

VILNIAUS UNIVERSITETAS

Darius Kazlauskas

DVIGRANDĖS DNR VIRUSŲ DNR REPLIKACIJOS
BALTYMŲ ANALIZĖ KOMPIUTERINIAIS
METODAIS

Daktaro disertacija
Fiziniai mokslai, biochemija (04 P)

Vilnius, 2014

Disertacija rengta 2010-2014 m. Vilniaus universiteto Biotechnologijos institute.

Mokslinis vadovas:

dr. Česlovas Venclovas (biomedicinos mokslai; biologija – 01 B ir fiziniai mokslai, biochemija – 04 P)

TURINYS

SANTRUMPOS.....	6
ĮVADAS	7
1. LITERATŪROS APŽVALGA	10
1.1. DNR replikacija	10
1.2. DNR replikacijos baltymai	12
1.2.1. DNR replikazės.....	12
1.2.2. Helikazės.....	18
1.2.3. Praimazės	21
1.2.4. Viengrandę DNR surišantys baltymai	22
1.2.5. Pradmenį pašalinantys baltymai	23
1.2.6. Ligazės	24
1.2.7. Topoizomerazės	25
2. METODAI	27
2.1. Pradiniai duomenys ir duomenų bazės	27
2.2. Replikacijos baltymų paieška	27
2.3. Baltymų sekų grupavimas.....	29
2.4. Replikacinių baltymų identifikavimas	30
2.5. Papildomos homologų paieškos	30
2.6. Sekų palyginiai	31
2.7. Antrinės struktūros ir nestruktūrizuotų regionų nustatymas.....	31
2.8. Genomų lyginimas ir genominės aplinkos analizė	31
2.9. Genomų grupavimas ir filtravimas	31
2.10. Homologinis modeliavimas ir struktūros analizė	32
2.11. Elektrostatinių savybių analizė	32

3. REZULTATAI IR JŲ APTARIMAS	34
3.1. DNR replikazių analizė.....	34
3.1.1. DNR replikazės komponentai ir genomo dydis.....	34
3.1.2. DNR polimerazės.....	38
3.1.1. Procesyvumo veiksniai	43
3.1.2. Žiedo užkėlimo kompleksai.....	49
3.1.3. DNR replikazių tyrimo rezultatų apibendrinimas ir sąsajos su ląsteliniiais organizmais.....	52
3.2. DNR replikacijos baltymų ir viruso genomo dydžio bei šeimininko taksonomijos ryšys.....	56
3.3. Replikacinių helikazių ir praimazių analizė	59
3.3.1. Replikacinių helikazių ir praimazių identifikavimas, įvairovė ir domeninė organizacija	60
3.3.2. Replikacinės helikazės – dažniausias dgDNR virusų replikacijos baltymas	63
3.3.3. Herpes virusų helikazės-praimazės UL8 subvienetas yra inaktyvuota B šeimos polimerazė.....	65
3.4. Viengrandę DNR surišantys baltymai	73
3.4.1. Keturiuose NCLDV šeimose rasti OB-sanklodos SSB baltymai	75
3.4.2. NCLDV SSB struktūra tinkama sąveikai su vgDNR	77
3.4.3. Raupų SSB baltymai (I3) negiminingi NCLDV SSB	80
3.4.4. NCLDV vgDNR surišančių baltymų taksonominis pasiskirstymas ir evoliucija.....	83
3.5. Pradmens pašalinimo baltymai	84
3.6. Ligazės	85
3.7. Topoizomerazės	85

IŠVADOS	86
MOKSLINIŲ DARBŲ SĄRAŠAS	87
FINANSINĖ PARAMA	88
GYVENIMO APRAŠYMAS	89
PADĖKA	90
LITERATŪROS SĄRAŠAS	91

SANTRUMPOS

AEP – archėjų-eukariotų praimazė
a-r. – aminorūgštis
bp – bazių pora; kbp – 10^3 bp; Mb – 10^6 bp
DB – duomenų bazė
dg – dvigrandė
DUF – nežinomos funkcijos domenas (domain of unknown function)
NCBI – JAV nacionalinis biotechnologinės informacijos centras (National Center for Biotechnology Information)
NCLDV – Branduolyje ir citoplazmoje besidauginantys didelio genomo DNR virusai (Nuclear and Cytoplasmic Large DNA Viruses)
NTP – nukleotidų trifosfatas
nr – NCBI nepasikartojančių baltymų sekų (non-redundant) duomenų bazė
nr70 – „nr“ duomenų bazė nufiltruota iki 70% identiškumo
OB – sąveikaujantis su oligonukleotidais (oligonucleotide-binding)
Pfam – baltymų šeimų (protein families) duomenų bazė
PCNA – besidauginančių ląstelių branduolio antigenas (Proliferating Cell Nuclear Antigen)
PV – procesyvumo veiksnys
RFC – replikacijos veiksnys C (Replication Factor C)
RFCL – replikacijos veiksnio C didysis subvienetas (Replication Factor C, large)
RFCS – replikacijos veiksnio C mažasis subvienetas (Replication Factor C, small)
SF – superšeima (superfamily)
SmpB – mažasis baltymas B (Small protein B) – bakterijų baltymas, surišantis tmRNR
SSB – viengrandę DNR surišantis baltymas (single-stranded DNA-binding)
Toprim – topoizomerazės-praimazės domenas
UDG – uracil-DNR glikozilazė
vg – viengrandė
ŽUK – žiedo užkėlimo kompleksas

ĮVADAS

Sugebėjimas daugintis ir evoliucionuoti yra pagrindinės gyvų esybių ypatybės. Genetinės informacijos kopijavimą atlieka replikacijos baltymai. Laisvai gyvenančių ląstelių organizmų replikacijos aparato sandara yra vienoda. Skiriasi tik atskirų šios sistemos komponentų kilmė. Pavyzdžiui, sekų ir struktūrų analizės rodo, kad eukariotų-archėjų ir bakterijų DNR polimerazės yra negiminingos (Leipe, Aravind et al., 1999; Bailey, Wing et al., 2006). Tuo tarpu virusų DNR replikacijos baltymai yra labai įvairūs. Jie gerai ištirti T7, T4 faguose, herpes, poliomos ir papilomos virusuose, tačiau šios grupės sudaro tik mažą dalį (~10%) visos tyrinėtos jų įvairovės. Kaip dauginasi mažiau žinomi dgDNR virusai? Ar jie naudoja jau žinomų replikacijos sistemų variacijas? O gal jiems būdingos specifinės, dar nežinomos DNR replikacijos strategijos? DgDNR virusai stebina ne tik savo įvairove, bet ir genomo dydžio svyravimu. Pavyzdžiui, mažiausių šios grupės virusų (poliomos) genomai yra 500 kartų mažesni už didžiausių Pandora virusų (genomas – 2500 kbp). Laisvai gyvenančių ląstelių organizmų genomo dydis taip pat varijuoja. Pavyzdžiui, mažiausio eukarioto (*Ostreococcus tauri*) ir žmogaus genomo dydis skiriasi ~260 kartų (Derelle, Ferraz et al., 2006). Nepaisant to, šių organizmų DNR replikacijos aparatas turi tuos pačius komponentus. Ar jie vienodi ir dgDNR virusuose? O galbūt replikacijos baltymų įvairovė ir pasiskirstymas genomuose priklauso nuo viruso genomo dydžio?

Siekiant atsakyti į aukščiau pateiktus klausimus, buvo atlikta išsami dgDNR virusų DNR replikacijos baltymų kompiuterinė analizė. Naudojant pažangiausius bioinformatikos metodus, identifikuoti ir charakterizuoti virusų replikacijos baltymai (DNR polimerazės, procesyvumo veiksniai, žiedo užkėlimo kompleksai, DNR praimazės, helikazės, viengrandę DNR surišantys baltymai, pradmens pašalinime dalyvaujanti nukleazės, DNR ligazės ir topoizomerazės) bei nustatyti jų pasiskirstymo genomuose dėsniumai.

DgDNR virusų replikacijos baltymų tyrimas vyko dviem etapais. Iš pradžių buvo tiriamos DNR replikazės (DNR polimerazės, procesyvumo

veiksniai, žiedo užkėlimo kompleksai). Nustatyta, kad šios sistemos komponentų buvimas genome ir jų tipai priklauso nuo viruso genomo dydžio. Maži virusai (<40 kbp) dažnai neturi DNR polimerazių arba turi B-šėimos polimerazę, pradmeniu naudojančią baltymą. Didėjant genomui virusai vis dažniau koduoja nuosavą DNR polimerazę, o patys didžiausi atstovai (>140 kbp) daugeliu atvejų papildomai turi procesyvumo veiksnį ir kartais jo užkėlėją. Vienintelė šios taisyklės išimtis buvo phiKZ fagai. Ji buvo pašalinta šiuose dideliuose virusuose radus B-šėimos DNR polimerazes. Nustačius DNR replikazės komponentų ir viruso genomo dydžio ryšį, buvo nuspręsta patikrinti ar minėtas dėsniumas galioja ir kitiems replikacijos baltymams. Paaiškėjo, kad didėjant genomui virusai dažniau koduoja nuosavus replikacijos baltymus. Tikrinant atvejus, nepaklūstančius šiai taisyklei, naujai rasti vgDNR surišantys baltymai didžiausiuose eukariotų virusuose (NCLDV). Nustebino tai, kad šie baltymai yra giminingi bakteriofago T7 SSB. Taip pat netikėta buvo nustatyti, kad replikacinė helikazė yra dažniausias virusų replikacijos baltymas. Ji buvo rasta net 70% visų atvejų. Detaliau analizuojant herpes virusų helikazę-praimazę nustatyta, kad mažiausiai charakterizuotas jos komponentas (UL8) yra inaktyvuota B-šėimos DNR polimerazė.

Darbo tikslas. Ištirti dvigrandės DNR virusų replikacijos baltymus kompiuteriniais metodais.

Uždaviniai:

1. Kompiuteriniais metodais analizuojant virusų genomų ir baltymų duomenų bazes identifikuoti ir charakterizuoti DNR replikacijos baltymus.
2. Nustatyti, ar yra ryšys tarp DNR replikacijos baltymų ir viruso genomo dydžio ir jei taip, pabandyti atskleisti šio ryšio dėsniumus.

