to have originally evolved in large DNA bacteriophages. Through an analysis of the contextual information provided by the predicted operons, in which the SSAPs occur, we predict several previously undetected functional connections of these proteins, which might shed new light on the corresponding DNA repair/recombination pathways.

Results

RecT and Red β are evolutionarily related and define a widespread family of DNA recombination proteins

Several lines of evidence, including genetic analyses, and similarities in biochemistry and quaternary structures, suggest that the *E. coli* RecT and phage λ Red β proteins are functionally equivalent as mediators of single-strand exchange in DNA recombination [10,12]. However, no sequence similarity has been detected between these proteins leaving their actual evolutionary relationships unresolved. In order to gain a better understanding of their functions and origins, we undertook a detailed sequence analysis of these two proteins using iterative sequence profile searches with the PSI-BLAST program with a inclusion threshold of .01 iterated until convergence. Such searches, with Redß proteins from different lambdoid bacteriophages as queries, retrieved not only other obvious Red β homologs, but also the RecT protein family. For example, searches initiated with the Red β homolog, PF161 protein (Genbank gi: 9836834 amino acids 1 to 188) from Borrelia hermsii[18], detected the E. coli RecT protein in the 5^{th} iteration with significant expectation (e) values (3×10^{-3}) . Subsequent iterations retrieved several more RecT-related proteins from diverse sources. Further, transitive searches with the proteins detected in the above searches resulted in the identification of more divergent homologs, such as a protein termed the 'enterohemolysin associated protein' (EHAP1) from E. coli[19] and its orthologs in Salmonella (Fig. 1). An examination of the pairwise alignments generated by these searches showed that all these proteins shared a characteristic set of residues, including two highly conserved aromatic residues at the Nand C-termini, respectively, and two consecutive acidic residues near the C-terminus. These observations strongly suggested that RecT and Red β , along with several other proteins, could be unified into a single protein superfamily with a core conserved domain of approximately 200 amino acid residues.

A multiple alignment of all members of the RecT/Red β superfamily was generated using the T_coffee program followed by adjustments based on the PSI-BLAST search results. This alignment was used to predict their secondary structure using the JPRED and PHD methods; these predictions pointed to an $\alpha + \beta$ domain with a core of five β -strands and five α -helices (Fig. 1). Some of the strongest conservation is concentrated in the long helices, and the pattern includes some charged or polar residues, suggest-