Mokslinis naujumas. Šis darbas yra pirma tokios didelės apimties analizė, kurioje ištirti visų žinomų dvigrandės DNR virusų replikacijos baltymai. Analizės metu nustatytas ryšys tarp viruso replikacijos baltymų ir genomo dydžio. Naujai atrasti ir charakterizuoti SSB baltymai NCLD virusuose, DNR

polimerazės phiKZ faguose. Nustatytas ir detaliai aptartas raupų viruso SSB (I3) panašumas į *E. coli* SmpB baltymą bei herpes virusų helikazės-praimazės komponento UL8 panašumas į B šeimos DNR polimerazes.

Praktinė nauda. Šio tyrimo metu nustatytas replikacijos baltymų ir genomo dydžio ryšys leidžia prognozuoti naujai sekvenuojamų virusų replikacijos aparato pilnumą (dideli genomai – pilnesnis replikacijos aparatas). Be to, dgDNR virusai yra dažni ligų sukėlėjai (herpes, papilomos, poliomos, raupų, adeno virusai), bei žemės ir vandens ūkių kenkėjai (afrikinio kiaulių maro, krevečių virusai), todėl jų tyrimai svarbūs ligų gydymui bei kovai su kenkejais. Naujai rastos DNR polimerazės *Pseudomonas aeruginosa* phiKZ faguose bei raupų viengrandę DNR surišančio baltymo (I3) ir herpes viruso helikazės-praimazės komplekso baltymo (UL8) charakterizavimas kompiuteriniais metodais galėtų padėti planuojant šių baltymų eksperimentinius tyrimus.

Ginamieji disertacijos teiginiai:

1. Replikazės komponentų (ne)buvimas genome ir jų tipai priklauso nuo viruso genomo dydžio: maži virusai (<40 kbp) dažnai neturi DNR polimerazių arba turi B-šėimos polimerazę, kaip pradmenį naudojančią baltymą. Didėjant genomui virusai vis dažniau koduoja nuosavą DNR polimerazę, o patys didžiausi atstovai (>140 kbp) dažnai papildomai turi procesyvumo veiksnį ir kartais jo užkėlėją.
2. PhiKZ fagai turi stipriai divergavusias B-šėimos DNR polimerazes.
3. Didėjant viruso genomui jame dažniau koduojami nuosavi replikacijos baltymai, o pats dažniausias virusų replikacinis baltymas yra helikazė.
4. Herpes virusų helikazės-praimazės subvienetas (UL8) yra inaktyvuota B šeimos DNR polimerazė.
5. Didžiausi eukariotų virusai (NCLDV) turi bent dvi evoliuciškai skirtingas SSB šeimas, negiminingas eukariotų baltymams.

1. LITERATŪROS APŽVALGA

1.1. DNR replikacija

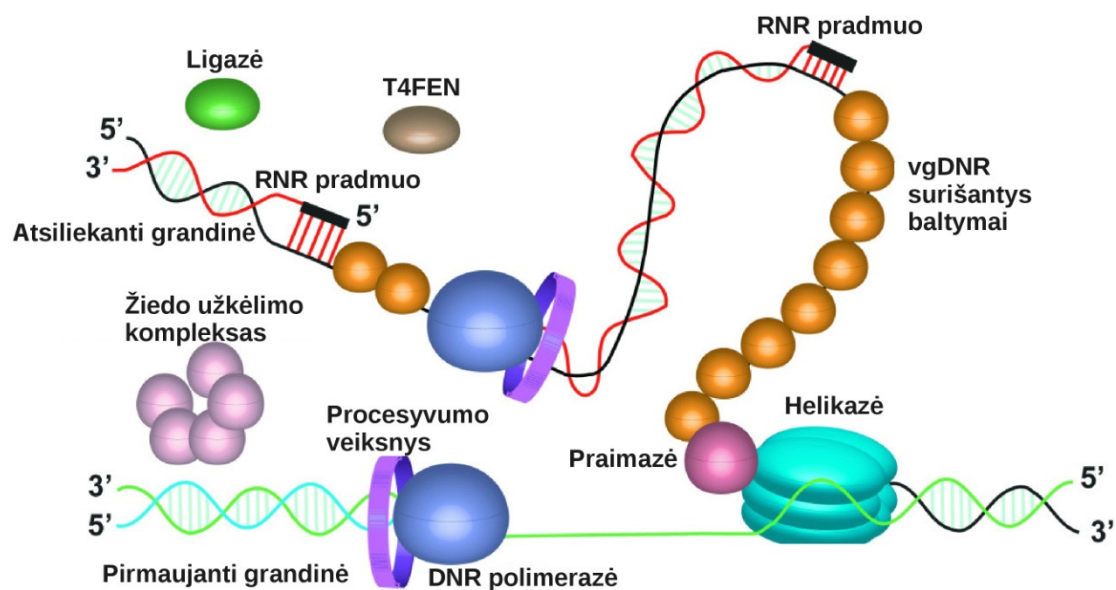
Genetinės informacijos kopijavimas būdingas visiems organizmams. Šiame procese dalyvauja daug tarpusavyje sąveikaujančių baltymų. DNR sintezės pradžiai replikacinės DNR polimerazės naudoja pradmenis. Pagal naudojamą pradmenį DNR replikaciją galima suskirstyti į du tipus: pradmens-baltymo ir pradmens-RNR (Lentelė 1.1). Pirmasis tipas būdingas kai kuriems virusams ir mažoms linijinėms bakterijų plazmidėms, o pradmens-RNR replikacija – ląsteliniams organizmams ir daugeliui virusų. Abiems tipams reikalingos DNR polimerazės ir vgDNR surišantys baltymai (SSB), tačiau kiti replikacijos baltymai skiriasi. Pradmens-baltymo atveju, prie geno 5' galų būna prijungtas terminalinis baltymas. Replikacija prasideda nuo abiejų geno galų, kuomet prie terminalinio baltymo serino/tirozino/treonino a-r. hidroksilo grupės DNR polimerazė prijungia nukleotidą (Salas, 1991). Replikacijos metu susidaranti vgDNR nuo pažeidimų apsaugoma ją padengiant SSB baltymais. DNR sintezė baigiama abejoms polimerazėms pasiekus geno galus.

Pradmens-baltymo atveju DNR sintetinama grandinės išstūmimo, o pradmens-RNR – dažnai replikacinės šakutės mechanizmu. Pastarasis yra sudėtingesnis. Vienu metu vykstant abiejų motininių grandinių komplementariam kopijavimui susidaro Y-formos struktūra, vadinama replikacine šakute (1.1 pav.). Kadangi replikacinės DNR polimerazės gali sintetinti tik 5'-3' kryptimi, viena grandinė (pirmaujanti) yra sintetinama nuolatos, o kita (atsilieanti) – fragmentiškai (1.1 pav.). Replikacinės šakutės atveju, DNR replikazę dažnai sudaro ne tik DNR polimerazė, bet ir jos sintezės efektyvumą užtikrinantys baltymai – procesyvumo veiksnys ir jo užkėlėjas (1.1 pav.). Šio tipo replikacijos pradžioje DNR helikazė išskiria motinines grandines, vgDNR padengiama SSB baltymais, o praimazė susintetina RNR pradmenį, nuo kurio DNR polimerazė pradeda DNR sintezę. Vėliau pradmuo pašalinamas nukleazėmis, o susidariusį tarpą užtaiso DNR ligazės (1.1 pav.).

Topoizomerazės, keisdamos DNR topologiją, pašalina įtampą, susidariusią priešais replikacijos šakutę. Tolimesniuose skyriuose visi minėti replikacijos baltymai yra aptariami detaliau.

Lentelė 1.1 DNR replikacijos tipai

Pradmuo-baltymas	Pradmuo-RNR
<u>Bendri baltymai</u>	
DNR polimerazė	
vgDNR surišantis baltymas	
<u>Specifiniai baltymai</u>	<u>Specifiniai baltymai</u>
Terminalinis baltymas	Helikazė
	Procesyumo veiksnys ir jo užkėlėjas
	Praimazė
	Pradmens pašalinimo baltymas



1.1 pav. T4 fago DNR replikacijos baltymai (adaptuota iš (Mueser, Hinerman et al., 2010)).

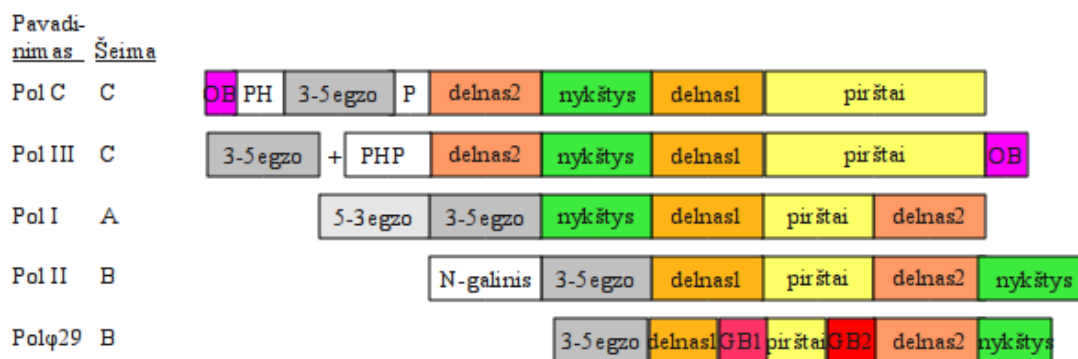
1.2. DNR replikacijos baltymai

1.2.1. DNR replikazės

Visų gyvų organizmų ir kai kurių virusų DNR sintezei naudojamos dažnai daugiakomponentės baltymų mašinos – DNR replikazės. Jos dažniausiai sudarytos iš DNR polimerazės, procesyvumo veiksnio ir žiedo užkėlimo komplekso.

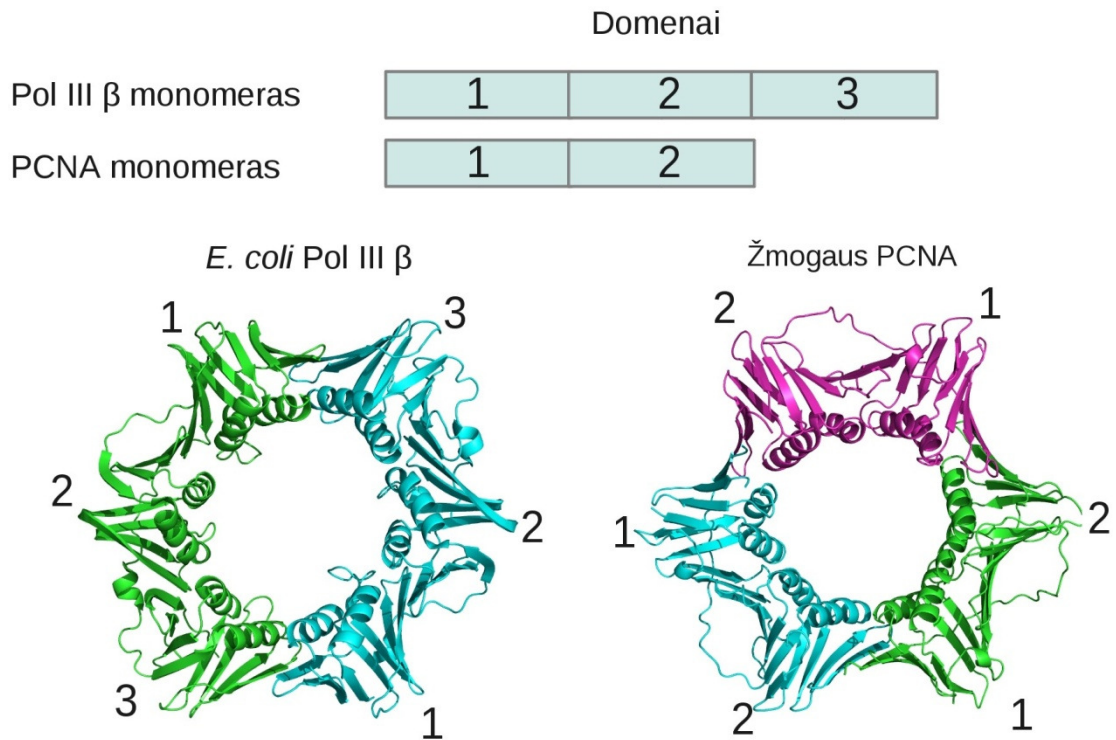
DNR polimerazės. DNR polimerazės yra fermentai sintetinantys DNR grandinę 5'-3' kryptimi. Šie baltymai dalyvauja ne tik DNR replikacijoje, bet ir reparacijos bei rekombinacijos procesuose. Remiantis tarpusavio panašumu išskiriamos šešios polimerazių šeimos (A, B, C, D, X, Y) (Rothwell ir Waksman, 2005). X ir Y šeimų atstovai dalyvauja DNR reparacijoje. Bakterijų DNR replikuoja C šeimos polimerazės. Jos negiminingos kitoms replikacinėms polimerazėms (Leipe, Aravind et al., 1999). Visiems eukariotams ir archėjoms būdingos B šeimos replikacinės polimerazės. D-šeimos polimerazės randamos tik archėjose (išskyrus *Crenarchaea*) (Sarmiento, Long et al., 2014). Virusų genomus dažniausiai replikuoja A ir B šeimų polimerazės, o C šeimos atstovai juose aptinkami retai (Koonin, 2006).

Replikacinių polimerazių iš A, B ir C šeimų struktūros ir domeninė organizacija yra žinomos (1.2 pav). Jų polimerazinis domenas sudarytas iš delno, nykščio ir pirštų subdomenų. Replikacijos tikslumą užtikrina 5'-3' egzozonukleazė. Beveik visos bakterinės C-šeimos polimerazės turi PHP ir OB domenus (Timinskas, Balvočiūtė et al., 2014), kurie atlieka atitinkamai 5'-3' egzozonukleazės ir viengrandės DNR surišimo funkcijas. Polimerazės, pradmeniu naudojančios baltymą (PolBp), papildomai turi terminalinio baltymo surišimo domenus (1.2 pav., GB1 ir GB2). Minėtas baltymas būna kovalentiškai prijungtas prie 5' genomo galų. Replikacija prasideda kuomet prie terminalinio baltymo serino/tirozino/treonino a-r. hidroksilo grupės prijungiamas nukleotidas (Salas, 1991).



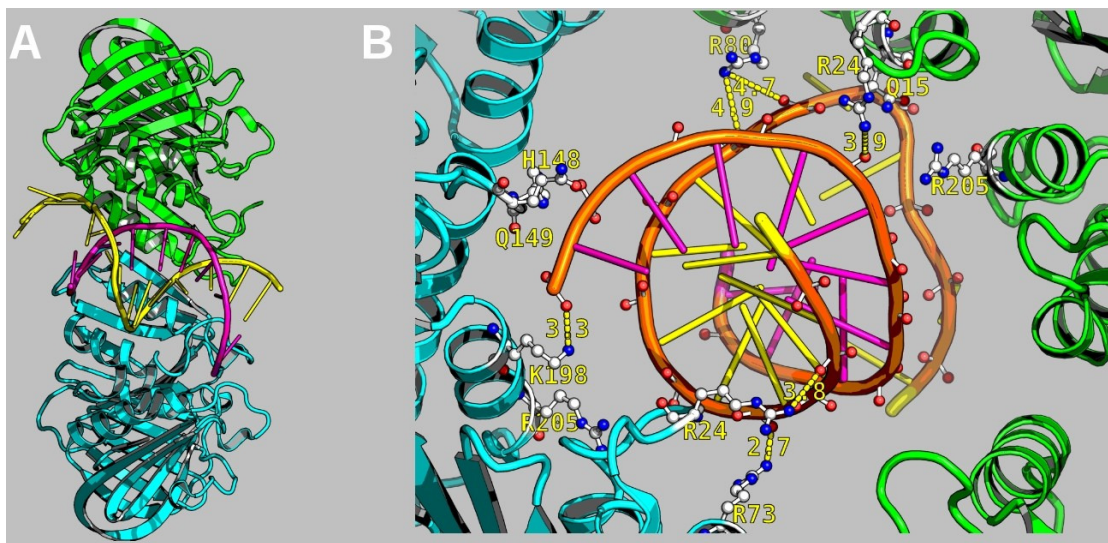
1.2 pav. Replikacinių DNR polimerazių domeninė organizacija

Procesyvumo veiksniai. Replikacijos efektyvumui užtikrinti visos ląstelinės replikacinės DNR polimerazės naudoja procesyvumo veiksnius (PV). Šie pagalbiniai baltymai neleidžia DNR polimerazei disocijuoti nuo DNR, tokiu būdu žymiai padidindami sintezės greitį (Johnson ir O'Donnell, 2005). Ląstelinių organizmų PV formuoja žiedą sudarytą iš šešių domenų (1.3 pav.). Eukariotams ir archėjoms būdingas trimerinis PCNA (Proliferating Cell Nuclear Antigen – besidauginančių ląstelių branduolio antigenas), bakterijoms – dimerinis PolIII β subvienetas (toliau - β subvienetas) (1.3 pav.). Šių baltymų aminorūgščių sekos labai skirtingos (identiškumas <15%), tačiau struktūros panašios (Kelman, 1997). Žinoma, kad PV dalyvauja ne tik DNR replikacijoje, bet, sąveikaudami su įvairiais baltymais, taip pat atlieka rekombinacijos, DNR reparacijos, ląstelės ciklo kontrolės funkcijas (Maga ir Hubscher, 2003). Pavyzdžiui, žmogaus DNR pažaidų taisyme dalyvauja specializuotas PV, kurį sudaro trys monomerai: Rad9, Hus1 ir Rad1 (Venclovas ir Thelen, 2000).



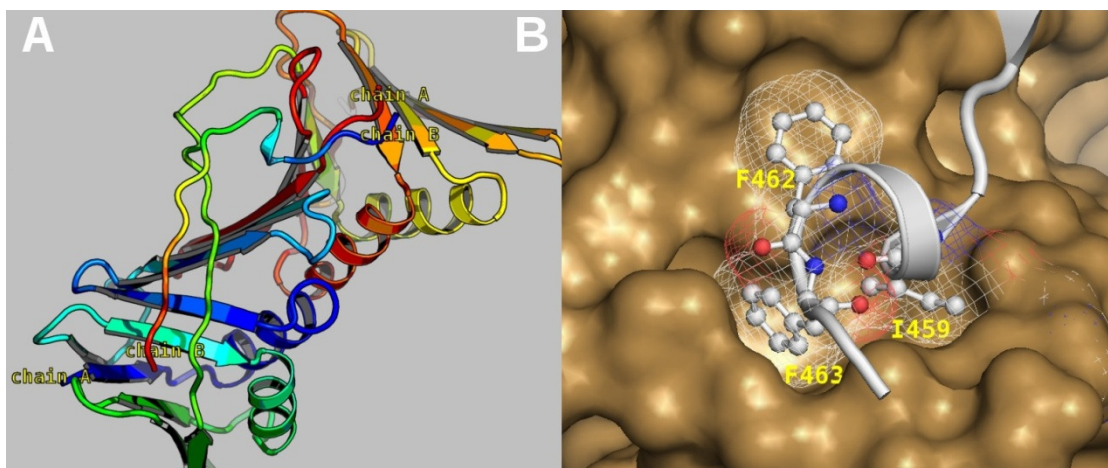
1.3 pav. PCNA ir β subvieneto domeninės organizacijos palyginimas.

Procesyvumo veiksniai sąveikauja su DNR ir baltymais. Sąveika su DNR yra nespecifinė (1.4 pav.). Vidinėje žiedo dalyje daugiau teigiamai įkrautų a-r., o žiedo išorėje – neigiamų. Manoma, kad teigiamas krūvis stabilizuoja sąveiką su DNR (Kelman ir O'Donnell, 1995). Priešingų krūvių santykis lemia mažą ląstelinių PV izoelektrinį tašką (pvz., PCNA – 4,5 (Delhon, Tulman et al., 2006)).



1.4 pav. β subvieneto sąveika su DNR. (A) β subvieneto struktūra kartu su DNR (PDB kodas: 3bep). (B) Aminorūgštys, dalyvaujančios sąveikoje, ir kontaktų ilgiai (Å) (sąveikaujančios a-r. pateiktos pagal (Georgescu, Kim et al., 2008)).

PV su kitais baltymais dažniausiai sąveikauja per išorinėje žiedo dalyje esančią hidrofobinę kišenę (1.5 pav.). Geriausiai ištirtas eukariotinių baltymų sąveikos su PCNA motyvas. Jis dar vadinamas PIP (PCNA-interacting peptide – su PCNA sąveikaujantis baltymas) sritimi. Ji dažnai būna baltymo C-gale ir yra sudaryta iš a-r. – Qxx(M/L/I)xxF(Y/F) (Bruning ir Shamoo, 2004). PIP nepolinės liekanos, formuojančios 3^{10} α spiralę, lyg kištukas įsitvirtina hidrofobiniame lizde (1.5 pav., B).