Noliz-1

Holix-2

Strand_2

	Hellx-1	NOTIN E	Scranu-1	beruna r		Strana S	
PHD Sec. Structure			EEEEE	EEEEE		EEEEEEEEE	
Red [lambda_137511	19 DSVDPQELITTLRQ	-TAFKGDASD-AQ <mark>F</mark> IAL <mark>L</mark> IVANQYG <mark>L</mark> NPW	FKEI <mark>YAF</mark> PDKQN	GIVPV <mark>V</mark>	GVDG <mark>W</mark> SRI <mark>I</mark> NENQQFD	-GMDFEQDNE	
Beta_Sd_6759957	19 DSVDPQELITTLRQ	-TAFKGDASD-AQ <mark>F</mark> IAL <mark>L</mark> IVANQYG <mark>L</mark> NPW	FKEI <mark>YAF</mark> PDKQN	GIVPV <mark>V</mark>	GVDG <mark>W</mark> SRI <mark>INENQQF</mark> D	-G <mark>M</mark> DFEQDNE	
lin1755_Li_16414242	24 GEEVKLSGNIIRDYL	-VSGNAEVTD-QE <mark>I</mark> IMF <mark>L</mark> QL <mark>C</mark> KYQK <mark>L</mark> NPFI	LNEA <mark>YLV</mark> KFKNTKGPDK	PAQII <mark>V</mark>	SKEA <mark>F</mark> MKR <mark>A</mark> ETHEQ <mark>Y</mark> D	-G <mark>F</mark> EAGIIVERNGEVVEIEGAVS	
ml17977_M1_13476603	7 YTHSPRQLALIQKT	VAKDCNT-DE <mark>F</mark> NLF <mark>V</mark> EV <mark>A</mark> RAKG <mark>L</mark> DPF·	LGQ <mark>II</mark> PMIFSKGDSNKR-	KMTII <mark>I</mark> SRDGÇ	QRVIAQRCGDYRP <mark>A</mark> SKPPS <mark>Y</mark> E	<mark>F</mark> DAELKSETNPQGIVSATVYL	
SPy0958 Strpy 13622119	15 TTDPTLLTGADIKKY	-FDPQNLLSE-KQ <mark>V</mark> GQA <mark>L</mark> AL <mark>C</mark> KGRN <mark>L</mark> NPF	ANEV <mark>YIV</mark> AYKNNSGT	DFSLIV	SKEA <mark>F</mark> MKRAERCEG <mark>Y</mark> D	-G <mark>F</mark> EAGITVMRNGEMVEIEGSLK	
BPphi31.1 Orf245 7239201	7 IFSVDNLNMTTIKQY	-LDGGGKASD-AELVLL <mark>I</mark> NL <mark>C</mark> KQNN <mark>M</mark> NPFI	MKEV <mark>YFI</mark> KYGNQ	PAQIVV	SRDF <mark>Y</mark> RKRAFONPN <mark>F</mark> V	-G <mark>I</mark> EVGVIVLNKDGVLEHNEGTFK	
XF1648 Xf 9106705	15 MTAEYOOSIRTALK	-TSLYPGASD-TSVDMVLSYCOAADLDPM	rkpv <mark>hiv</mark> pmwipekkvdgrv	VSSAGMRDVIM	PGIELYRTKAHRTGEYA	-GODEAVFGDTLCETLGGVOIR	
UU154 Uu 14195377	29 EINOI-TRAVLTIO	GIDLKAIDLNOAAOITYFCOANNLNPL	NKEVYLTOMGN	RLAPIV	GIHTMTERAYRTERI.V	GTVOSYNDVNKSA	
PF161 Borhercp 9836834	1 -NSSNIVEVWEAVKS	MHGLKSMDTOSEREILTLLOVNNLNPF	KKEAYTTPFNG	RYAVVV	AYOTLLIRAY RAGYSKY	-SLEFKEEMVKTIKIDSKGNKMV	
Room Fo 16129310	52 AFPMIDIATTEIDE		- AL CHAVLL PRONUNEVSCUV		CYRCMIDI ARRCOTA	CI CARWAREOPEREFECT DEV_I THROCE	
1;n2412 1; 16901475	49 ABORT TELL NI VNC	- DDVI OKTOPMTUUTS AMUA ATI DI - DI	S-ALGHATIBBFFGHKHEKSGKK-	RAOFOT			
ODE40 DD110 16709925	49 APOPLESLENDING	DDVI OV TOPMI VISAMVAAIDDD-FI	DENILGIANI VIIKG	KAQIQD		TWITEVROOFILL KINDLEETELDLON	
ORF48_BPA118_10/98835	49 APOPLISLENLING	DDILOKTOPMIVVISAMVAATLOL-PI	DKNLGIAWIVPIKG	RAOFOL	GIKGIIQLALKIGQIK-	-SINVIEVROGELLKWNRLIEEIELDLDN	
BH3543_Bh_15616105	51 APOFMISIINLISN	DSGLQKCDPMTVISSAMVAASLDL-PI	DKNLGIAWIVPIIDRKTKSI	RAQFQL	GYRGYIQLALRSGQYR-	-YINAIPVRKGELIKWDPLTEEIEIDFEA	
Ydak_B5_160/9681	48 ATOFTASILSLYNS	EQMLQRTDPMSVISSAMVAATLDL-PI	DKNLGIAWIVPIGG	KAQFQL	GYRGYIQLALETGQYR-	-SINCIPIHEGELQKWNPLTEEIEIDFEK	
RecT_BPb1L309_13095817	29 TEGFVASLLSVV-G	NSNLKNADANS <mark>V</mark> MTAAMKAATLDL-PII	EPSLGFA <mark>YVI</mark> PYGR	EAQFQ <mark>I</mark>	GYKG <mark>F</mark> IQLALRSGQLT-	-GLNCGIVYESQFVSYDPLFEELELDFSQ	
ORFC_Lp_13186144	40 PERMARIAMTELRK	TPKLQECDPLS <mark>F</mark> IAS <mark>I</mark> MQ <mark>A</mark> AQLG <mark>L</mark> EPG	ILGSC <mark>YLI</mark> PFWNSKLGKF	ECTFMP	GYRG <mark>F</mark> LDL <mark>A</mark> RRSGQ <mark>I</mark> V	-S <mark>L</mark> VARSVYENDEFSYEFGLKEN-IIHKPAM-	
gene35_SPP1_540750	52 ADRLSRIAMNVIRT	NPKLLECDTAS <mark>L</mark> MGA <mark>V</mark> LESAKLG <mark>V</mark> EPGI	LLGQA <mark>YIL</mark> PYTNYKKKTV	EAQFI <mark>L</mark>	GYKG <mark>L</mark> LDL <mark>V</mark> RRSGH <mark>V</mark> S	-T <mark>I</mark> SAQTVYKNDTFEYEYGLDDK-LVHRPAPF(G
orf43_BPPVL_9635208	45 PSNAMKQAWLQISQ	DNKLMSCNDTSKANA <mark>L</mark> LD <mark>M</mark> VTQG <mark>L</mark> NPAI	KNQC <mark>YFI</mark> PYGN	KMQLQR	SYHGNVMMLKRDAG <mark>A</mark> Q	-D <mark>V</mark> VAQVIYKGDTFKQEMGETGR-IKAIKHEQ!	DFFN
SA1794_SaN315_15927560	37 PENAMKSAMLQLQELKGSKKDO	GYKPALEFATSTS <mark>I</mark> ANA <mark>L</mark> MD <mark>M</mark> VVQG <mark>L</mark> NPAI	KNQG <mark>YFI</mark> MYGD	KVQFQR	SYHGTMAVTKRVAG <mark>A</mark> E	-E <mark>I</mark> NAEVIFEGDEVKYKTKNGKI-VELEHTQS	FGN
ORF10_BPR1T_1353527	37 AEGALGYTALAI	VNSGFTVSKEV <mark>I</mark> VDT <mark>L</mark> IK <mark>V</mark> ASKG <mark>L</mark> DPRI	KDQL <mark>YVI</mark> PNKK	GQVML <mark>M</mark> E	SYFG <mark>Y</mark> EKL <mark>A</mark> YDIPE <mark>I</mark> ER(3S <mark>V</mark> FAEVVRQGETVSFQGRTLEHEKAFEAIDN	DIIGAYAKV
SPy1477 Strpy 15675383	31 AEQFTTSLLSIISN	NNLLAKATSES <mark>I</mark> MGA <mark>A</mark> MK <mark>A</mark> AVLN <mark>L</mark> -PII	EPSLGFA <mark>YVV</mark> PYNRNYKDG	NRWIT <mark>V</mark> NEAQE	7QIGYRG <mark>L</mark> IQL <mark>A</mark> QRSGQ <mark>V</mark> R	-N <mark>I</mark> EHGIIYEEEFLGYDKIRGQL-KLTGDY	
EHAP1 EC 484496	4 LVRFAELMSQSKATV	PKHLESKP-AD <mark>C</mark> LAVTMQ <mark>A</mark> AQWG <mark>M</mark> NP	LPVAQ	KTHV <mark>V</mark> N	GTLG <mark>Y</mark> EAQLVNAVVSSS	-SLATRLNYRWSGDW	
STY2074 Salent 16760818	25 IOTFSOVMASGMATV	PEHLRGNP-SDCMAITMOAMOWOMNP	YAVA0	KTFVVN	GVLG <mark>YEAOL</mark> VNAVISTR-	-GPLTGRIEYDWFGPWEKIIGKFEIRKNDK	
STM2633 StLT2 16765953	48 LTAFANLMADSOVTV	PAHLAGKP-ADCMAIVMOAMOWGMNP	<mark>YAV</mark> AO	KTHLVN	GVLGYEAOLVNAVIASS	-SAIHGRFHYRYGGDWERCTRTOEITRDKNGK	NGKY
consensus/85%			ahh.b	h	ushbh.psh	.sh	
	Strand-4	Strand-5		Helix-4			
PHD Sec Structure	Strand-4	Strand-5	нн	Helix-4			
PHD Sec. Structure	Strand-4	Strand-5 EEEEEE	НН	Helix-4			205
PHD Sec. Structure Red β _lambda_137511 Rota_54 6759677	Strand-4 EEEEEEEEEE SCTCRIYRKDRNHPJ	Strand-5 EEEEEE ICVTEWMDECRREPFKTREGRE	НН 	Helix-4 HHHHHHHHHHHHH RMLRHKAMIQCARI		VENTAYTAERQPERDITEVNDETMQEI	205
PHD Sec. Structure Redβ_lambda_137511 Beta_Sd_6759957	Strand-4 EEEEEEEEEE SCTCRIYRKDRNHPI SCTCRIYRKDRNHPI	Strand-5 EEEEEE		Helix-4 HHHHHHHHHHHHH RMLRHKAMIQ <mark>CA</mark> RI RMLRHKAMIQCARI	-AFGFAGIYDKDEAERI	VENTAYTAERQPERDITPVNDETMQEI VENTAYTTERQPERDITPVNEETMSEI DWN ETSUDERTVONMEN	205 205
PHD Sec. Structure Redβ_lambda_137511 Beta_Sd_6759957 lin1755_Li_16414242	Strand-4 SEEEEEEEEE SCTCRIYRKDRNHPJ SCTCRIYRKDRNHPJ LDKDKLLGGWAKVFRKDRSF	Strand-5 EEEEEEE ICVTEWMDECRREPFKTREGRE ICVTEWMDECRRAPFKTREGRE RVSVR <mark>I</mark> SEREFNKRQ	HH ITGP <mark>W</mark> QSHPK ST <mark>W</mark> NAMPL	Helix-4 IHHHHHHHHHHHH IRMLRHKAMIQ <mark>CA</mark> RI IRMLRHKAMIQCARI TMMRKTAVVNAMREA	-AFGFAGIYDKDEAERI -AFGFAGIYDKDEAERI -PDNLGAMYTESEQGSL	VENTAYTAERQPERDITPVNDETMQEI VENTAYTTERQPERDITPVNEETMSEI DNNETSVQEEIKQNANTEMLDIPAQ	205 205 226
PHD Sec. Structure Redβ lambda 137511 Beta_5d_6759957 lin1755_Li_16414242 ml17977_ML_13476603	Strand-4 SCTCRIYRKDRNHPJ SCTCRIYRKDRNHPJ LDKDKLLGGWAKVFRKDRSS WKQDAKTANFFVAGQSYNDE	Strand-5 EEEEEE ICVTEWMDBCRREPFKTREGRE RVUSVRISEREPIKRQ RVUSVRISEREPIKRQ 		Helix-4 IHHHHHHHHHHHHH IRMLRHKAMIQCARI IRMLRHKAMIQCARI ITMMRKTAVVNAMREA ILMIAKCAEMQALRAG	-AFGFAGIYDKDEAERI -AFGFAGIYDKDEAERI -PDNLGAMYTEEEQGSL -PEQFTGLYDEEEMDRA -PEQFTGUYDEEEMDRA	VENTAYTAERQPERDITPVNDETMQEI VENTAYTTERQPERDITPVNEETMSEI NNETSVQEEIKONANTENLLIPAQ KVLEMAASEIVAHEQEEN	205 205 226 237
PHD Sec. Structure Redβ lambda 137511 Beta_Sd_6759957 lin1755_Li_16414242 ml17977_ML_13476603 SPy0958_Strpy_13622119	Strand-4 EEEEEEEEEE SCTCRIYRKDRHPJ SCTCRIYRKDRNHPJ LDRCHLLGGWAKVFRKDRSF WKQDAKTAAFFEVAGSYHDE LPDDVLIGGWAIVYRKDRSHN	Strand-5 EEEEEEE- ICVTEWHDECRREPFKTREGRE ICVTEWHDECRRAPFKTREGRE RPVSVRJSEEPNKRQ PAPISYDAKWNDTGETWEDSGKPKKKR YKVTVD <mark>P</mark> NEYVKLDKYGN		Helix-4 IHHHHHHHHHHHHH IRMLRHKAMIQCARI IRMLRHKAMIQCARI IMMRKTAVVNAMRAA ILMIAKCAEMQALRAG ITMIRKTALVQTLREA	AFGFAGIYDKDEAERI AFGFAGIYDKDEAERI PDIGAMYTER2GSL PDGFLGLYDERMDRAI PDELGNMYTDIDGGDT	VENTAYTAERQPERDITPVNDETMQEI VENTAYTYERQPERDITPVNETMSEI DNNETSVQEEIKQNANTEMLDIPAQ KVLEMAASEIVAHEQEEN PDQEVRA	205 205 226 237 214
PHD Sec. Structure Red _b lambda 137511 Beta Sd 6759957 lin1755_ti 16414242 ml19977_ML_13476603 SPY0958_Strpy_13622119 BPphi31.1_0cf245_7239201	Strand-4 	Strand-5 -EEEEEEE CUTEMNDECRRAPPKTREGRE- CUTEMNDECRRAPPKTREGRE- PAPISVELSERENKRO- PAPISYUNGSERENKRO- KUTUDKEUKUNUTGETMEDSGKPKKKR KUTUDKEUKUNUTGETMEDSGKPKKKR	-HH -ITGPWQ=HPK TGPWQ=HPK ULRDGATPQLD-DSGNWC=MPR -PRSTWK=MPG -PNKMWTNKPC	Helix-4 IHHHHHHHHHHHH IRMLRHKAMIQCARI IRMLRHKAMIQCARI IRMLRHKAMIQCARRI IMMLRKTAVVNAMREA ILMIAKCAEMQALRAG ITMIRKTALVQTLREA ITMIRKTALVQTLREA ITMICKVAESQALRMA	-AFG <mark>P</mark> AGI <u>V</u> DKDEAERI -AFGPAGI <u>V</u> DKDEAERI -PDNLGAM <u>YTEEEQGSL</u> -PDELG	VENTAYTAERQPERDITPVNDETMQEI VENTAYTERQPERDITPVNETMSEI SUNETSVQEEIKQNANTEMLDIPAQ KVLEMAASEIVAHEQEEN VDTQEEVRA KEPREVNGKEPDRAQISFSPKE	205 205 226 237 214 209
PHD Sec. Structure Red ₀ lambda 137511 Beta_Sd 6759957 lin1755_ti_16414242 ml17977_ML_13476603 SPy0958_Strpy_13622119 BPphi31.1_0rf245_7239201 XF1648_Xf_9106705	Strand-4 EEEEEEEEE 	Strand-5 -EEEEEEE CUTEMU & CRREPFKTREGRE- CUTEMU & CRRAPFKTREGRE- PAVSVR.SEEENKRQ- PAVSVR.SEENKRQ- VAVSVDEVVLDRVAN- VYVAVSXDEVVQMKD-GH- PAVTVHLAVTARAKDSPA-		Helix-4 IHHHHHHHHHHHH IRMERHKAMIQCARI IRMERHKAMIQCARI IMMERKAVVNAMEBA ILMIARCAEMQALRAG ITMERKALVQTLREA ITMEGRVAESQALRMAI GQLEKCAEALALRKA	-AFG <mark>F</mark> AGI <u>V</u> DKDEAERI -AFG <mark>F</mark> AGI <u>V</u> DKDEAERI -PDRIGAMYTEEQGSL -PEQTGLYDEA'MDRAI -PDELGNMYTDIGGDT -PAE <mark>F</mark> SGT <mark>Y</mark> GEEEYPED -PEAVGAQTAEENDAG	VENTAITAERQPERDITPVNDETMQEI DENTAITERQPERDITPVNDETMQEI NMESTQUEEKQNANTEALDIPAQ CV	205 205 226 237 214 209 226
PHD Sec. Structure Red:_lambda_137511 Beta_Sd_6759957 lin1755_Li6414242 ml17977_M_3476603 SPy0958_Strpy_13622119 BPphi31.l_0rf245_7239201 XF1648_Xf_9106705 UU154_U_14195377	Strand-4 EEEEEEEEEE 	Strand-5 -EEEEEEE CUTEMNDECRRAPPKTREGRE- ICUTEMNDECRRAPPKTREGRE- PATSVDSISEERENKRO- PATSVDATKNDTGETVEDSGRPKKKR- VIVAVSTDPIEVKLDKIGN VIVAVSTDEVVQNKD-GH PATVIWLEATATARKDSPA	НН — ТТСРИО-НРК — ТТСРИО-НРК — STWNAMPL — STWNAMPL — PRSTMK ВИРО — PRSTMK ВИРО — PNKHMT: КРС — PNKHMT: КРС — PNKHMT: КРС — NKHMT: КРС — NKHMT: КРС — NKHMT: КРС	Helix-4 IHHHHHHHHHHHH IRMLRHKAMIQCARI IRMLRHKAMIQCARI ITMRKTAVVNAMEAI ILMIAKCAEMQALRAG ITMIKKTALVQTLEBA ITMIGKVAESQALRMAI GQLEKCAEALALRKAI ITMLKKVSLAHALRI	-AFGRAGIYDKD BARNI -AFGRAGIYDKD BARNI -PDNLGAMYTER'QGSL -PDELG	VENTAITAERQPERDITPVNDETMQEI VENTAITERQPERDITPVNDETMQEI NMETSVQEEKNANTENLDIPAQ VVLEMAASEIVAHEQEEN PDAHEQUENTQEEVRA EKEPREVNGVKEPPAQEGETHVADIEGETHVADIEGETHVADIEGETHVADIEVRQUVR NHTI	205 205 226 237 214 209 226 206