1.5 pav. PCNA ir baltymų sąveika. (A) Žmogaus PCNA (chain A) ir p21 PIP motyvo (chain B) kompleksas (PDB kodas: 1axc). Molekulės nudažytos vaivorykštės spalvomis (N-galas – mėlyna, C-galas – raudona). (B) Žmogaus PCNA hidrofobinė kišenė ir joje esantis p66 baltymo C-galo fragmentas (PDB kodas: 1u76).

Virusai pasižymi didele PV įvairove. T4 fago ir herpes virusų PV struktūriškai panašūs į ląstelių organizmų PV (Wiederstein, Gruber et al., 2014). T4 gp45 formuoja žiedą, o herpes virusų UL42 ir UL44 veikia atitinkamai kaip monomeras ir homodimeras (Lentelė 1.2). UL42 ir UL44 būdingas padidėjęs teigiamas krūvis sąveikos su DNR pusėje (Appleton, Loregian et al., 2004). Raupų virusų polimerazės procesyvumą užtikrina kompleksas sudarytas iš uracil-DNR glikozilazės ir A20 baltymo. Deja, pilna A20 struktūra ir domeninė organizacija nežinoma (Contesto-Richefeu, Tarbouriech et al., 2014). T7 fagas kaip procesyvumo faktorių naudoja *E. coli* tioredoksiną. Šis baltymas jungiasi prie polimerazės nykščio subdomeno ir manoma, kad užlinksta ant DNR, neleisdamas jai disocijuoti (Akabayov, Akabayov et al., 2010).

Lentelė 1.2 Procesyvumo veiksmų įvairovė ir sąlybės.

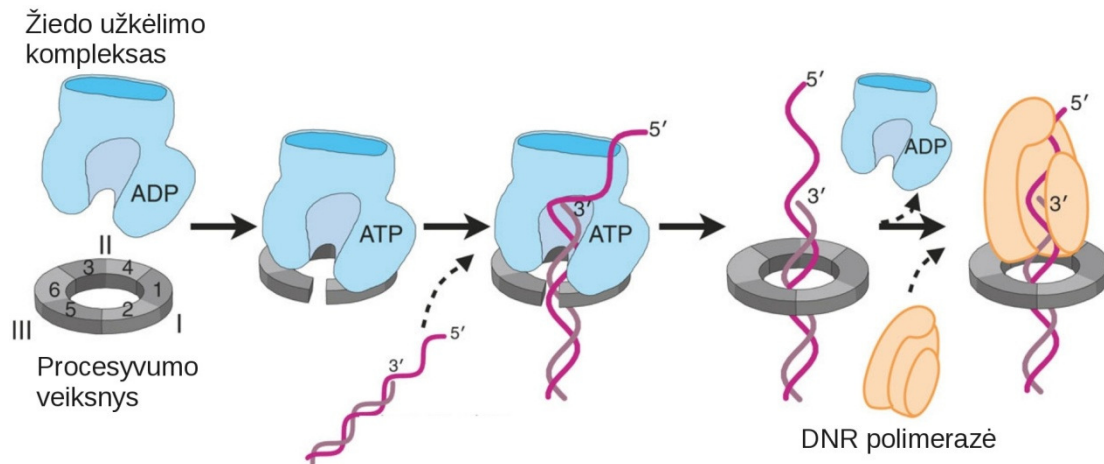
Procesyvumo veiksnys	Organizmas	4'inė struktūra	Sąveikos su PV motyvas*
PCNA	Eukariotai, archėjos, kai kurie virusai	homotrimeras	Qxx(M/L/I)xxF(Y/F) ¹
PolIII β	Bakterijos	homodimeras	Qxxhh ²
GP45	T4 fagas	homotrimeras	SLDFLF ³
UL42	Žmogaus herpes virusas 1	monomeras	RLRTAGFGAVGA ⁴
UL44	Žmogaus herpes virusas 5	homodimeras	LEPAFLPYSVKA ⁵
BMRF1	Žmogaus herpes virusas 7	?	?
PCNA123	<i>Crenarchaea</i>	heterotrimeras	Qxxhxxaa ⁶
Rad9-Hus1-Rad1	Eukariotai	heterotrimeras	n.d.

*x – bet kokia a-r., h – hidrofobinė a-r., a – aromatinė a-r., pabrauktos a-r. sudaro hidrofobinį „kištuką“

¹(Zhuang ir Ai, 2010); ²(Dalrymple, Kongsuwan et al., 2001); ³(Berdis, Soumillion et al., 1996); ⁴(Zuccola, Filman et al., 2000); ⁵(Appleton, Brooks et al., 2006); ⁶(Imamura, Fukunaga et al., 2007)

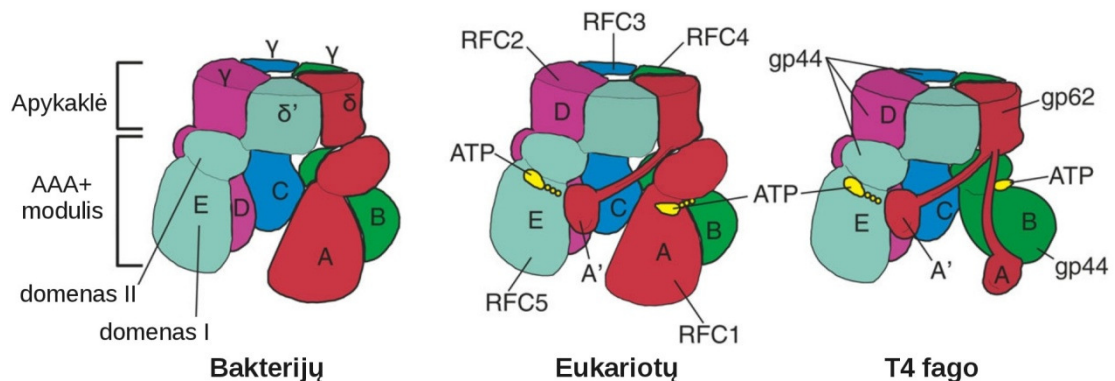
Žiedo užkėlimo kompleksai. Ląstelių organizmų ir T4 fagu procesyvumo veiksniai formuoja žiedą. Šių PV užkėlimui ant DNR naudojamas pentamerinis baltymas – žiedo užkėlimo kompleksas (ŽUK). ŽUK subvienetai turi ATP surišimo ir hidrolizės motyvus bei priklauso AAA+ ATPazių šeimai (Iyer, Leipe et al., 2004). ŽUK reikia prisijungti ATP, kad

galėtų jungtis su PV ir DNR (1.6 pav.) (Kelch, Makino et al., 2012). Hidrolizavus ATP ŽUK disocijuoja, o jo vietą užima DNR polimerazė.



1.6 pav. Procesyvo veiksnio užkėlimo mechanizmas (adaptuota iš (Kelch, Makino et al., 2012)).

ŽUK sudarytas iš keturių tarpusavyje panašių mažųjų ir vieno didžiojo subvieneto (eukariotuose – RFC1, archėjose – RFCL) (1.7 pav.). ŽUK subvienetams (išskyrus T4 gp62) būdinga trijų domenų organizacija. Subvienetai tarpusavyje jungiasi per apykaklės domeną. Domenai I ir II sudaro AAA+ modulį. Procesyvo veiksnys sąveikauja su ŽUK subvienetu, esančiu A, C ir E pozicijose (1.7 pav.), pirmu domenu.



1.7 pav. Žiedo užkėlimo kompleksų įvairovė, sandara ir subvienetų domeninė organizacija (adaptuota iš (Kelch, Makino et al., 2012)).

1.2.2. Helikazės

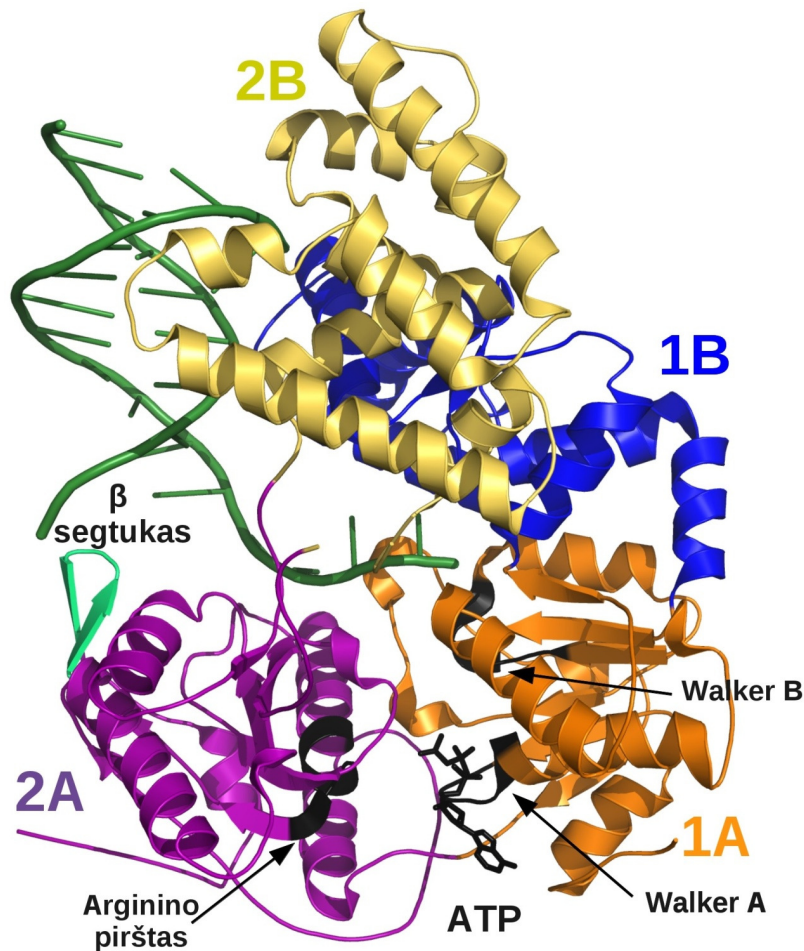
Visos helikazės turi RecA arba AAA+ NTP surišantį domeną, kuriam būdinga penkių lygiagrečių β juostų lakštas, apsuptas α spiralių (Singleton, Dillingham et al., 2007). Helikazės pagal tarpusavio panašumą skirstomos į šešias superšeimas (Lentelė 1.3, SF(1-6)). Pirmos ir antros superšeimos baltymai dažniausiai turi du sujungtus RecA domenų ir veikia kaip monomerai arba dimerai (Fairman-Williams, Guenther et al., 2010). SF3-6 šeimų atstovai dažnai formuoja heksamerinius žiedus. SF3 ir SF6 šeimos turi AAA+ domeną, o SF4 ir SF5 – RecA (Singleton, Dillingham et al., 2007). Replikacinių helikazių yra visose šeimose, išskyrus penktą. Šiai superšeimai priklauso bakterijų transkripcijos veiksnys *rho*.