PHD Sec. Structure Redβ_lambda 137511 Beta 8d 675957 11n775 Li 6414242 ml17977 ML_13476603 SPy0958 6trpy 13622119 BPphi31.1_orf245 7239201 W1548 xf 306705 UU154 Ju_14195377 PF161_Bohercp_9836834	Strand-4 SEEEEEEEE 	Strand-5 -EEEEEEE CUTEMID & CRRAPPKTREGRE- CUTEMID & CRRAPPKTREGRE- AP ISY TO AYKAVDTGETWEDSGKPKKKK KVTVDN & YVLAVSK VYVAVSK DE YVQMKD-GH- AATVY KLAZATARKDSPA- VERGVFLE & YSTMK- VERGVFLE & YSTMK- SFSVL.FWEYKNS-	HH TTGP40_HPK STWNAMPL URDGATPQLD—DSGNNC MPR PRSTWL MPG PNKNWT INFO PNSNW0, RPP NLML INFT PT SR INFO	Helix-4 IHHHHHHHHHHHHH- IRMIRHKAMIQCAR IRMIRHKAMIQCAR ITMMRKTAVVNAMREA ILMIAKCAEMQALRAG ITMIRKTALVQTIREA ITMIKKALVQTIREA ITMIKKAVSLAHALRI ITMIKKKVSLAHALRI	-AFGFAGIVDKOBAERI -AFGFAGIVDKOBAERI -PDRIGAMTTEIGGSL -PDEGTCLUDERINDBA -PDEGG	VENTAITAERQPERDITPVNDETMQEI DENTAITERQPERDITPVNEETMSEI NMESTVQEEKONANTEKLOIPAQ VVLENMASEIVAHEQEEN DFD	205 205 226 237 214 209 226 206 188
PHD Soc. Structure RedB_lambda_137511 Betts 5d, 675957 lin1755_Li_16414242 nl.7977_M13476603 SPy0958_Strpy_13622119 Bephi31_0.0rf245_7239201 XP1648_Xf_9106705 UU154_U_14195377 PF161_Borhercp_9836834 ReeT_Ec.16129310	Strand-4 	Strand-5 = EBEELE - EUTENHD KCREAPFKTREGRE- IUTENHD SCREAPFKTREGRE- PPVSVH 5 SHEFNKRO- AT 15 XDATKWDTGETHEDSGRPKKKR KWTUDPNE VKLDRYGN- 7AATVWLEAYATARKOSPA- PAATVWLEAYATARKOSPA- STSFSUT-RESTRK- STSFSUT-RESTRK- STSFSUT-RESTRK-	HH ITGPWG HFR ITGPWG HFR STNAMU- VLRDGATQLD-DSGNWC HFR -PRSTW KMOG -PNSMGT KFC -PNSMG RFP -STWK KFV -FTWK KFV -NNGPW THWS	Helix-4 IHHHHHHHHHHHHHHH IRHKANIQCAI IKHRIKANIQCAI IKHRIKANIQCAI IKHRICANNAMEAI IKHIKIVAENAALAG IKHIKIVAENAALAIKAI IKHIKIVAENAALAIKAI IKHIKIVAENAALAIKAI IKHIKIVAENAALAIKAI	-AFGPAGIYDKN MAERI -AFGPAGIYDKN MAERI -PFDFLG	VENTAYTALEROPERDITFVM	205 205 226 237 214 209 226 206 188 267
PHD Sec. Structure Redβ_lambda 137511 Beta 8d 675957 11n1755_L16414242 ml17977_ML_13476603 SPy0586_Strpy_13622119 BPphi31.1_orf245_7233201 W1544_U_14195377 PF161_Bohrercp_9836834 RedT_Ec_16129310 11n2413_L16801475	Strand-4 Stepresesses	Strand-5 -EEEEEEE CUTEMU & CRRAPPKTREGRE- CUTEMU & CRRAPPKTREGRE- APISVIS & SERENKQ- PAVSVIS & SERENKQ- VAUSTO & VUDVENKO- VIVAUSTO & VUDVENKO- VIVAUSTO & VUDVENKO- VIVAUSTO & VUDVENKO- VEAUVIE & STINK- STOFEVUT RIVEL - SKAG- 	IIIOPNO HER IIOPNO HER IIOPNO HER STRINATE VLRDGATDD-DSGNKC MER PRSTKK MCC PRSTKK MCC PRSTKK MCC PRSTKC MCC PRSTKC V NLL KET TR KEY NNOPV HEE SDFGKK DD	Helix-4 HHHHHHHHHHHHHHHHH KIMFRIKANGCAI RMIRIKANGCAI IMMRIKANGCAI IMMRIKANGCAG TMMRIKANGCAG TMFRIKALAGA GOLEICARALALIKA GOLEICARALALIKA MIKIVASLALA FMKRICASSCLCIT-I BEMAKITAIRKEF()	- AFGFAGI 2DK) AERI - AFGFAGI 2DK) AERI - FPOH_G	VENTAITAERQPERDITPVNDETMQEI DENTAITERQPERDITPVNEETMSEI VMESTQUEEKQNANTEKLOIPAQ VVLENMASEIVAHEQUEEN DFD	205 205 226 237 214 209 226 206 188 267 253
PHD Soc. Structure Redl lambda 137511 Betts 26 d55957 lin1755_Li 16414242 ml17977_M_1_3476603 SPy0958_Strpy_13622119 Bephi31_0rf245_7239201 XF1648_Xf_9106705 UU154_U_14195377 PF161_Borbercp_9836834 Rest_Ec_16129310 lin2413_Li_16801475 ORF48_BPA118_16798835	Strand-4 	Strand-5 EBEEFE (UVTFMD)CCRRPFKTRGOR	HI TTOPNO HPA TTOPNO HPA TTOPNO HPA TTOPNO HPA TTOPNO HPA PRSTWK MPC PRSTWK MPC PRSTWK MPC PRSTWC TTOPNO PRSTWC TTO PRSTWC TTO PRSTWC PRSTWC TTO PRSTWC TTO PRS	Helix-4 HHHHHHHHHHHHHHHHHH KIKIRIKANIQCA:I KHIRIKANIQCA:I YHRITAVANICA LINTAYCARMOAL:AG TIMIGIVAESQAL:MAI GOLE:CREALAL:KAI TIMIGIVAESQAL:MAI GOLE:CREALAL:KAI TIMIKIVSLARAL: FMLRICAVSCLCTT- FMLRICAVSCLCTT- AMAKITVLRNML:K-		VENTATTARROPERDITPVA	205 205 237 214 209 226 206 188 267 253 253
PHD Sec. Structure Redβ_lambda 137511 Beta 8d 6759957 lin1755_Li6414242 ml17977_ML_13476603 SPy058_6rtpy_13622119 BPphill.lorf245_7233201 W154& T_9106705 UU154_U_14195377 PF161_Borhercp_9836834 RedT_Ec_16129310 lin2413_Li6801475 ORF48_BPA18_16579835	Strand-4 EEEEEEEEE- 	Strand-5 -EEEEEEE CUTEMUD ECRRAPPKTREGRE- CUTEMUD ECRRAPPKTREGRE- APISVIS SERENKRQ- PAVSVIS SERENKRQ- APISVIS VDAYKNUDYCOH- VIVAVSVDEYVUNKUDKYCOH- VIVAVSVDEYVUNKUDKYCOH- VIVAVSVDEYVUNKUDKYCOH- VEREVFLGEYSTINK- STSVLPUTYKINS- STSVLPUTYKINS- STSVLPUTYKINS- STSVLPUTYKINS- SKAG- PEKTYVYKRKEIEA-HKKK-FSK PEKTYVYKRKEIEA-HKKK-FSK	IITOPNO HER IITOPNO HER IITOPNO HER STRVANEL PRESTKA INCO PRESTKA INCO PRESTKA INCO PRESTKA INCO PRESTKA INCO IITOR INCO IITOR INCO SECON DU SECON DU SECON DU SECON DU	Helix-4 HHHHHHHHHHHHHHHHHHHHHHH KRHRIKANIQCA:I RWIRIKANIQCA:G TWIRIKANIQCA:-AG TWIRIKANIQCA:-AG TWIRIKANIQCA:-AG TWIRIKALAG TWIRIKANIGCA CARAGA TWIRIKANIGCA TATATANIGCA TWIRIKANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATANIGCA TATATATATATANIGCA TATATATATATATATATATATATATATATATATATAT	- AFGFAGI 2DK) AERI - AFGFAGI 2DK) AERI - FCGFA	VENTAITAERQPERDITPVNDETMQEI TENTAITERQPERDITPVNEETMSEI NMESTVQEEK (RANNTEKLOIPAQ (VLENMASEIVAIRQEEN KEPREVNGVKEPDAQ	205 205 226 237 214 209 226 206 188 267 253 253 261
PHD Soc. Structure Red lambds 137511 Borta Sd 675957 lin1755_Li 16414242 ml17977_ML_13476603 SPy0958_Strpy_13622119 BPhi311_Orf245_7339201 XFJ648_Xf_9106705 UU1154 UD_14195377 FFJ61_BOThercp 9836834 ReeT Ec 1629310 lin2413_Li 1698835 BH3543_Bh 1556105 YvaR Bm 156796811	Strand-4 	Strand-5 - EFEFEFE - STATUTE -		Helix-4 HHHHHHHHHHHHHHHHHHHHHH KIKIR:KANIQCA: KRHR:KANIQCA: THR:TAVWAM:EA LHTA:CAEMOAL:AG TTMIC:TAVGTL:EA TTMIC:VAESQAL:MA GGLE:CAEALAL:KAR TTMIC:VAESQAL:MA GGLE:CAEALAL:KAR TTMIC:VAESQAL:MA HHK:VVLRML:F AMAK:TVLRML:K- AMAK:TVLRML:K- AMAK:TVLRML:K-	- AFGPAGIVDK: AENI - AFGPAGIVDK: AENI - AFGPA	VENTATTAREQUEED ITPVA. DETROET VENTATTERQUEED ITPVA. ESTASI NIM-ETSIQUEI KOMANET. NILDIPAO PD-ALKOVEPDE. QUEANASE IV ALKOVEPDE. QUEANASE IV ALKOVEPDE. DISPATS VENTATTARE VENTATION OF AN AND VENTATION OF AN AND VENTATION OF AN AND NETTON OF AN AND NETON OF AN AN AND NETON OF AN AND NETON OF AN AN	205 205 226 237 214 209 226 206 188 267 253 253 253 261 251
PHD Sec. Structure Red@ lambda 137511 Beta 8d 6759957 lin1755 Li 6614242 ml17977 ML_13476603 SPy058 6 kryp 13622119 BPphill lorf245 7233201 W1544 X f Ju6705 UU154 Ju 14195377 PF161 Borhercp 9836834 ReeT Ec_16129310 elin2413 Li 6801475 ORF48 BPA18 16579833 PH3543 Bh 15516105 yqak Be 16079681 per mPF11300 13095817	Strand-4 	Strand-5 -EEEEEEEE CUTEMUD ECRREPFKTREGRE- CUTEMUD ECRRAPFKTREGRE- APISVD SERENKRQ- PAPISVD SERENKRQ- APISVD STOATKNUDTGH- ATATVYL ATARKDSPA- VFLASVDETVGKTRAFACSA VFLASVTLSETSTNK- SFSVLPHETYLNS- SFSVLPHETYLNS- SFSVLPHETEA-HEKK-FSK- PERTYVYREREIEA-HEKK-FSK- PERTYVYREREIEA-HEKK-FSK- PERTYVYREREIEA-HEKK-FSK- PERTYVYREREIEA-HEKK-FSK-	IIIOPNO HPN IIIOPNO HPN IIIOPNO HPN IIIOPNO HPN PRIVIN HPN PRIVIN HPN PRIVIN HPN IIIOPNO HPN IIIOPNO HPN IIIOPNO HPN Sprokk DDD Sprokk DDD Sprokk DDD Sprokk DDD	Helix-4 HHHHHHHHHHHHHHHHHHHHHHH KHLRIKANIQCA IRMLRIKANIQCA IRMLRIKANIQCA IXMRITAVINAMIBAL LMTAICAEMOALJAG UTHRITALAUNAMIBAL HTMLRICAEMOALJAG GOLECCAEALAUKAI TMLKIVAEALAUKAI HTMLRIVAEALAUKAI AMAKITVLRNHLIK- AMAKITVLRNHLIKAU AMAKITVLRNHLIKAU	- AFGFAGI ZDK/I AERI - AFGFAGI ZDK/I AERI - FFQFI	VENTATTAERQPERDITPVNDETMQEI VENTATTERQPERDITPVNEETMSEI VOIETMSEI VOI	205 205 226 237 214 209 226 206 188 267 253 253 261 251 251 239
PHD Soc. Structure Red. lambda 137511 Beta 56 4575957 lin1755_115614442 n17977_M1_15476603 BPphi31_10+7745 Pphi31_10+7745 Phi31_10+7745 Pilos Phi121_10+7245 Phi121_11 Pilos Phi131_11 Pilos Pilos <tr< td=""><td>Strand-4 </td><td>Strand-5 - EEEEEE - CYTENHD CRRAPYKTREGRE CYTENHD CRRAPYKTREGRE CYTENHD CRRAPYKTREGRE - CYTENHD CRRAPKTREGRE - CYTENHD CRRAPKTREGRE - CYTENHD CRRAPKTREGRE - CYTENHD CRRAPKTREGRE - CYTENH - CYTENH</td><td>III TOPPQ HPA III TOPPQ HPA III TOPPQ HPA III TOPPQ HPA III TOPPQ HPA III TOPPQ HPA IIII TOPP IIII TOPP IIIII TOPP IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII</td><td>Helix-4 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</td><td>- AFGPAGI DKC MAERI - AFGPAGI DKC MAERI - FONG</td><td>VENTATIARROPERDITEVIA DETROIT VENTATTARROPERDITEVIA DETROIT VENTATTEROPERDITEVIA VENTATTEROPERDITEVIA VENTATTEROPERDITEVIA VENTATIARROPERDITE VENTATIAROPERDITE VENTA</td><td>205 205 226 237 214 209 226 266 267 253 253 251 251 251 254</td></tr<>	Strand-4 	Strand-5 - EEEEEE - CYTENHD CRRAPYKTREGRE CYTENHD CRRAPYKTREGRE CYTENHD CRRAPYKTREGRE - CYTENHD CRRAPKTREGRE - CYTENHD CRRAPKTREGRE - CYTENHD CRRAPKTREGRE - CYTENHD CRRAPKTREGRE - CYTENH - CYTENH	III TOPPQ HPA III TOPPQ HPA III TOPPQ HPA III TOPPQ HPA III TOPPQ HPA III TOPPQ HPA IIII TOPP IIII TOPP IIIII TOPP IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	Helix-4 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	- AFGPAGI DKC MAERI - AFGPAGI DKC MAERI - FONG	VENTATIARROPERDITEVIA DETROIT VENTATTARROPERDITEVIA DETROIT VENTATTEROPERDITEVIA VENTATTEROPERDITEVIA VENTATTEROPERDITEVIA VENTATIARROPERDITE VENTATIAROPERDITE VENTA	205 205 226 237 214 209 226 266 267 253 253 251 251 251 254
PHD Sec. Structure Red8_lambda_137511 Beta_58_d575957 lin1755_Li_16414242 ml17977_ML_13476603 SPy0558_Strpy_13622119 BPphi31_corf245_7239201 XF1648_Xf_9106705 UU154_UL_14195377 PF161_Borbercp_9336834 Rec7_Ec_16129310 lin2413_Li_1601631 BH3543_Bh_1561615 BH3543_Bh_1561615 BH3543_Bh_1561615 BH3543_Bh_1561615 BH3543_Bh_1561615 BH5545_B07561 PFC_Lp_13186144 BFC_Lp_1540750	Strand-4 	Strand-5 -EEEEEEE CUTEMUD ECRREPFKTREGRE- CUTEMUD ECRRAPFKTREGRE- - APISVIS SEENKRQ- - APISVIS SEENKRQ- - APISVIS SUDAYKUNDYCAH- - ATTYVILDEVKLORYCA- VFLASVID EVYCHLORYCA- VFLASVID EVYCHLORYCA- VFLASVID EXTRACTOR SFSVLPHYTKINS- SFSVLPHYTK	IIIOPRO HER IIIOPRO HER IIIOPRO HER STRINATU /LROGATPOLD-DSONKC MER PRISTIK INCO PRISTIK INCO PRISTIK INCO IIIONIC MER IIIONIC MER SEGUE SEGUE SEGUE SEGUE ANOPRE HER SEGUE SEGUE NUCPRO HER NUCPRO HER NUC	Boliz-4 HINDHINHHINHHINHHINHHINHHINHHINHHINHHINH	- AFGFAGI ZDK/I AERI - AFGFAGI ZDK/I AERI - FCGT	VENTAITAERQPERDITPVNDETMQEI VENTAITERQPERDITPVNEETMSEI VENTTERQPERDITPVNEETMSEI VANESTVQEEKOAITEGALDIPAQ VV	205 226 237 214 209 226 206 188 265 253 253 253 253 251 251 239 254
PHD Soc. Structure Red. lambda 137511 Beta_564 675957 lin1755_11 1641442 all7977_ML_15476603 BPPhi21-10c7245 723201 Pris12-10c7245 723201 Pris12-10c7245 723201 Pui54 yf 9106705 OrtHeat yf 9106705 OrtHeat 112-111 Broncerce 9036834 Rect Ec 1629310 Nat18 16 16798035 PH3547_186746 Robyrak Brits Broncerce 1136744 Groverce 1136514 Grover 1136514 Grover 1136514	Strand-4 	Strand-5 - EEEEEE - CYTENHD CRAPPKTREGRE - CYTENHD CRAPPKTREGRE - CYTENHD CRAPPKTREGRE - CYTENHD CRAPKRAC - CYTENHD CRAPKAR - AF 18 TO A RAWYLOK CR - AF 18 TO A RA	HIGH CARLS AND	Reliz-4 HIGHIMMERIENEN HIGHI	- AFGPAGI DKC BAERT - AFGPAGI DKC BAERT - AFGPAGI DKC BAERT - PSNLG	VENTATTAERQPERDITFVM	205 226 237 214 209 226 206 188 253 253 253 261 251 251 254 254 264
PHD Sec. Structure RedB. lambda 137511 Beta Sd 675957 lin1755_Li 16414242 all7977_ML_13476603 SPy0558_Strpy_13622119 BPhi31_Lorf245_7239201 XF1648_Xf_9106705 UU154_UL_14195377 PF161_Borhercp_9836834 ReeT_Ec_16129310 lin2413_Li_16801475 ORF48_BFA18_1679681 ReeT_BP51L309_13095817 ORFC16_D136614 gene35_SPP1_5407508	Strand-4 	Strand-5 EBERETE CUTEMO CRERAPYTREGRE- CUTEMO SCREAPYTREGRE- PYSVB15 BREPHKBQ- APISYDAYKWYDGETWEDSGRPKKRR KWTUDPHEYVKLDRYGN- PATYUNAEATATARKDSPA- EAFYLGETSTRK- STSFSUTRUEATATARKDSPA- STSFSUTRUEATATARKDSPA- STSFSUTRUEATATARKDSPA- FEKTYURREELEA-HKK-FYKS- FEKTYURREELEA-HKK-FYKS- FEKTYURREELEA-HKK-FYKS- FKKUTRKEELEA-HKK-FYKS- SKGPUUSEELEA-HKK-FYKS- STMFL/N5GQOVEK-HKK-FKS- SKGPUUSEELGA-SKSKNEGO			- AFORA	VENTAVIAERQPERDITPVM	205 205 226 237 214 209 226 188 266 188 267 253 253 253 253 251 251 251 254 268 264 261
PHD Scc. Structure Red. lambda 17511 Betu-ed 675957 lin1755_Li 16414442 ml1797_ML 18476603 SP0958_5crp 13622119 BPphi31.1_0:rf245_r239201 UU154_UL41953_r036834 RecT_BP161_B076801 RecT_BP1130_Li 16801475 OR746_BPA166169835 BM3543_mL5616105 yqaK_Bz_16579681 AecT_BP511309_13095817 OR7CLp_131861495 OR745_BPFL_032508 Sa1794_SAM315_1592750	Strand-4 SEEEEEE	Strand-5 -EBEEEE 	III TOPPQ BPA III TOPPQ BPA III TOPPQ BPA STNMATL LRNGATPLD-DSGNC MPA PNNGT KC PNNGQ NP NIGHT KC PNNGQ NP III KK NNGPV INS SDFGK DD SDFGK DD SDFGK DD SDFGK DD SDFGK DD OSDFGK DD OSDFGK DD OSDFGK DD SDFGK DD SDF	Relix-4 HINDHIMMENTHEMEN- MICHAENTOCA INTERNETAVIAL AND AND AND AND AND INTERNETAVIAL INTE	- AFGFAGI DKC BAERT - AFGFAGI DKC BAERT - FCGA	VENTANTAERQPERDITPVN DETNQEI VENTANTRAPERDITPVN BETNEEI VENTANTERQPERDITPVN BETNEEI VENTANTERQPERDITPVN BETNEEI VO LENNAEU MEGUEEN VO LENNAEU MEGUEEN VENTANTERQPERDITPV AUTOPO AUTOPO VENTONEDORQ LESFORE BENNENCOLOURIENDAR VENTONEDORUM BENNENCOLOURIENDAR BENNENCOLOURIENDAR VENTONEDORUM BENNENCOLOURIENDAR BENNENCOLOURIENDAR RANDOLUMINENDAR BENNENCOLOURIENDAR BENNENCOLOURIENDAR RANDOLUMINENDAR BENNENCOLOURIENDAR BENNENCOLOURIENDAR RANDOLUMINENCOLOURIENDAR BENNENCOLOURIENDAR BENNENCOLOURIENCOLOURIENCOLOURIENDAR RAN	205 226 237 214 206 286 267 253 261 251 251 254 264 254 264 261 251 261 261 261
PHD Soc. Structure RedB lambda 137511 Beta 58 d55957 lin1755_Li 16414242 all7977_M_13476603 SPy0958_Strpy_13622119 BPh3131_orf245 7239201 XP1648_Xf_9106705 UU154_UL_14195377 PF161_Bornercp_9836834 RecT_Ec_1629310 lin2413_Li_16801475 ORf48_BPA118_16769881 RecT_BP5L309_13095817 ORf40_BPA1186144 gene35_SPP1_540750 orf43_BP47_9633135_15927560 SA1794_SaN315_15927560	Strand-4 	Strand-5 EBERETE EDETET (UTENDO CARREPFKTREGRE- LIVTENDO CARREPFKTREGRE- PFVSVB15 STAREPKTREGRE- PFVSVB15 STAREPKTREGRE- PATTVKLEATATARKOSPA- PATTVKLEATATARKOSPA- PATTVKLEATATARKOSPA- STAFUTUELEL- KSTSSULGUEL-VRSL- SKAGUTELE- FKTVYTREELEA-HKK-FVK- SK- PEKTVYTREELEA-HKK-FVK- SK- PKTVYTREELEA-HKK-FVK- SK- PKTVYTREELEA-HKK-SK- SK- PKTVYTREELEA-HKK-SK- SK- PKTVYTREELEA-HKK-SK- SK- PKTVYTREELEA-HKK-SK- SK- SK- SK- SK- SK- SK- SK- SK- S	Н - ТТОРРО НРА - ТТОРРО НРА - ТТОРРО НРА - STM NAUL - STM NAUL - STM NAUL - STM - STM - STM - STM - STM - STM - S	Boliz-4 Internet Market Con- Internet Market Con- Market Con- Mark	- AFOÜAOITDK: LAERI - AFOÜAOITDK: LAERI - PDRIGOITDK: LAERI - PDRIG	VENTATTARROPERDITPVM	205 205 225 237 214 209 226 266 267 253 251 253 251 251 251 259 254 268 261 278
PHD Scc. Structure Red: _imbds 137511 Beta_56 _ 675957 11n175_1.34416603 BPy059 _Strpy _13622119 BPphill. 0.ef244 _ 7239201 VII448 _ VI4195377 PF161 _BORECTD 9836834 RecT _BP11.268014755 OR746 _BPA18 _ 16516105 yqaK _Bs_16079631 RecT _BP11.309_13095817 OR7CLp_13186144 gene33_BPV_540550 orf43 _BPVL 983208 Sa1794 _San315_15927560 ORF10 _BPRLT_133527 Spr1477_Strpy_156753333	Strand-4 StreeEeEE	Strand-5 -EEEEEE 	II TOPPQ BPA II TOPPQ BPA II TOPPQ BPA STNAMU /LROGATQLD-DSGNCCMPA PNKMC MPA PNKMC MPA PNKMQ NPC -PNKMQ NPC -PNKMQ NPC -PNKMQ NPC -PIR KVP -NNOPV HE SDFGK DD SDFGK DD SDFGK DD -SDFGK DD -SDFGK DT -SDFGK DT -SDFK -STFCPA -STFC	Relix-4 HINDHIMMENTHEMENT HINDHIMMENTHEMENT HINDER KANTGCA INTER TAVAN THE ANALON HINDHIMMENT HINDHIME	- AFGFAGI VDKO BAENT - AFGFA	VENTAITAERQPERDITPVN	205 205 226 237 214 209 226 266 265 253 253 253 251 251 254 254 261 261 261 261 278 261
PHD Soc. Structure Red Lambda 137511 Beta 56 d57957 lin1755_Li 16414242 ml17977_M_13476603 SPy0958_Strpy_13622119 BPh3131_Orf245 7239201 XF1648_Xf_9106705 UU154_U_14195377 PF161_BOTHercp_9836834 RecT Ec 1629310 lin2413_Li 16801475 ORf48_BPA118_1679681 RecT BebL1309_13095817 ORf48_BPA1186144 gene35_SPP1_540750 Orf43_BPPU_9635208 SA1794_SaM315_15927560 ORF10_BF11_1355278 SA1794_SaM315_15673833	Strand-4 	Strand-5 EDEDEDCU EDITED SUPERIO CAREPYTREGRE- LIVERIO CAREPYTREGRE- PYSVELSTERPHER PYSVELSTERPHER PATTY IS THAL THE SAMPTICE THE START AND THE SAMPTICE PATTY WAYS DE TYME TATTY WAYS DE TYME TATTY WAYS DE TYME THE START PETTY WARE DE LE VASL-SKAC- PETTY WARE DE LE VASL-SKAC- SKAUT ME SEQULA-HEKK-FVKS- BADY IE WAT DE URGANG-SKANT WAS SKAUT WARE DE URGANG-SKAUT WAS WAS	TIGPNO HIM TIGPNO	Boliz-4 Internet Internet Internet Internet Construction Internet	L AFGÜAGITOK BAERI AFGÜAGITOK BAERI -PRIGGITOK BAERI -PRIGGITOK BAERI -PRIGGITOK BAERI -PRIGGITOK BAERI -PRIGGITOK BAERI -PRIGGITOK BAERI -PRIGGARANSM KERL -PRIGGARANSM KERL GILSG DGRANSM KERL LISSS DGRANSM KERL LISSS D	VENTAYTAERQPERDITPVN	205 205 225 237 214 209 226 266 267 253 253 253 253 251 251 251 254 268 261 278 268 261 278 261 278
PHD Scc. Structure Red: Lambda 137511 Beta 5d 675957 11n1755.L16414242 ml17977 ML13476603 SP40959 Strpy 13622119 BPphill.l0.ef245 7239201 W1548 4W 14195377 PF161 Borkercp 9836834 RecT Ec_16129310 0R748 BPA18 16516105 yqaK Bs_16079681 RecT BP51L309 13095817 OR7C4 Bs_15096144 gene33_SP15_540750 0rf43 BPFU_9635208 SA1794 San15 15527560 OR710 BFRIT_135327 SP1477, Strpy 156753333 EHAP1 Ec_484456 ST12074, Salem_16760818	Strand-4 EEEEEEEEE- 	Strand-5 -EEEEEE 	IITOPHQ BPA IITOPHQ BPA IITOPHQ BPA STNANT, ILRDATTQLD DSGNC MPA PNNMQ NPA PNNMQ NPA PNNMQ NPA IIL KFI PNNMQ NPA INOPY NPA SDFGKK DTD SDFGKK DTD SDFGKK DTD SDFGKK DTD SDFGKK DTD SDFGKK DTD SDFGKK DTD SDFGKK DTD SDFGKK DTD DNOPY NIT VICPA HTP VICPA STAN PK SDFGK DTD DTS IIFKKLIZATNING. FPF DFK JFCFFA ITD STLA DFK	Relix-4 HINHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	- AFGFAGI VDKO - AERT - AFGFA	VENTAITAERQPERDITPVNDETNQEI TENTAITTERQPERDITPVNDETNQEI TENTAITTERQPERDITPVN	205 205 226 237 214 209 226 268 263 253 253 253 254 261 254 261 261 278 278 278 278 278 278
PHD Soc. Structure Red Lambda 137511 Beta 56 d57957 lin1755_Li 16414242 ml17977_ML_13476603 SPy0958_Strpy_13622119 Pphi31_0rf245_7239201 xF1648_xf_9106705 UU154_UL_14195377 PF161_Borhercp_9836834 RecT Ec_16129310 lin2413_Li_1607661 RecT BrbI13091186144 gene35_SPP1_540750 orf43_BPP1_9635208 SA1794_SaN315_15927560 SA1794_SaN315_15927560 SA1794_SaN315_15927560 SA1794_SAN315_15927563 SA1794_SAN315_1575333 EHAP1_Ec_484296 STU2074_Salent_16760318	Strand-4 	Stind-5 Stind-5 ICVTFMD CARPYTREGRE- ICVTFMD CARPYTREGRE- PVSVL5 SHEPNERG- PVSVL5 SHEPNERG- PVSVL5 SHEPNERG- PVSVL5 SHEPNERG- PVSVL5 SHEPNERG- PVSVL5 SHEFNERG- PATVYNERGUSTEL-VRSL- SKOTSHE- PETVVFREIE-HERG-FSK- PETVVFREIE-SKG- PETVVFREIE-SKG- PETVVFREIE-SKG- PETVVFREIE-SKG- PETVVFREIE-SKG- PETVVFREIE-SKG- PETVVFREIE-SKG- SKS- PETVVFREIE-SKG- SKS- SKSVT5- SKSVT- S	ITGP 00 HI ITGP 00 HPA ITGP 00 HPA ITGP 00 HPA VLRDGATPQL- PESTWK. MPG PESTWK. MPG PPG VLRDGATPQL- PESTWK. MPG PHSM00 RFP VLRDGATPQL- PESTWK. MPG PHSM00 RFP VLLL IFI SDFGWK. DTD SDFGWK. DTD SDFGWK. DTA SDFGWK. DTA SDFKWK. DFA SDFGWK. DTA SDFKWK. DFA SDFKWK. DFA SDFKWK. DFA SDFKWK. DFA SDFKWK. DFA SDFKWK. DFA SDFKWK. DFA SDFKWK. DFA </td <td>Reliz-4 Internet Methods and Anti- Reliance of the second secon</td> <td>L AFGÜAGITXK: BAERI - AFGÜAGITXK: BAERI - AFGÜAGITXK: BAERI - AFGÜA</td> <td>VENTATTARROPERD ITPVM. DETNOET UENTATTEROPERD ITPVM. DETNOET NIM-ETSUYDET KOMMTES. ULD IPMO V. L. STANDER V. MEDGEN D. ALKOVETPOS. DETOE V. L. SEET. HVAPI. DETOE HETI - EGET. HVAPI. DETVISON HETI - EGET. HVAPI. DETVISON HETI - EGET. HVAPI. DETVISON HETI - EGET. HVAPI. DETVISON HETI - STANDE - SITO ITAD PROMOSTPEV. DETVISON R. KOVETDE-SITO ITAD PP. ADNVVE- EFFKPEVP PP. ADNVVE- EFFKPEVP RR. KOVETDE-SITO ITAD PP. ADNVVE- EFFKPEVP RR. KOVETDE-SITO ITAD PP. ADNVVE- EFFKPEVP RR. KOVETDE-SITO ITAD PP. ADNVVE- EFFKPEVP RR. KOVETDE-SITO ITAD RR. KEVLDAE-VEENANGE RR. KEVLDAE</td> <td>205 205 226 237 214 209 226 188 253 253 251 251 251 259 254 268 261 278 268 261 278 268 261 278 268 261 278 263</td>	Reliz-4 Internet Methods and Anti- Reliance of the second secon	L AFGÜAGITXK: BAERI - AFGÜAGITXK: BAERI - AFGÜAGITXK: BAERI - AFGÜA	VENTATTARROPERD ITPVM. DETNOET UENTATTEROPERD ITPVM. DETNOET NIM-ETSUYDET KOMMTES. ULD IPMO V. L. STANDER V. MEDGEN D. ALKOVETPOS. DETOE V. L. SEET. HVAPI. DETOE HETI - EGET. HVAPI. DETVISON HETI - EGET. HVAPI. DETVISON HETI - EGET. HVAPI. DETVISON HETI - EGET. HVAPI. DETVISON HETI - STANDE - SITO ITAD PROMOSTPEV. DETVISON R. KOVETDE-SITO ITAD PP. ADNVVE- EFFKPEVP PP. ADNVVE- EFFKPEVP RR. KOVETDE-SITO ITAD PP. ADNVVE- EFFKPEVP RR. KOVETDE-SITO ITAD PP. ADNVVE- EFFKPEVP RR. KOVETDE-SITO ITAD PP. ADNVVE- EFFKPEVP RR. KOVETDE-SITO ITAD RR. KEVLDAE-VEENANGE RR. KEVLDAE	205 205 226 237 214 209 226 188 253 253 251 251 251 259 254 268 261 278 268 261 278 268 261 278 268 261 278 263