Lentelė 1.3 Helikazių įvairovė ir savybės.

Superšeima	Judėjimo kryptis	Tipiški atstovai arba jų grupės	NTPazės domenai	4'inė struktūra
SF1	3'-5' (SF1A); 5'-3' (SF1B)	HerpesUL5/RecD/Pif1/Dda; Rep/UrvD; Upf1-like	2xRecA	(mono/di)merai
SF2	3'-5' (SF2A); 5'-3' (SF2B)	RecG; RecQ; Rad3/XPD; Ski2; T1R; Swi/Snf; RIG-I; DEAD- box; DEAH/RHA; NS3/NPH-II	2xRecA	(mono/di)merai
SF3	3'-5'	LTag; E1; Rep/NS1; Rep40	AAA+	heksamerai
SF4	5'-3'	DnaB; T7gp4; T4gp41; TWINKLE	RecA	heksamerai
SF5	5'-3'	<i>rho</i>	RecA	heksamerai
SF6	3'-5'	MCM, RuvB	AAA+	heksamerai

SF1 ir SF2 superšeimos. SF1 ir SF2 atstovai dalyvauja daugelyje DNR ir RNR metabolizmo žingsnių. Pagal sekų panašumą išskiriamos trys SF1 ir net dešimt SF2 šeimų (Lentelė 1.3) (Fairman-Williams, Guenther et al., 2010). Pirma ir antra helikazių superšeimos struktūriškai panašios. Jos neformuoja žiedo bei turi du RecA domenų (Singleton, Dillingham et al., 2007). NTP surišimo ir hidrolizės motyvai (1.8 pav., Walker (A/B) ir arginino pirštas) išsidėstę tarp RecA domenų (1.8 pav., 2A ir 1A) (Singleton, Dillingham et al.,

2007). Šios helikazės dgDNR grandinių atskyrimui naudoja specialią struktūrą – β segtuką (Buttner, Nehring et al., 2007). Papildomi domenai, įsiterpę į RecA domenus (pvz., 1.8 pav., 2B ir 1B) arba esantys jų galuose, lemia didelę šių superšeimų baltymų atliekamų funkcijų įvairovę (Fairman-Williams, Guenther et al., 2010).

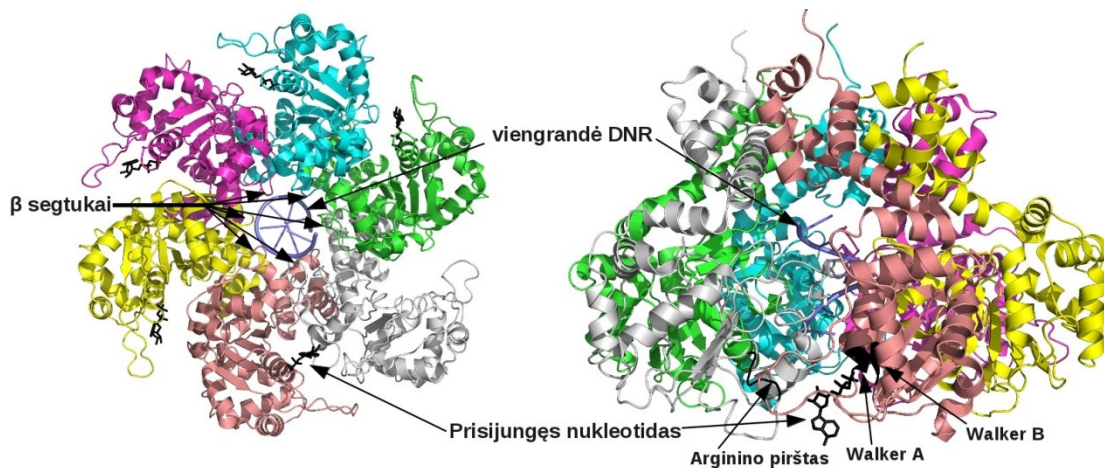


1.8 pav. SF(1/2) helikazių atstovo PcrA struktūra (PDB kodas: 3pjr). Domenai nuspalvinti skirtingomis spalvomis, ATP surišimo ir hidrolizės motyvai bei ATP pažymėti juoda spalva.

SF1 ir SF2 superšeimose žinomi tik keli replikacinių helikazių pavyzdžiai. Eksperimentiniai tyrimai rodo, kad N15 ir PY54 fagų SF2 šeimos helikazės (RepA) yra replikacinės (Mardanov, Strakhova et al., 2004; Ziegelin, Tegtmeyer et al., 2005). Geriausiai ištirta yra herpes virusų pirmos superšeimos replikacinė helikazė (UL5). Ji yra gimininga mielių Pif1 ir bakterijų RecD (Lentelė 1.3) (Hodgman, 1988). UL5 kartu su baltymais UL8 ir UL52 sudaro helikazės-praimazės kompleksą. UL52 yra archėjų-eukariotų

praimazių (AEP) superšeimos atstovas (Iyer, Koonin et al., 2005), o UL8 baltymas neturi katalizinio aktyvumo ir jo struktūra bei domeninė organizacija nežinomos (Muylaert, Zhao et al., 2012).

Heksamerinės helikazės. SF3-6 atstovai yra heksamerinės helikazės. SF4 ir SF6 superšeimų baltymai yra pagrindinės replikacinės helikazės visuose ląsteliniuose organizmuose (McGlynn, 2013). SF3-6 baltymams būdinga tai, kad jie formuoja heksamerinius žiedus, o NTP jungiasi tarp subvienetų esančioje kišenėje (1.9 pav.). Pagrindiniai NTP surišimo ir hidrolizės motyvai išsidėstę gretimuose monomeruose (arginino pirštas viename, o Walker A ir Walker B – kitame (1.9 pav)). Dvigrandės DNR išvyniojimas vyksta „traukiant“ vieną iš grandžių per heksamerinio žiedo centrą. Traukimą atlieka β segtukai, kontaktuojantys su vgDNR (1.9 pav.). Pastarųjų konformacija kinta priklausomai nuo NTP hidrolizės stadijos (Lyubimov, Strycharska et al., 2011).



1.9 pav. Papilomos viruso (BPV-1) helikazė (E1) (PDB kodas: 2gxa) (kairėje – vaizdas iš viršaus, dešinėje – iš šono).

Heksamerinės helikazės tarpusavyje skiriasi savo NTPazės domenais bei judėjimo ant vgDNR kryptimi (Lentelė 1.3). Bakterijų replikacinė helikazė (DnaB), priklausanti ketvirtai superšeimai, turi RecA domeną bei juda 5'-3' kryptimi, o eukariotų ir archėjų MCM baltymui (SF6 superšeima) bei papilomos, poliomos ir raupų virusų replikacinėms helikazėms (SF3 superšeima) būdingas AAA+ domenas ir priešinga judėjimo kryptis (Medagli

ir Onesti, 2013). Žiedą formuojančios helikazės užsėda ant vgDNR pačios (pvz., T7 fago gp4) arba naudoja specialius baltymus – helikazės užkėlėjus (pvz., *E. coli* DnaC arba Cdc6/Orc1 archėjose) (O'Shea ir Berger, 2014). Replikacijos metu, helikazėms išvyniojus dgDNR, vyksta pradmens sintezė. Šiam procesui reikia tikslaus helikazės ir praimazės veiksmų koordinavimo (Patel, Pandey et al., 2011). Minėti baltymai sąveikauja tarpusavyje bakterijose, o kai kuriuose faguose (T7, P4) ir NCLD virusuose – koduojami vienoje polipeptidinėje grandinėje (Iyer, Koonin et al., 2005).

1.2.3. Praimazės

DNR polimerazės sintetinti pradeda prijungdamos nukleotidą prie hidroksilo funkcinės grupės. Ląsteliniuose organizmuose ir daugelyje DNR virusų naudojamas praimazės susintetintos RNR 3'-OH galas. Praimazės, pagal jų turimą katalizinį domeną, skirstomos į dvi superšeimas: archėjų-eukariotų praimazės (AEP) ir topoizomerazės-praimazės domeno (Toprim) praimazės (Lentelė 1.4) (Aravind, Leipe et al., 1998; Iyer, Koonin et al., 2005). Šie domenai struktūriškai nepanašūs, bet jiems abiemis būdingas aktyvusis centras su konservatyviu DxD motyvu, naudojamu Mg²⁺ surišimui (Aravind, Leipe et al., 1998; Iyer, Koonin et al., 2005).

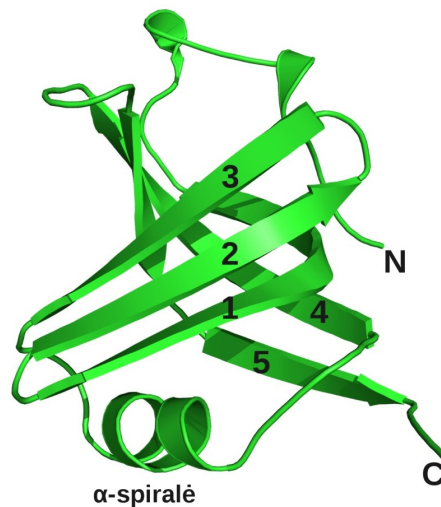
Praimazės gali turėti įvairių papildomų domenų ir dažnai būna aktyvios tik kompleksuose su kitais baltymais (Lentelė 1.4) (Kuchta ir Stengel, 2010). Pavyzdžiui, Toprim praimazės dažnai turi Zn²⁺ surišimo domeną, kuris jungiasi prie DNR. Taip pat praimazės gali turėti įvairių helikazių superšeimų domeną (Lentelė 1.4, T7, P4 fagai, raupų virusai). Archėjos ir eukariotai turi didįjį (PriL) ir mažąjį (PriS) praimazės subvienetus. Eukariotuose į šį kompleksą dar įeina DNR polimerazė (Pol α) ir p70 pagalbinis baltymas. Aktyvus bakterijų praimazės kompleksas susidaro tada, kai trys DnaG praimazės jungiasi su heksamerine helikaze (DnaB) (Kuchta ir Stengel, 2010). Herpes virusuose AEP praimazė (UL52) sudaro kompleksą su UL5 helikaze bei UL8 baltymu (Weller ir Coen, 2012).