Strand-1

Strand_2

Holiz-2

Figure I

Multiple sequence alignment of the RecT/Red β **superfamily of proteins.** Proteins are denoted with their gene names, species abbreviation and gi numbers. The coloring reflects the amino acid conservation at 85% consensus. The consensus abbreviations and coloring scheme are as follows: h: hydrophobic residues (L,I,Y,F,M,W,A,C,V), l: aliphatic (L,I,A,V) and a: aromatic (F,Y,W,H) residues shaded yellow; o: alcohol (S,T), colored blue, c: charged (K,E,R,D,H) residues, +: basic (K/R/H) residues, -: acidic (D,E) residues, and p: polar (S,T,E,C,D,R,K,H,N,Q) residues colored purple; s: small (S,A,C,G,D,N,P,V,T) and u:tiny (G,A,S) residues, colored green; b: big (L,I,F,M,W,Y,E,R,K,Q) residues shaded gray. Secondary structure assignments are as follows: H: Helix, E: Extended (Strand). Species abbreviations are as follows: Bh: Bacillus halodurans, Borhercp: Borrelia hermsii circular plasmid, BPA118: Bacteriophage A118, BPbIL309: Bacteriophage blL309, BPphi31_1: Bacteriophage phi31.1, BPPVL: Bacteriophage PVL, BPR1T: Bacteriophage R1T, Bs: Bacillus subtilis, ec: Escherichia coli, lambda: Bacteriophage λ , Li: Listeria innocua, Lp: Legionella pneumophila, MI: Mesorhizobium loti, Salent: Salmonella enterica subsp. enterica serovar Typhi, SaN315: Staphylococcus aureus N315 subsp. aureus N315,Sd: Shigella dysenteriae, SPP1: Bacteriophage SPP1, StLT2: Salmonella typhimurium LT2, Strpy: Streptococcus pyogenes, Uu: Ureaplasma urealyticum, Xf; Xylella fastidiosa

ing that they are probably exposed and participate in the protein-protein and protein-DNA interactions that are typical of this superfamily (helices 2,3, 4 in Fig. 1). The conserved, regularly spaced hydrophobic residues in the RecT/Red β superfamily are predicted to be buried, allowing these domains to assume a globular structure. Experimental studies have shown that the strand transfer reaction mediated by RecT and its binding to dsDNA are sensitive to Mg⁺² concentrations and it was proposed that the levels of free Mg⁺² could regulate RecT activity [4]. Similarly, Red β has been shown to promote single strand annealing in a Mg⁺²-dependent manner [20]. In this context, the conservation of the two C-terminal acidic residues in the majority of members of this superfamily suggests that these might be involved in the coordination

of Mg⁺² and implies that the metal ion-dependent conformational switching is likely to be a generic feature of this family.

Phylogenetic analyses of the RecT/Red β superfamily using the least squares and maximum likelihood methods distinguished three distinct groups, namely the RecT-like, the Red β -like and the EHAP1-like families (Fig. 2). Previously, the RecT proteins have been known from very few bacteria and Red β has only been detected in λ and closely related phages. However, we showed that the Red β family is widespread in bacteria, such as *Borrelia hermsii*, *Xylella*, *Ureaplasma*, *Listeria*, *Streptococcus pyogenes*, *Mesorhizobium loti*. The RecT family is predominantly seen in the low-GC Gram-positive bacteria, such as *Bacillus*, *Streptococcus*, *Lac*-



Maximum likelihood tree for the RecT/Red β **superfamily of proteins.** The internal branches with RELL bootstrap support >70% are indicated by blue circles. Proteins are designated by their gene names and species abbreviations as in Fig. 1. The gene neighborhoods of the RecT/Red β superfamily genes are shown in association with the corresponding branches whenever they they contained genes for proteins with plausible functional connections with SSAPs. The hatched boxes represent fragments of ERF (Indicated by E) and SSB (indicated by S) genes encoding C-terminal regions as described in the text. Gene abbreviations are as follows: GP46: GP46 of bacteriophage PSA, GP32:GP32 of bacteriophage PSA, ORF15:ORF15 of *Streptococcus thermophilus* bacteriophage 7201, ORF40: ORF40 of bacteriophage PVL ORF86:ORF86 of *Staphylococcus aureus* temperate phage ϕ SLT, Orf100a:0rf100a of *Staphylococcus aureus* temperate phage ϕ SLT, ORF364: ORF364 of bacteriophage ϕ 31.1, Ec2360: b2360 of *E. coli*, AcylTr: N-Acyltransferase, Bro: Bro-N domain fused to XF0704, Met: DNA Methyltransferase

tococcus and Listeria, and their phages (Fig. 2). *E. coli* and *Legionella pneumophila* are the only two γ -proteobacteria that possess this protein, suggesting that they might have acquired RecT via a relatively recent horizontal transfer from Gram-positive bacteria. The sporadic distribution of the RecT and Red β family proteins in bacterial genomes and their presence in phages suggest that these proteins ultimately are of phage origin and have been co-opted by the bacterial DNA recombination/repair systems. Consistent with this, practically all the bacterial members of these families appear to belong to prophages or their remnants,

as they are mostly in the neighborhood of what appear to be clearly phage-derived genes. The EHAP1-like family is extremely divergent and represented thus far only in *E. coli* and the closely related *Salmonella*. Practically all members of the RecT/Red β superfamily are single-domain proteins showing extended similarity to each other throughout their globular regions. The only exceptions are the *E. coli* EHAP1 and the PF161 protein encoded in the *Borrelia hermsii* circular plasmid, which are fused to C-terminal fragments of the ERF protein (see below).