Lentelė 1.4 Praimazių įvairovė ir savybės

Atstovai	Praimazės domenas	Aktyvus praimazės kompleksas	Sujungta su helikaze
T7 gp4	Toprim	-	+ (SF4)
T4 gp61	Toprim	gp61-gp41	-
P4 α	Toprim	-	+ (SF3)
<i>E. coli</i> DnaG	Toprim	DnaG-DnaB	-
Archėjų PriS	AEP	PriS-PriL	-
Eukariotų PriS	AEP	PriS(p49)-PriL(p58)-Pol α -p70	-
Raupų D5	AEP	-	+ (SF3)
Herpes UL52	AEP	UL52-UL5-UL8	-

1.2.4. Viengrandę DNR surišantys baltymai

DNR replikacijos metu susidaro ilgi viengrandės DNR (vgDNR) fragmentai. Jiems apsaugoti nuo pažeidimų naudojami vgDNR surišantys baltymai – SSB (single-stranded DNA-binding). Šiems baltymams dažniausiai būdingas vienas arba keli OB (oligonucleotide/oligosaccharide-binding – sąveikaujantis su oligonukleotidais) domenai (Lentelė 1.5), prie kurių jungiasi vgDNR. OB domenai sudaryti iš penkių β juostų statinaitės, uždengtos α spirale (1.10 pav.) (Murzin, 1993). Bakterijų bei T4 ir T7 fagų SSB būdingas rūgštinis C-galas. Jis būtinas SSB jungimuisi su kitais replikacijos baltymais, taip pat jis moduliuoja sąveiką su vgDNR (Marintcheva, Marintchev et al., 2008).



1.10 pav. Tipiškas OB domenas iš archėjų aspartil-tRNR sintetazės (PDB kodas: 1b8a).

Ląstelinių organizmų SSB baltymams (išskyrus *Thermoproteales* archėjų SSB (Paytubi, McMahon et al., 2012)) irgi būdingas OB domenas. Minėti baltymai tarpusavyje skiriasi OB-domenų kiekiu ir ketvirtine struktūra (Lentelė 1.5). Pavyzdžiui, *E. coli* SSB sudarytas iš keturių subvienetų, turinčių po vieną OB domeną, o eukariotams būdingas heterotrimerinis Rpa, turintis šešis OB domenų, iš kurių keturi sąveikauja su DNR (Lentelė 1.5). Archėjų Rpa būdingas įvairus OB domenų kiekis ir ketvirtinė struktūra (Lentelė 1.5) (Sarmiento, Long et al., 2014). Virusų SSB pasižymi didele įvairove. Jiems būdingi tiek OB domenų turintys SSB (T4, T7, herpes virusų atvejais), tiek SSB baltymai, kuriuose OB domeno nėra arba nerasta (phi29, adeno, raupų, bakulo virusai) (Lentelė 1.5).

Lentelė 1.5 VgDNR surišančių baltymų įvairovė ir savybės.

Atstovai	OB domenų kiekis monomere/komplekse	Ketvirtinė struktūra
T7 gp2.5	2	homodimeras
T4 gp32	1	monomeras
phi29 Gp5	?	monomeras
Adeno DBP (E2A)	-	?
Raupų I3	?	?
Bakulo virusų LEF3	?	?
Herpes ICP8	2	dimeras
Eukariotų Rpa	6 (keturi iš jų – sąveikai su vgDNR)	heterotrimeras (Rpa 70, 32 ir 14)
<i>Sulfolobus solfataricus</i> Rpa	1	monomeras
<i>Pyrococcus</i> Rpa	3	heterotrimeras (Rpa 31, 32, 14)
<i>Thermoproteales</i> ThermoDBD	-	dimeras
<i>E. coli</i> SSB	4	homotetrameras

1.2.5. Pradmenį pašalinantys baltymai

Ląstelinių organizmų genomas sudarytas iš DNR, todėl praimazės susintetintas RNR pradmuo turi būti pašalintas. Šią funkciją atlieka pradmenį pašalinantys (PP) baltymai. Jie yra nukleazės, pradmenį šalinančios 5'-3' kryptimi, bei katalizei naudojančios du metalo jonus, kuriuos suriša konservatyvios asparto arba glutamo rūgšties liekanos (Yang, 2011). Nepaisant to, kad PP baltymai atlieka tą patį darbą, jų struktūros skirtingos. Pagal

tarpusavio panašumą išskiriamos trys superšeimos. Joms priklauso įvairias DNR/RNR metabolizmo funkcijas (DNR replikacija, rekombinacija, reparacija, restrikcija-modifikacija, mRNR brendimas, RNR interferencija) atliekantys baltymai (Yang, 2011). Dažniausiai pradmenį pašalinantys baltymai yra FEN superšeimos atstovai. Šiai šeimai priklauso eukariotų ir archėjų bei T4 ir T7 fagų pradmens pašalinimo baltymai (Lentelė 1.6). Bakterijų DNR replikacijos metu pradmenį pašalina DNR polimerazė (PolI), savyje turinti FEN domeną. RnaseH superšeimoje pradmens pašalinimo funkciją atlieka RNaseH(I/1) ir RNaseH(II/2) grupių atstovai (Lentelė 1.6) (Yang, 2011; Makarova ir Koonin, 2013). Eukariotų RNaseH žymimos arabiškais skaitmenimis, o prokariotų – romėniškais (Cerritelli ir Crouch, 2009). PD-(D/E)XK superšeima pasižymi funkcijų įvairove. Joje išskiriama net 121 grupė baltymų (Steczkiewicz, Muszewska et al., 2012). PD-(D/E)XK superšeimos 5'-3' egzonukleazės vykdo lambda fago ir bakterijų DNR rekombinaciją (Lentelė 1.6) (Yang, 2011). Žinoma, kad eukariotų Dna2 DNR replikacijos metu pašalina ilgus RNR pradmenis (Kim, Kim et al., 2006). Herpes virusų šarminė egzonukleazė (UL12), kuriai nustatytas 5'-3' RNazės aktyvumas (Knopf ir Weissart, 1990), manoma, kad gali atlikti pradmens pašalinimo funkciją (Boehmer ir Lehman, 1997).

Lentelė 1.6 Pradmenį pašalinančių baltymų įvairovė ir savybės.

Superšeima	Atstovai
FEN	Archėjų-eukariotų FEN1, T4FEN, T7 gp6, <i>E. coli</i> DNA PolI
RNaseH	Žmogaus RNaseH1, bakterijų RNaseHI, HIV atvirkštinė transkriptazė, <i>A. fulgidus</i> RNaseHII, bakterijų RNaseHII, HIV integrazė, RuvC, UvrC, <i>P. furiosus</i> argonautas
PD-(D/E)XK	λ fago egzonukleazė (Red rekombinazė), <i>E. coli</i> RecE, RecB, <i>B. subtilis</i> AdnAB, Dna2, Herpes UL12

1.2.6. Ligazės

Pašalinus pradmenį lieka tarpas tarp gretimų 5'-fosforil ir 3'-hidroksil grupių. DNR ligazės fosfodiesteriniu ryšiu sujungia minėtas grupes. Pagal naudojamą kofaktorių išskiriami du DNR ligazių tipai: NAD ir ATP. Šios dvi ligazių grupės yra tarpusavyje tolimai giminingos (Aravind ir Koonin, 1999).

ATP ligazės dažniausiai sutinkamos archėjose ir eukariotuose (Lig1), o bakterijų DNR replikacijai būtinos NAD ligazės (Leipe, Aravind et al., 1999). DgDNR virusams būdingos abiejų grupių DNR ligazės (Benarroch ir Shuman, 2006; Yutin ir Koonin, 2009). NAD ir ATP ligazės sąveikai su DNR dažnai turi papildomus domenų ir(arba) motyvus, pavyzdžiui, OB domeną, cinko pirštų, spiralė-segtukas-spiralė (Helix-hairpin-Helix (HhH)) motyvus (Tomkinson, Vijayakumar et al., 2006).

1.2.7. Topoizomerazės

Replikacijos metu helikazė išskiria dgDNR grandines, o prieš replikacijos šakutę esanti dgDNR susivynioja (didėja teigiama superspiralizacija). Dėl dgDNR susivyniojimo atsiradusi įtampa gali lėtinti arba sustabdyti DNR replikaciją. Įtampą pašalina DNR topoizomerazės – fermentai, keičiantys DNR topologiją. Jie gali kirpti vieną arba abi DNR grandines, pridėti arba sumažinti DNR sūkių skaičių, tam darbui naudodami NTP energiją arba ne. Taip pat DNR topoizomerazės naudojamos išskirti katenuotas DNR molekules, susidariusias replikacijos arba rekombinacijos metu.

IA/IIA/IIB, IB ir IC topoizomerazės turi skirtingus DNR skėlimo domenų (Lentelė 1.7), tačiau visoms joms būdingas konservatyvus tirozinas aktyviajame centre (Chen, Chan et al., 2013). Jis atlieka nukleofilo funkciją DNR skėlimo ir sujungimo reakcijose. Topoizomerazės pagal DNR skėlimo pobūdį skirstomos į dvi grupes: skeliančios vieną grandinę priklauso I grupei, o skeliančios abi grandines – II (Lentelė 1.7). Visų grupių (išskyrus IA) atstovai dalyvauja mažinant teigiamą superspiralizaciją, susidarančią DNR replikacijos metu (Vos, Tretter et al., 2011).

Lentelė 1.7 Topoizomerazių įvairovė ir savybės.