		Helix-1		Helix-	2 Helix-	3 Strand-	-1 Strand-2	2 Strand-3	
PHD Sec. Structure		- ННННННННН	[HHHHH	НННННН	HEEEE	EEEEE	EEEEEEEEE-	
ERF P22 9635505	5	FYARLAE IQEHL	NAPKNQY <mark>N-SF</mark> GK	(Y <mark>ky</mark> rsced <mark>i</mark>	LEG <mark>V</mark> KP <mark>L</mark>	LKG <mark>L</mark> FLSIS-	DEIVLI	-GDRYY <mark>V</mark> KATATITE	GENS
ERF ⁻ HK97 9634188	5	FYARLAAIQENL	NAPKNQY <mark>N-SF</mark> GK	(YK <mark>Y</mark> RSCED <mark>I</mark>	LEG <mark>V</mark> KP <mark>L</mark>	LNG <mark>L</mark> FLSIS-	DEVVLI	-GDRYY <mark>V</mark> KATATITE	GENS
orf17 phiPV83 9635694	10	INKA <mark>M</mark> VA FR KE <mark>V</mark>	KQPLK-DKNN-PFFK	(SK <mark>Y</mark> VPLEN <mark>V</mark>	VEA <mark>I</mark> DE <mark>A</mark>	ATPH <mark>GL</mark> SYT	QWALNDVD-	- G RVG- <mark>V</mark> ATMLMHES	GEYI
lmo2318 Lm 16804357	10	IAKA <mark>M</mark> AAIQKE <mark>V</mark>	KPLEKSAAN-PFTI	OS <mark>KY</mark> TPLDK <mark>I</mark>	VEA <mark>I</mark> FK <mark>V</mark>	APGH <mark>GV</mark> SFT	QWPISGDN-	- G TIG- <mark>I</mark> GTMLMHES	GEWI
orf216 phiSLT 12719409	7	LYQK <mark>I</mark> AD <mark>VK</mark> AN <mark>I</mark>	AGFTKDTK-GY-N	IFS <mark>Y</mark> VSGSQ <mark>I</mark>	LHR <mark>I</mark> REKI	MIEH <mark>NL</mark> LLVPN'	LNNEWTTHTF	-KNKKGQEVTEFI	VEMDLNYTWINAD
ORF4 Bbcp8.3 458219	18	FISD <mark>M</mark> KTLRMNL	AGIDKNLK-GY-G	GY <mark>K<mark>Y</mark>QNFNE<mark>I</mark></mark>	VRE <mark>I</mark> KN <mark>V</mark>	INKH NL YLDFK-	QFPTFTVVE	E G QQV- <mark>L</mark> HVVRTTFY	-STNTGYKDSFDTPIL
BBL29 Bbcp32-8 11497308	41	FRKDMKTLKMNL	PGIDKSLK-GY-G	GY <mark>K</mark> YQNFNE <mark>I</mark>	v re<mark>i</mark>kn<mark>v</mark>	IKKH <mark>NL</mark> ELDIE-	QYPISIE	-GQYGI <mark>V</mark> DYIRTTFY	-STSTGYEFSFDTRIP
BB029 Bbcp32-7 11497280	36	FRKDMKTLKMNL	PGIDKSLK-GY-G	GY <mark>K<mark>Y</mark>QNFNE<mark>I</mark></mark>	VRE <mark>I</mark> KN <mark>V</mark>	IKKH <mark>NL</mark> ELDIE-	QYPISIE	-GQYGI <mark>V</mark> DYIRTTFY	-STSTGYEFSFDTRIP
BBR29 Bbcp32-4 11497168	47	FRKDMSTLIRNI	PRIDKSLK-GY-G	GY <mark>ky</mark> qdfnd <mark>i</mark>	VEV <mark>I</mark> YS <mark>V</mark>	IDKH <mark>NL</mark> DLFFT-	QAPISVE	- G QYGI <mark>V</mark> DYIRTTFY	-STSTVYKYSFDTRIH
BBC05 Bbcp9 11527300	19	FLRDMETLRMNI	PGIDKNLK-GY-G	GY <mark>KY</mark> QNFNE <mark>I</mark>	A <mark>re</mark> ikk <mark>v</mark>	IBKH <mark>NL</mark> CLDFK-	QFPTFTVV-	-GEQQV <mark>L</mark> HVVRTTFY	-STNTGYKDSFDTPIL
ORF7 ST7201 9634634	2	EEMTFTELQOKI	QLEKKKE-GTAK	(YASRHVED <mark>I</mark>	YNVFKNL	KS NW NVVV	NYELVEF	- S GKTF <mark>I</mark> KAIATASN	KDEK
Orf52 D3 9635643	85	FNASMAAMQSE1	ERGPSIAAITVNGQKF	RSN <mark>Y</mark> ATFED <mark>I</mark>	NDI <mark>V</mark> KP <mark>I</mark> I	MQRF <mark>GF</mark> AVSFR-	VETVQT	-GVS <mark>V</mark> TGILMHCA	GHRE
Orf223 Lj771 13491642	13	WAMHY AQVKANI	KOPEKTHKVT V SGKTKOGT PY- S	SYD <mark>Y</mark> NYADLN	DIDAAVMDGIK	KVTDKDGNVVF-	SYFFDIRTE	e n ntve <mark>v</mark> otilvdss	GFTL
ERF Unk 6015511	7	LYEALAETONNI	EOPKK-DASN-PMFK	(S <mark>S</mark> YVTLDA <mark>V</mark>	INA <mark>I</mark> VK <mark>A</mark>	RKAS <mark>GA</mark> KFFF	TNIVQD	-GVM <mark>F</mark> TRIIGYGE	TLDL
ORF244 BPmv4 11138335	10	IFGALSKFRAQV	KQPAK-TAKN-PYFN	ISN <mark>Y</mark> VTLEG <mark>V</mark> I	MQS <mark>I</mark> DA <mark>A</mark>	LPGT <mark>GL</mark> AYC	QLVENGDN-	-GVS <mark>V</mark> STLITHSS	
ERf bIL67 9627938	1	MESK <mark>I</mark> LKLINEI	EVPKSOYN-SYGK	(YN <mark>F</mark> RNNED <mark>I</mark>	0TALKPL	LLKY <mark>GL</mark> MEVA	GTEMLEM	- n nelm <mark>l</mark> hvhveife	PENPNDV
ERf c2 9628668	1	MESKVLKLINEI	KVPKSOYN-SFGK	(YN <mark>F</mark> RNNED <mark>I</mark>	0TALKPL	LLOF <mark>GL</mark> MEKA	TTEMLEM	- n nelm <mark>l</mark> hvhidife	PDNPNDI
ERF MMI 15088753	2	ADLTFAELQRKM	1QIEKQTK-QGVK	(YP <mark>f</mark> rtaed <mark>i</mark>	NNK <mark>F</mark> KSL	DS GW SVSFP-	EDDIIQK	-GDKLY <mark>Y</mark> KAVAVVKR	ESDGTI
consensus/85%		hhhb.pl	p.sb	.pa.s.p.l	p.hh	sh		.sh	
		1		1					

	Strand-4	He	lix-4	Helix-5
PHD Sec. Structure	EEE	HHHHHH	ННННННННН	НННННННН
ERF_P22_9635505	HSASAIAREEENKK	GMDAAQVTGATSS <mark>YA</mark>	RKYCLNG <mark>LFGI</mark> DD-A	KDADTEEHKQQQNAAPAKQTKSSPSSPAPEQVLKAFS 157
Gp40 HK97 9634188	HTATALAREEESKK	GMDSAQVTGATS <mark>SYA</mark>	RK <mark>YCL</mark> NG <mark>LFGI</mark> DD-A	KDADTDEHKHQQNAAAKQSKPSPTPEQVLKAFT 153
orf17 phiPV83 9635694	EYDPVFMNAEK	NTPQGAGSLISYL	KR <mark>YSL</mark> SA <mark>IF</mark> G <mark>I</mark> TS-D	QDDDGNEASGKNNNPKQQTRTQWASSETIGILRKEVISFTKL 164
1mo2318 Lm 16804357	EYDPLYMTVITNKK	MSSAQEAGGTIT <mark>YA</mark>	kr <mark>yvi</mark> aa <mark>vf</mark> givs-d	EDKDGNIQSSPSKYPKNN-SYKSN-NYNQNSNSQQQTKQSQQNDN 169
orf216 phiSLT 12719409	KPEEQYEVSYHAYGQQ	NDISQAHGTALT <mark>YA</mark>	er <mark>y</mark> f <mark>l</mark> mk <mark>ffni</mark> ptde	DDADAKQKQDKYSTVSQEFKDILTKEVN 165
ORF4 Bbcp8.3 458219	TENLKWYNENGAKN-	VNTVPQLVGSSITYF	KR <mark>YAL</mark> VA <mark>YLNI</mark> ES-E	VDTDAAPINNNYKNEN-SNMPSKQVIVNQEQKKDINQNQNQIK 189
BBL29 Bbcp32-8 11497308	TENLQWNNENGSKV-	TNTVYQMFGSGIT <mark>YV</mark>	<mark>kr<mark>y</mark>alva<mark>algi</mark>es-e</mark>	IDTDAAPIYNNHENENSMSSKQVSVNQKQEQKREQKQEINQ 210
BB029 Bbcp32-7 11497280	TENLQWNNENGSKV-	TNTVYQMFGSGIT <mark>YV</mark>	<mark>kr<mark>y</mark>alva<mark>algi</mark>es-e</mark>	IDTDAAPIYNNHENENSMPSKQSSVNQKQEQKREQKQEINQ 205
BBR29 Bbcp32-4 11497168	TDKLQWNSENGSKN-	MNTMPQFVGSAITYF	KR <mark>Y</mark> ALVGH <mark>LCI</mark> RS-E	MDTDAAPIYNNYENRNSMPSKQSSVNQKQEQKREQKQEINQ 216
BBC05 Bbcp9 11527300	TENLKWNNENGSKNV	VNTVPQLVGSSIT <mark>YF</mark>	KR <mark>Y</mark> ALVA <mark>YL</mark> NIES-E	VDTDAAPIYNNHENENSMPSKQAGVNQNQIKNFDKKLKTGK 190
ORF7 ST7201 9634634	VQAQAFAELSPVPILKTRNGE	LKQMNEPQWVGAVQ <mark>S</mark> YA	GK <mark>YAL</mark> QA <mark>LF</mark> A <mark>I</mark> GE	EDVDHFEVAEQSMRPNQNHNQMQSHQQANYINQQQH 160
Orf52 D3 9635643	QTTMLVPLDTSGSK	NAVQSLGSSVSYG	KR <mark>YVL</mark> SA <mark>LLNI</mark> TTRG	EDDDGNAAVPPKKLITKAQAQQLKALLSQCL 234
Orf223 Lj771 13491642	KTNKVVFQNNK	AWDAQATASLIS <mark>YA</mark>	KR <mark>YSL</mark> SG <mark>AFGI</mark> A	ADNDDDAQDQKTIYEPKILTKQELEDYKVYYNG 174
ERF Unk 6015511	AGSKVADDLGNR	GTNSAQAEGSALTYA	RR <mark>YSL</mark> SM <mark>AFGI</mark> A	SDVDDDGNSAGGSKRKPEAPKTISQEKLVLLEKLITDTS 157
ORF244 BPmv4 11138335	GEWMIVGPLTL-APTK	RDPQGQGSAITYA	KRYQLAS <mark>AFGI</mark> S	SDIDDDGNACSFGEDRQSGYQRQSAQNKSYRGQNANQGNRQ 160
ERf bIL67 9627938	TSGDGWAVIDVNKK-	GMDKAQATGASQS <mark>YA</mark>	skyaygqalkldd-i	KDADSTNKGQNNAQRPKAVPKASYQYKLSDLKKMVAN 158
ERf c2 9628668	ASGDGWAVIDINKK-	GMDKAQATGASQS <mark>YA</mark>	skyaygqalkldd-t	KDADSTNKGPNNATQMKSRPKSNYQYNLSDLKKKVAN 158
ERF MM1 15088753	EKAIG W AREEDVPIFH TQK GD	VKQMQDPQWTGAVG <mark>S</mark> YA	RK <mark>Y</mark> ALQG <mark>LF</mark> AIGG	EDVDEYPVEESQEQGQNNQQQKPNNQQAQGQNQVRYIDNTQYQEI 172
consensus/85%	b	p.sQGss.oYh	p+YslhhsI	.DsDsp.pp

Multiple sequence alignment of the ERF protein superfamily. The coloring reflects the amino acid conservation at 85% consensus. The coloring scheme and secondary structure assignment abbreviations are as in Fig. I. Species abbreviations are as follows: Bbcp32-4: Borrelia burgdorferi cicular plasmid (cp) 32-4, Bbcp32-7: B. burgdorferi cp 32-7, Bbcp32-8: B. burgdorferi cp 32-8, Bbcp8.3: B. burgdorferi cp 8.3, Bbcp9: B burgdorferi cp 9, blL67: bacteriophage IL67, c2: Lactococcus phage c2, D3: Pseudomonas phage D3, HK97: bacteriophage HK97, Lj771: Lactobacillus johnsonii prophage Lj771, Lm: Listeria monocytogenes, MM1: Streptococcus pneumoniae bacteriophage MMI, BPmv4: Bacteriophage mv4, P22: Bacteriophage P22, phiPV83: Bacteriophage φPV83, phiSLT: Staphylococcus aureus temperate phage φSLT, ST7201: Streptococcus thermophilus bacteriophage 7201, Unk: Unknown.

ERF defines a superfamily of SSAPs that are evolutionarily distinct from the RecT/Red β super family

The ERF protein of phage P22 is involved in the circularization of the linear dsDNA phage genome upon entry into the host cell [21-23]. Experimental studies have shown that, mutations in ERF are complemented by Red β and that in vitro ERF adopts quaternary structures analogous to those of Red β and RecT [14,17,24,25]. However, in the comprehensive analysis of the RecT/Redß superfamily no statistically significant similarity could be detected between these proteins and the ERF proteins. To explore the evolutionary affinities of the ERF domains, we carried out a sequence profile analysis as described above for the RecT case using transitive PSI-BLAST analysis. As a result of these searches, homologs of ERF encoded in several bacterial and phage genomes from diverse taxa were identified.

The alignments generated in these searches consistently pointed to a region of approximately 150 amino acids that is conserved in all these proteins, with a characteristic motif of the form: GuXXoYhp + YXhXXhh (where G is glycine, Y-tyrosine, u is a tiny residue, h-hydrophobic, p is a polar residue, o is an alcohol residue, + is a basic residue, and X is any residue; Fig. 3). This suggested that ERF was the prototype of a family of conserved bacterial domains.

Secondary structure prediction based on the multiple alignment of the ERF domain suggests a globular α + β fold with five helices and three or four strands (Fig. 3). The above-mentioned motif that is typical of this family is associated with helix 4 of this domain; given the presence of conserved basic residues, it may be critical for DNA-binding and strand-transfer activity of the ERF-like proteins.



Maximum likelihood tree for the ERF superfamily of proteins. The designations, gene names and species abbreviations are as in Fig. 2A. The internal branches with RELL bootstrap support >70% are indicated by blue circles. The gene neighborhoods of the ERF proteins are shown whenever they contained gene coding for proteins with potential functional relevance. Gene abbreviations are as in Fig. 1B

Additionally, in the loop between helices 4 and 5 of the ERF domain there is a universally conserved acidic motif of the form DXD. Analogous to the RecT superfamily, this acidic dyad might coordinate a divalent cation and undergo a conformational change dependent on metal-binding. However, the average size of the core domains, the patterns of conserved residues, and the predicted secondary structures of the RecT/Red β and ERF domains show no correspondence to each other, implying that there is no direct evolutionary link between these protein groups.

ERF homologs are encoded by the genomes of several temperate phages of Gram-positive bacteria and γ -proteobacteria; additionally, we detected members of this superfamily in *Listeria* and in all the circular plasmids and one linear plasmid of *Borrelia burgdorferi* (Fig. 4). Thus, like the RecT/Red β superfamily, the ERF family is likely to have emerged in the temperate phages, and was disseminated to the *Borrelia* circular plasmids and some bacterial genomes via prophages.

Detection of bacterial homologs of RAD52 and identification of an aberrant HhH domain in these proteins

The baker's yeast protein RAD52 and its paralog RAD59 define a small family of proteins thus far represented in fungi, vertebrates and the early-branching ameboid eukaryote, *Entamoeba histolytica*. Rad52 functions in conjunction with the RecA ortholog, the RAD51 recombinase in double-strand break repair and meiotic recombination [6]. RAD52 binds ssDNA during recombination and also



Multiple sequence alignment of the RAD52 protein superfamily. The coloring reflects the consensus at 90% conservation. The coloring scheme and secondary structure assignment abbreviations are as in Fig. 1. Species abbreviations are as follows: BPul36: Bacteriophage ul36. Cab: *Clostridium acetobutylicum*, Dr: *Deinococcus radiodurans*, Hs: Homo sapiens, Kla: *Kluyveromyces lactis*, NC: *Neurospora crassa*, Sc: *Saccharomyces cerevisiae*, Sp: *Schizosaccharomyces pombe*, Sparatyphi: *Salmonella paratyphi* A, Ralbus: *Ruminococcus albus*. The Shiga toxin-converting phage RAD52-like protein (gi: 17977996) is nearly identical to the *Salmonella paratyphi* A RAD52 like protein. The RAD52-like proteins from Bacteriophage ul36 (gi: 8248159) and *Ruminococcus albus* are respectively adjacent to genes encoding the single-strand binding protein and the λ -type exonuclease.

shows a quaternary organization similar to those of RecT/ Red_β and ERF [16,26]. However, RAD52-like proteins showed no detectable sequence similarity with either the ERF or the RecT/Redβ-like proteins. Sequence searches initiated with the conserved globular region of the eukaryotic RAD52 proteins readily detected their homologs from other eukaryotes and, at convergence, also retrieved from the database certain bacterial proteins, such as DR0423 from Deinococcus and CAC1936 from Clostridium respectively, with border-like statistical significance ($e \sim .05$). These bacterial proteins form a small family that is additionally represented in Salmonella paratyphi A, the temperate bacteriophage u136 of Lactococcus lactis (ORF252encoded protein) and a Shiga toxin-converting phage from E. coli. Iterative profile searches initiated with CAC1936 from Clostridium acetobutylicum and its S. para-

typhi A ortholog correspondingly retrieved S. cerevisiae RAD52 and its eukaryotic homologs, with borderline evalues at convergence (~0.043). The alignment between these bacterial proteins and the eukaryotic Rad52 homologs was co-linear throughout the entire length of their shared globular region and the Gibbs sampling procedure detected two motifs of greater than 20 residues, with a probability of chance occurrence in these proteins less than 10⁻¹⁸ (Fig. 5). In addition to the similar conservation pattern, separate secondary structure predictions for both the eukaryotic RAD52 family and their potential bacterial homologs showed a complete concordance of the predicted structural elements between RAD52 and the bacterial proteins, strongly suggesting that they all belong to a single homologous superfamily (hereinafter the RAD52 superfamily).

The secondary structure predictions showed that the Rad52 superfamily proteins adopt a structure with interspersed α -helices and β -strands (Fig. 5). Additionally, fold predictions using 3DPSSM (E-value=.0085, corresponding to a 90% confidence in the prediction) and the hybrid fold method (Z-score = 19.5) predicted the presence of a potential Helix-hairpin-Helix (HhH) fold in members of the RAD52 superfamily. The HhH domain is a small nucleic acid-binding module comprised of two helices joined by a central loop (hairpin), which functions as the DNA-binding moiety of numerous repair and recombination proteins[27,28]. Two HhH modules are predicted in the core conserved domain of the RAD52 family, the first one bounded by the predicted helices 2 and 3, and the second one bounded by helices 5 and 6 (Fig. 5). Although these predicted HhH modules are very divergent in sequence from the typical versions, the hairpin in both HhH modules of the RAD52 family proteins is bounded by small residues, typically glycine; this conforms to the signature motif characteristic of the classical HhH modules [28,29]. However, in the case of the RAD52 superfamily the predicted HhH modules appear to have been welded into a large globular superstructure that maintained its evolutionary distinctness over time. The conservation pattern and predicted structural elements of the RAD52 superfamily are distinct from those predicted for the ERF and RecT/Red β superfamilies (Fig 1, 3, 5), supporting the lack of a direct evolutionary relationship between these proteins.

The RAD52 superfamily shows a sporadic phyletic distribution, and even in the crown-group eukaryotes, might have been secondarily lost in certain lineages, such as plants, nematodes and insects. The sporadic distribution of this family among phylogenetically distant bacteria, along with its presence in several prophages, suggests that, like the RecT/Red β and ERF superfamilies, at least the bacterial RAD52-proteins might be of predominantly phage origin. The core of the eukaryotic recombination system appears to have been inherited from the system present in the common ancestor shared with the archaea [29]. However, RAD52 is thus far absent in all archaeal genomes and is restricted to a single orthologous group in the eukaryotes [29]. Thus, it appears plausible that eukaryotic RAD52 was ultimately derived through lateral transfer either from a bacterial genome or directly from a viral source, at a point at least predating the divergence of the crown group eukaryotes and Entamoeba.

Contextual information from gene neighborhoods provides details regarding functional interactions of the SSAPs with DNA recombination pathways

The clustering of functionally related genes in prokaryotic genomes into co-transcribed and co-regulated units, operons, often allows functional assignments through the principle of 'guilt by association' [30–32]. Generally, genes whose products physically interact to form a complex or are involved in successive steps in a biochemical pathway form operons that are conserved over large evolutionary distances [30]. On previous occasions, we have used gene neighborhoods or operons to predict novel DNA repair complexes and their components [33]. Accordingly, a similar approach was applied to the three families of SSAPs (RecT/Redβ, ERF, Rad52), to shed light on their functional links.

Notably, the genes encoding the three evolutionarily distinct SSAPs co-occurred with similar sets of DNA repair/recombination-related proteins (Figs. 2,4). In at least one case, each of them was found adjacent to the gene for the single-strand-binding protein (SSB), an OB-fold protein that binds ssDNA (Figs. 2,4). This association ties in with the function of the SSAPs in single-strand annealing, suggesting that they closely interact with SSB. It has been suggested in the case of RecT that it may compete with SSB for binding single strand overhangs and thereby make them available for the annealing process [3]. Similar interactions between other SSAPs and SSB, that probably coats the ssDNA generated by nucleases, appear likely. Genes for SSAPs from all the 3 distinct superfamilies may also occur adjacent to or in the vicinity of genes encoding nucleases or Holliday junction resolvases (HJRs). Genes for RecT/Redß superfamily proteins are associated with genes encoding a λ -type exonuclease (LE) of the type II restriction enzyme fold, RecE, which also might be a divergent member of this fold, and a nuclease of the Endonuclease VII (EndoVII) fold [7,34] (Fig. 2). The ERF superfamily genes are associated with a RusA superfamily nuclease/ HJR and EndoVII fold nucleases (Fig. 4) [34]. Furthermore, the Borrelia plasmids that encode ERF, also almost always additionally encode a λ -type exonuclease, even if it is not the adjacent gene. In a single instance, in the Grampositive bacterium Ruminococcus albus, the gene encoding a RAD52 superfamily protein occurs adjacent to a gene for a λ -type exonuclease. These nucleases probably contribute to the repair process, in which SSAPs are involved, by providing the initial break in the dsDNA and/or in digesting the nicked target to generate ssDNA.

The RecT and Red β family proteins often co-occur with the SbcC gene that encodes an ABC ATPase with a large coiled-coil segment. These proteins are known to cooperate with SbcD, nuclease of the calcineurin-like phosphoesterase superfamily and to degrade dsDNA in the 3' \rightarrow 5' direction generating ssDNA [35,36]. It seems likely that RecT/Red β proteins, at least in certain cases, function in conjunction with the SbcCD-pathway, by utilizing the single-stranded regions generated by the SbcCD nuclease. Additionally, several genes, whose functions are less clear, tend to co-occur with the genes coding for the SSAPs. These include DNA methyltransferases and the primosomal protein DnaD from low-GC Gram-positive bacteria [37] that co-occur with both ERF and RecT superfamily members (Figs. 2,4). The poorly characterized phage-or prophage-specific genes that are frequently observed in these neighborhoods include ORF15 (*Streptococcus thermophilus* bacteriophage 7201), ORF86, ORF100a (*Staphylococcus aureus* temperate phage ϕ SLT) and ORF364 (bacteriophage ϕ 31.1) (Figs. 2,4). Secondary structure predictions indicate a high α -helical content for these proteins. It is likely that these α -helical proteins are phage innovations that could function as adaptors in the recombination pathway either as accessory protein-protein interacting domains or as DNA-binding domains.