Grupė	Pogrūpis	Ketvirtinė struktūra	DNR skėlimo domenas	Funkcija	Eukariotai	Archėjos	Bakterijos
I	A	mono	toprim/CAP	Mažina neigiamą superspiralizaciją	Topo III(α/β)	TopoIII; atvirkštinė DNR girazė	Topo (I/III/III β); atvirkštinė DNR girazė
I	B	mono	IB/Int	Mažina/didina neigiamą/teigiamą superspiralizaciją	Topo I		Topo IB
I	C	mono	Nepanašus į kitas topoizomeras ¹	Mažina/didina neigiamą/teigiamą superspiralizaciją, DNR reparacija		Topo V	
II	A	di/tetra	toprim/CAP	Mažina/didina neigiamą/teigiamą superspiralizaciją	Topo II(α/β)	girazė	Topo IV; girazė
II	B	tetra	toprim/CAP	Mažina/didina neigiamą/teigiamą superspiralizaciją	Top6(A/B)*	Topo VI	

*tik augaluose

¹(Taneja, Patel et al., 2006)

2. METODAI

2.1. Pradiniai duomenys ir duomenų bazės

Nepasikartojančių sekų (non-redundant („nr“)) baltymų duomenų bazė parsųsta iš NCBI („ftp://ftp.ncbi.nlm.nih.gov/blast/db/“). Dvigrandės DNR virusų genomų sekos parsųstos iš:

„<http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid=10239>“.

Virusų baltymai paimti iš genomų GenBank failų. Siekiant išvengti genų praradimo dėl nepilnai anotuotų genomų, visi genomai buvo nutransliuojami šešias rėmeliais, naudojant programas Virtual Ribosome (Wernersson, 2006) arba Bioperl (Stajich, Block et al., 2002). Visi anksčiau neanotuoti atviro skaitymo rėmeliai, ilgesni nei 60 a-r., buvo įtraukti į tolimesnę analizę. Virusai iš *Polydnaviridae* šeimos nebuvo tirti, nes jiems būdinga netipiška genomo organizacija bei žinoma, kad jie dalyvauja tik kaip vektoriai parazitinių vapsvų genomo pernašai (Federici ir Bigot, 2010).

2.2. Replikacijos baltymų paieška

Replikacijos baltymų paieškai naudoti standartiniai profilio-sekos (pvz., PSI-BLAST (Altschul, Madden et al., 1997)) ir jautrūs profilio-profilio (pvz., HHsearch (Söding, 2005)) lyginimo metodai. Visų pirma, remiantis literatūros duomenimis, kiekvienai replikacijos baltymų grupei (DNR polimerazės ir jų pagalbiniai baltymai (žiedai ir jų užkėlėjai), viengrandę DNR surišantys baltymai, DNR praimazės, helikazės, pradmens pašalinime dalyvaujančios nukleazės, DNR ligazės bei topoizomerazės) atrinkti reprezentatyvūs atstovai. Siekiant papildyti reprezentatyvių atstovų sąrašą, kiekvienai iš jų atliktos sekų paieškos su Jackhmmer (Eddy, 2011) arba PSI-BLAST programomis, o atstovams su žinomomis struktūromis alikta panašių struktūrų paieška Dali serveryje (Holm ir Rosenstrom, 2010). Galiausiai, reprezentatyvūs atstovai buvo naudojami sekų paieškai per Pfam (Punta, Coggill et al., 2012), SCOP (Murzin, Brenner et al., 1995) ir PDB (<http://www.pdb.org/>) duomenų bazes.

Remiantis minėtų paieškų rezultatais atrinkti Pfam, SCOP ir PDB replikacinių baltymų profiliai.

Naudojant reprezentatyvius replikacinių baltymų atstovus, atliktos jų homologų paieškos dgDNR virusų baltymų ir atviro skaitymo rėmelių duomenų bazėje. Šiam tikslui pasitelktos PSI-BLAST ir Jackhmmer programos. Sekos profilis sugeneruotas iteruojant iki jo konvergencijos (tol, kol nerandama naujų sekų) ir naudojant griežtą naujų sekų įtraukimo ribą (E-value=0.001). Analizuojant paieškų rezultatus pildytas reprezentatyvių replikacijos baltymų sąrašas bei sudarytas preliminarus dgDNR virusų replikacijos baltymų sąrašas, kuris vėliau pildytas jautrių homologų paieškų rezultatais.

Jautri homologų paieška. Jautriai homologų paieškai buvo naudoti profilių lyginimo metodai: HHsearch (Söding, 2005) ir Condor (<http://mindaugas.ibt.lt/condor/>). Šie metodai veikia dviem etapais. Iš pradžių sekų duomenų bazėje ieškoma homologų duotajai sekai. Duotoji seka ir sekos, reikšmingai panašios į ją, išlyginamos ir sukuriamas profilis. Sekų profiliai sudaryti visiems virusų baltymams vykdant dvi arba tris iteracijas per nufiltruotas duomenų bazes (Lentelė 2.1). Antrame etape profiliai lyginami su profiliais, sudarytais struktūrų (PDB ir SCOP) bei baltymų šeimų (Pfam) duomenų bazėse esančioms sekoms. Tolimesnei analizei naudoti tie rezultatai, kurių panašumo į replikacinių baltymų profilius įverčiai buvo geresni už reikšmingumo įverčio ribą (Lentelė 2.1).

Lentelė 2.1 Jautrios homologų paieškos metodai ir naudoti parametrai.

Metodas	Profilio sudarymo programa	Duomenų bazė	Iteracijų skaičius	Reikšmingumo įvertio riba (E-value)	Profilijų duomenų bazės	Profilijų lyginimo programa	Reikšmingumo įvertio riba
HHsearch	HHblits	nr20	2	0,001	Pfam26/SCOP1.75/PDB70_120616	HHsearch	Tikimybė=20 %
Condor	Jackhammer	nr70	2	0,001	Pfam26/SCOP1.75/PDB70_130311	Condor	E-value=10
HHsearch	PSI-BLAST	nr90	3	0,001	PDB70_2011/replika zės komponentų profiliai	HHsearch	Tikimybė=50 %

2.3. Baltymų sekų grupavimas

Paieškų metu rastos virusų potencialių replikacijos baltymų sekos sugrupuotos kartu su atitinkamomis replikacinių baltymų grupėmis naudojant CLANS (Frickey ir Lupas, 2004) programą. CLANS sekų tarpusavio panašumas įvertinamas visoms grupuojamoms sekoms atliekant BLAST arba PSI-BLAST paieškas. BLAST algoritmas naudotas tuomet, kada sekos tarpusavyje buvo artimai giminingos. Stipriai divergavusioms sekoms grupuoti buvo naudojamas CLANS su PSI-BLAST. Iš pradžių, PSI-BLAST programa sugeneruodavo profilį atlikdama, nuo dviejų iki šešių iteracijų su E-value=0.001 įtraukimo riba, paiešką per nr70 sekų duomenų bazę. Tuomet profilis buvo naudojamas paieškai per grupuojamų sekų duomenų bazę. Gautus sekų tarpusavio panašumo įverčius CLANS programa perskaičiuodavo į P-įverčius (P-value) atsižvelgdama į sekos ilgį ir duomenų bazės dydį. Pagal minėtus įverčius sekos buvo išdėstomos erdvėje. Analizuojant susidariusius klasterius identifikuoti replikacinių baltymų homologai.

2.4. Replikacinių baltymų identifikavimas

Identifikuojant replikacijos baltymus arba esant keliems kandidatams į tam tikrą replikacinių baltymų grupę, atrenkant vienintelį kandidatą, vadovautasi šiais kriterijais (mažėjančio prioriteto tvarka):

1. Panašumas į reprezentatyvius replikacijos baltymus (nustatyta iš HHsearch/Condor ir CLANS grupavimosi rezultatų). Esant dvejonėms dėl panašumo, buvo papildomai leidžiama paieška HHpred (Söding, Biegert et al., 2005) (PSI-BLAST, dvi iteracijos, įtraukimo riba – 0.001 ir(arba) Jackhmmer paieška) ir(arba) GeneSilico (Kurowski ir Bujnicki, 2003) serveriuose.
2. Aktyviojo centro ir konservatyvių regionų (ne)buvimas (kiekvienai replikacijos baltymų grupei sudaryti daugybiniai palyginiai).
3. Identifikuojamame baltyme yra kito replikacijos baltymo domenas (pavyzdžiui, helikazės ir praimazės atvejais).
4. Genome nėra kito baltymo atliekančio tą pačią funkciją.
5. Baltymas genome yra koduojamas replikacijos baltymų aplinkoje.

2.5. Papildomos homologų paieškos

Tyrimo metu pastebėta, kad kuo didesnis genomas, tuo pilnesnis jo replikacijos aparatas. Todėl sudarius baltymų rinkinius buvo bandyta ieškoti „trūkstančių“ replikacinių baltymų didelių genomų (virš 100 kbp) virusuose. Naudojant CLANS programą sugrupuotos virusų proteomos ir jau charakterizuoti ląstelių organizmų bei jau identifikuoti virusiniai replikacijos baltymai. Manoma, kad replikacinių genų nerandama dėl nejautraus sekų profilio, naudoto paieškai. Profilio jautrumui padidinti reikia, kad jį sudarytų daugiau sekų, todėl nuspręsta virusų sekų profilius generuoti ieškant homologų per duomenų bazę, papildytą metagenominiais duomenimis (nr80_env_nr80), naudojant keturias PSI-BLAST iteracijas. Reikšmingo panašumo į jau žinomus replikacijos baltymus gali būti nerandama todėl, kad buvo pasirinkta netinkama pradinė seka profiliui generuoti. Todėl nutarta profilio sudarymui naudoti jau

anksčiau rastų virusinių replikacijos baltymų homologus. Sugrupavus virusų proteomų ir replikacinius baltymus CLANS programa buvo identifikuoti reikšmingi tarpusavio panašumai ir papildytas virusų replikacijos baltymų sąrašas.

2.6. Sekų palyginiai

Daugybiniai sekų palyginiai sudaryti naudojant MAFFT (Kato ir Standley, 2013) (pasirinktas L-INS-i parametras) arba MUSCLE (Edgar, 2004) su standartiniais nustatymais. Jeigu lyginamos sekos turėjo homologų, kurių struktūros yra žinomos, tuomet daugybinio sekų palyginio sudarymui naudotas PROMALS3D (Pei, Kim et al., 2008) serveris. DaliLite serveris (Holm ir Park, 2000) buvo naudojamas sudaryti sekų palyginiams užklojant atitinkamų baltymų struktūras.