Evidence for convergent operon evolution and in situ nonorthologous displacement of genes in operons encoding SSAPs

A superposition of the gene neighborhood information upon the phylogenetic trees for the SSAP superfamilies provides insights into the evolutionary processes that led to the emergence of the operons that include the SSAP genes. As discussed above, the RecT/Red β superfamily clearly splits into three distinct families (Fig. 2). The phylogenetic tree shows that SbcC co-occurs with the SSAP once within Redβ-family and once within the RecT-family. An examination of the tree and the respective gene neighborhoods suggests that independent juxtaposition of SbcC with Redß-like and RecT-like genes on two separate occasions is the most parsimonious explanation. The alternative explanation, namely that the gene coding for the common ancestor of the Redβ and RecT already co-occurred with the *sbcC* gene is far less likely because it would require over 10 independent losses of this, apparently, functionally advantageous organization in different bacterial and bacteriophage lineages. Likewise, the observation that, in one or more cases, genes encoding each of the SS-APs co-occur in the same predicted operon with SSB or a λ -type exonuclease, suggests that similar operon structures may also emerge independently in evolution. Thus, the same or analogous operon organizations may emerge convergently on multiple occasions, probably due to the selective pressure arising from the strong interactions between the SSAPs and their functional partners such as SbcC, SSB and LE.

The distribution of the gene neighborhoods, in which a member of the RecT superfamily occurs next to the RecE on the phylogenetic tree of the RecT/Red β superfamily, indicates that the RecE-RecT combination was probably the ancestral state for at least the RecT and EHAP1 families (Fig. 2). This implies that, on at least two occasions, the gene for λ endonuclease displaced the functionally analogous *recE* gene and became the adjacent gene to RecT (Fig. 2). That this displacement might have occurred by *in situ*

insertion of a non-orthologous gene is suggested by the detection, on three separate occasions, of unusual remnants of pre-existing genes. The RecT/Redß superfamily members, namely EHAP1 from the enterobacteria and PF161 from Borrelia hermsii, contain a small, C-terminal fragment of the core conserved domain of the ERF superfamily, which is located C-terminal of their bona fide RecT/ Redß domains. These fragments of the ERF protein are closely related to other ERF domains from related organisms and are unlikely to fold into the native conformation characteristic of the full-length ERF domain. For example the ERF fragment fused to the EHAP1 RecT/Red β domain is closely related to the P22 phage ERF domain. This suggests that in each of these cases a RecT superfamily gene was inserted in frame into a pre-existing ERF gene leaving behind only a non-functional fragment of it (Fig. 2). In a very similar case, the bacterial RAD52-like protein from a Shiga toxin encoding temperate phage is fused to an extreme C-terminal fragment that is nearly identical to the C-terminal most portion of the P22 ERF protein. In this case, it appears that the pre-existing ERF gene was displaced through the insertion of a bacterial RAD52-like gene. Interestingly, and in the same vein, the RecT proteins from Bacillus species contain a short C-terminal acidic module that is missing in other RecT proteins, but is highly similar to the C-terminal region of SSBs, particularly those from Gram-positive bacteria (data not shown). This suggests that, at some stage in their evolution, the Ba*cillus recT* gene protein has recombined with the gene coding for SSB, which might even have resulted in a functional replacement of an SSB with an SSAP.

Thus it appears likely that functionally equivalent genes may displace their analogs in operons via insertion into the same position.

Conclusions

We show that functionally similar SSAPs belong to at least three evolutionarily distinct superfamilies. We unify the Red β and RecT proteins and their homologs, which have not been reported as being related at the sequence level, into a single superfamily, supporting the notion that these proteins share a similar mechanism of action. The second superfamily typified by the ERF proteins is predominantly found in bacteriophages and is also present on all circular plasmids from Borrelia, suggesting a role in the recombination of these plasmids. The third superfamily, typified by the yeast RAD52 protein and previously detected only in eukaryotes, was shown to include bacterial and phage homologs and to contain a modified HhH domain. By comparing the gene neighborhoods of the SSAPs, we show that the predicted operons that include the SSAP genes evolve according to the "LEGO" principle. In these operons, the SSAP genes are linked to the genes for various DNAses and DNA repair related proteins, such as SSB

and SbcC, which implies functional connections between the encoded proteins. Evidence is presented of convergent emergence of similar SSAP-encoding operons in different lineages and of *in situ* non-orthologous displacement of functionally similar genes in these operons.

Materials and Methods

Sequence searches of the non-redundant (NR) and the unfinished genomes databases, were done using the gapped BLAST and PSI-BLAST programs [38]. Iterative PSI-BLAST searches used for in-depth sequence analysis were done with the profile inclusion cutoff expectation value (E value) set at 0.1. Multiple sequence alignments were generated using the T_Coffee program [39] and the output was adjusted using PSI-BLAST search results and secondary structure predictions, which were conducted using the PHD [40,41] and Jpred [42] programs. Fold predictions were done using the 3-D position specific score matrix (3DPSSM) [43] and the Hybrid fold method [44]. Phylogenetic analysis was carried out using the neighbor-joining algorithm, with subsequent local rearrangements using the maximum likelihood algorithm [45]. The robustness of tree topology was assessed with 10000 Resampling of Estimated Log Likelihoods (RELL) bootstrap replicates. The MOLPHY and Phylip software packages were used for the analyses [46,47].

Authors' contributions

Author 1 (LMI) contributed to the discovery process, preparation of the manuscript and multiple sequence alignments. Author 2 (EVK) contributed to the analysis of the predicted operons encoding RecT and SSB proteins from Gram positive bacteria, Author 3 (LA) contributed to the discovery process, preparation of the manuscript and conceived the study.

References

- Roca Al, Cox MM: RecA protein: structure, function, and role in recombinational DNA repair. Prog Nucleic Acid Res Mol Biol 1997, 56:129-223
- Bianco PR, Tracy RB, Kowalczykowski SC: DNA strand exchange proteins: a biochemical and physical comparison. Front Biosci 1998, 3:D570-603
- Kuzminov A: Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. Microbiol Mol Biol Rev 1999, 63:751-813
- Noirot P, Kolodner RD: DNA strand invasion promoted by Escherichia coli RecT protein. J Biol Chem 1998, 273:12274-12280
- Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM: Biochemistry of homologous recombination in Escherichia coli. Microbiol Rev 1994, 58:401-465
- 6. Paques F, Haber JE: Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 1999, 63:349-404
- 7. Chang HW, Julin DA: Structure and function of the Escherichia coli RecE protein, a member of the RecB nuclease domain family. J Biol Chem 2001, 276:46004-46010
- Barbour SD, Nagaishi H, Templin A, dark AJ: Biochemical and genetic studies of recombination proficiency in Escherichia coli. II. Rec+ revertants caused by indirect suppression of recmutations. Proc Natl Acad Sci U SA 1970, 67:128-135

- Kushner SR, Nagaishi H, Templin A, dark AJ: Genetic recombination in Escherichia coli: the role of exonuclease I. Proc Natl Acad Sci USA 1971, 68:824-827
- Kolodner R, Hall SD, Luisi-DeLuca C: Homologous pairing proteins encoded by the Escherichia coli recE and recT genes. Mol Microbiol 1994, 11:23-30
- 11. Hall SD, Kane MF, Kolodner RD: Identification and characterization of the Escherichia coli RecT protein, a protein encoded by the recE region that promotes renaturation of homologous single-stranded DNA. J Bacteriol 1993, 175:277-287
- Hall SD, Kolodner RD: Homologous pairing and strand exchange promoted by the Escherichia coli RecT protein. Proc Natl Acad Sci USA 1994, 91:3205-3209
- Mortensen UH, Bendixen C, Sunjevaric I, Rothstein R: DNA strand annealing is promoted by the yeast Rad52 protein. Proc Natl Acad Sci USA 1996, 93:10729-10734
 Passy SI, Yu X, Li Z, Radding CM, Egelman EH: Rings and filaments
- 14. Passy SI, Yu X, Li Z, Radding CM, Egelman EH: Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. Proc Natl Acad Sci U S A 1999, 96:4279-4284
- Thresher RJ, Makhov AM, Hall SD, Kolodner R, Griffith JD: Electron microscopic visualization of RecT protein and its complexes with DNA. J Mol Biol 1995, 254:364-371
- Shinohara A, Shinohara M, Ohta T, Matsuda S, Ogawa T: Rad52 forms ring structures and co-operates with RPA in singlestrand DNA annealing. Genes Cells 1998, 3:145-156
- Poteete AR, Sauer RT, Hendrix RW: Domain structure and quaternary organization of the bacteriophage P22 Erf protein. J Mol Biol 1983, 171:401-418
- Stevenson B, Porcella SF, Oie KL, Fitzpatrick CA, Raffel SJ, Lubke L, Schrumpf ME, Schwan TG: The relapsing fever spirochete Borrelia hermsii contains multiple, antigen-encoding circular plasmids that are homologous to the cp32 plasmids of Lyme disease spirochetes. Infect Immun 2000, 68:3900-3908
- Stroeher UH, Bode L, Beutin L, Manning PA: Characterization and sequence of a 33-kDa enterohemolysin (Ehly I) – associated protein in Escherichia coli. Gene 1993, 132:89-94
- Kmiec E, Holloman WK: Beta protein of bacteriophage lambda promotes renaturation of DNA. J Biol Chem 1981, 256:12636-12639
- 21. Botstein D, Matz MJ: A recombination function essential to the growth of bacteriophage P22. J Mol Biol 1970, 54:417-440
- 22. Weaver S, Levine M: Recombinational circularization of Salmonella phage P22 DNA. Virology 1977, 76:29-38
- 23. Poteete AR: Location and sequence of the erf gene of phage P22. Virology 1982, 119:422-429
- 24. Poteete AR, Fenton AC: Lambda red-dependent growth and recombination of phage P22. Virology 1984, 134:161-167
- Poteete AR, Fenton AC: Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages lambda and P22. Genetics 1993, 134:1013-1021
- Stasiak AZ, Larquet E, Stasiak A, Muller S, Engel A, Van Dyck E, West SC, Egelman EH: The human Rad52 protein exists as a heptameric ring. Curr Biol 2000, 10:337-340
- 27. Shao X, Grishin NV: Common fold in helix-hairpin-helix proteins. Nucleic Acids Res 2000, 28:2643-2650
- Doherty AJ, Serpell LC, Ponting CP: The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. Nucleic Acids Res 1996, 24:2488-2497
- Aravind L, Walker DR, Koonin EV: Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Res 1999, 27:1223-1242
- Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV: Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res 2001, 11:356-372
- Huynen M, Snel B, Lathe W 3rd, Bork P: Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res 2000, 10:1204-1210
- Dandekar T, Snel B, Huynen M, Bork P: Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 1998, 23:324-328
- 33. Aravind L, Koonin EV: Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku pro-

tein and prediction of a prokaryotic double-strand break repair system. Genome Res 2001, 11:1365-1374

- Aravind L, Makarova KS, Koonin EV: Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res 2000, 28:3417-3432
- Connelly JC, Kirkham LA, Leach DR: The SbcCD nuclease of Escherichia coli is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. Proc Natl Acad Sci U S A 1998, 95:7969-7974
- Connelly JC, de Leau ES, Leach DR: DNA cleavage and degradation by the SbcCD protein complex from Escherichia coli. Nucleic Acids Res 1999, 27:1039-1046
- Bruand C, Ehrlich SD, Janniere L: Primosome assembly site in Bacillus subtilis. Embo J 1995, 14:2642-2650
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-3402
- Notredame C, Higgins DG, Heringa J: T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000, 302:205-217
- Rost B, Schneider R, Sander C: Protein fold recognition by prediction-based threading. J Mol Biol 1997, 270:471-480
 Rost B, Sander C, Schneider R: PHD – an automatic mail server
- Rost B, Sander C, Schneider R: PHD an automatic mail server for protein secondary structure prediction. Comput Appl Biosci 1994, 10:53-60
- Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ: JPred: a consensus secondary structure prediction server. Bioinformatics 1998, 14:892-3
- Kelley LA, MacCallum RM, Sternberg MJ: Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 2000, 299:499-520
- 44. Fischer D: Hybrid fold recognition: combining sequence derived properties with evolutionary information. In: Pacific Symposium on Biocomputing, Hawaii 2000, 119-130
- Wolf YI, Rogozin IB, Grishin NV, Tatusov RL, Koonin EV: Genome trees constructed using five different approaches suggest new major bacterial clades. BMC Evol Biol 2001, 1:8
- Adachi J, Hasegawa M: MOLPHY: Programs for Molecular Phylogenetics. Tokyo: Institute of Statistical Mathematics; 1992
- Felsenstein J: Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* 1996, 266:418-427