2.7. Antrinės struktūros ir nestruktūrizuotų regionų nustatymas

Baltymo antrinė struktūra, esant žinomai jo kristalinei struktūrai, suskaičiuota su DSSP (Joosten, te Beek et al., 2011). Genesilico metaserveriu (Kurowski ir Bujnicki, 2003) buvo atliktas antrinės struktūros ir nestruktūrizuotų regionų nusakymas. Papildomai nestruktūrizuoti regionai nustatyti PrDOS (Ishida ir Kinoshita, 2007) ir MetaDisorder3D (Kozlowski ir Bujnicki, 2012) serveriais.

2.8. Genomų lyginimas ir genominės aplinkos analizė

Genomai išlyginti ir genų aplinka nustatyta naudojant ACT (Artemis Comparison Tool) programą (Carver, Rutherford et al., 2005).

2.9. Genomų grupavimas ir filtravimas

Virusų genomai sugrupuoti pagal jų genomų ir baltymų sekų tarpusavio panašumą naudojant atitinkamai LAST (Frith, et al., 2010) ir CLANS

programas. Grupių pavadinimai sudaryti atsižvelgiant į jų taksonominę klasifikaciją. Grupė virusų, nepriskirtų jokiai žemesnei taksonominei kategorijai, buvo pavadinama pagal seniausią sekvenuotą jos atstovą. Pradiniam genomų rinkinyje buvo daug tarpusavyje panašių genomų. Norint gauti reprezentatyvesnę genomų imtį jie buvo tarpusavyje palyginti LAST ir CLANS programomis ir atmesti tie, kurių lokalūs nukleotidų sekų panašumai buvo >70% arba kurie turėjo >70% homologiškų baltymų.

2.10. Homologinis modeliavimas ir struktūros analizė

Pradinis modeliuojamos sekos (taikinio) ir jai giminingos struktūros (šablono) palyginys sudarytas remiantis PSI-BLAST-ISS (Margelevičius ir Venclovas, 2005), COMA (Margelevičius, Laganeckas et al., 2010), GeneSilico, I-TASSER (Roy, Kucukural et al., 2010), RaptorX (Kallberg, Wang et al., 2012), pGenTHREADER (Lobley, Sadowski et al., 2009), FFAS-3D (Xu, Jaroszewski et al., 2013), HHpred (Söding, Biegert et al., 2005) programų ir serverių pateiktais palyginiais. Abejotini taikinio-šablono palyginio regionai pataisyti rankiniu būdu iteratyvaus modeliavimo procedūros (Venclovas ir Margelevičius, 2009) eigoje. 3D modelio sudarymui naudota Modeller programa (Šali ir Blundell, 1993). Gauto modelio kokybė įvertinta ProSA-web serveryje (Wiederstein ir Sippl, 2007) bei palyginta su kokybe šablono, naudoto modeliavimui, kurio trūkstami struktūros regionai buvo įmodeliuoti prieš vertinant kokybę. Modelių a-r. konservatyvumas analizuotas ConSurf serveriu (Ashkenazy, Erez et al., 2010) ant struktūros paviršiaus užklojant sekos pozicijų konservatyvumo įverčius, gautus iš daugybinių sekų palyginių.

2.11. Elektrostatinių savybių analizė

Sekos teorinio izoelektrinio taško (pI) skaičiavimai atlikti programa Isoelectric point iš EMBOSS programų paketo (Rice, Longden et al., 2000). Prieš skaičiuojant pI nekonservatyvūs sekų N ir C galai buvo pašalinti.

Paviršiaus elektrostatinio potencialo žemėlapiai suskaičiuoti APBS programa (Baker, Sept et al., 2001), kuri yra PyMol programos (Schrodinger, 2010) įskiepis. Prieš pradėdant skaičiavimus iš PDB struktūrų pašalinti visi hetero atomų ir vandens molekulių įrašai. Modelių ir PDB struktūros paruoštos naudojant PDB2PQR (Dolinsky, Nielsen et al., 2004) su Amber jėgų lauku.

3. REZULTATAI IR JŲ APTARIMAS

DgDNR virusų replikacijos baltymų tyrimas vyko dviem etapais. Iš pradžių buvo tirtos DNR replikazės. Nustatytas ryšys tarp genomo dydžio ir replikazės komponentų (ne)buvimo genome. Ši išvalga įgalino atidžiau išanalizuoti vienus iš didžiausių phiKZ fagų ir juose naujai rasti DNR polimerazes. Kadangi DNR replikacijoje dalyvauja ne tik replikazės komponentai, bet ir DNR praimazės, helikazės, pradmens pašalinime dalyvaujančios nukleazės, DNR ligazės bei topoizomerazės, todėl tolimesniame etape nuspręsta tyrimą papildyti minėtais virusų DNR replikacijos baltymais. Pastarojo tyrimo eigoje naujai rasti SSB baltymai NCLD virusuose, nustatytas raupų viruso SSB (I3) panašumas į *E. coli* SmpB, paaiškėjo, kad herpes virusų helikazės-praimazės komponentas UL8 yra inaktyvuota B-šeimoms DNR polimerazė, o helikazės yra dažniausias virusų replikacijos baltymas.

3.1. DNR replikazių analizė

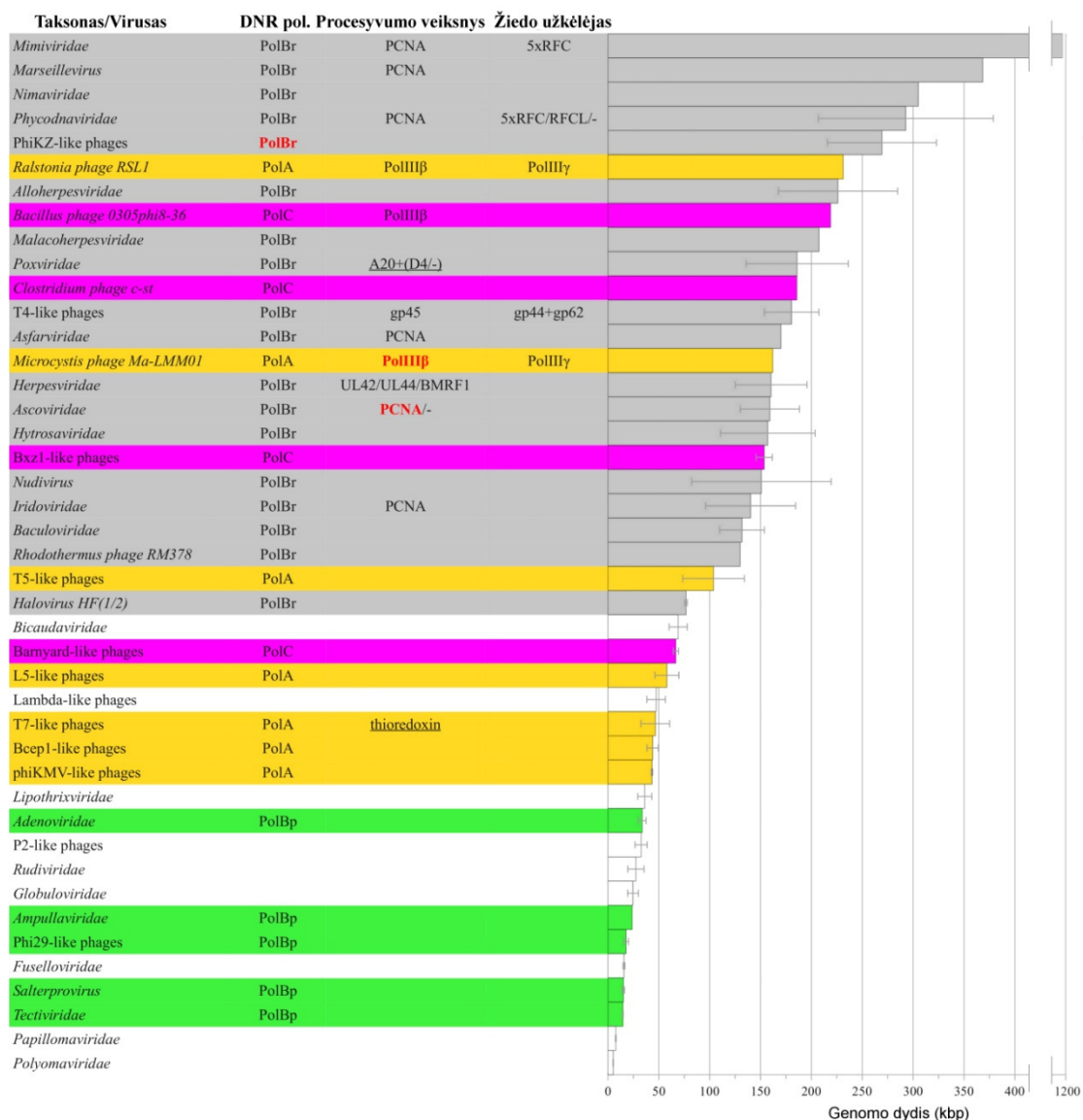
Analizuojant gausią informaciją, esančią dvigrandės DNR virusų genomų ir baltymų duomenų bazėse, buvo tirta šių organizmų replikazių įvairovė, jų savybės, taksonominis paplitimas bei sąsajos su genomo dydžiu.

3.1.1. DNR replikazės komponentai ir genomo dydis

Pradiniame šio darbo etape buvo tirtos dgDNR virusų replikazės. DNR replikazės komponentų ieškota 808 sekvenuotuose dgDNR virusų genomuose. 57% visų jų priklausė bakteriofagams, 39% – eukariotų ir tik 4% – archėjų virusams. Genomuose buvo ieškoma DNR polimerazių, procesyvumo veiksmų ir jų užkėlėjų homologų. Beveik pusė tirtų virusų turėjo DNR polimerazę. PhiKZ faguose ji buvo rasta pirmą kartą (žr. priedą). Procesyvumo veiksmų homologus turėjo žymiai mažiau virusų (<20%). Šie baltymai naujai rasti Ma-LMM01 fage ir *Ascoviridae* šeimoje. Žiedus formuojantys procesyvumo veiksniai (PCNA, polIIIβ, gp45) naudoja pentamerinius užkėlėjus. Kaip ir

galima buvo tikėtis, pastarieji buvo rasti tik virusuose, turinčiuose procesyvumo veiksnį. Įdomu tai, kad ne visi virusai turintys PCNA ar polIIIβ homologus turėjo baltymus, reikalingus jų užkėlimui ant DNR.

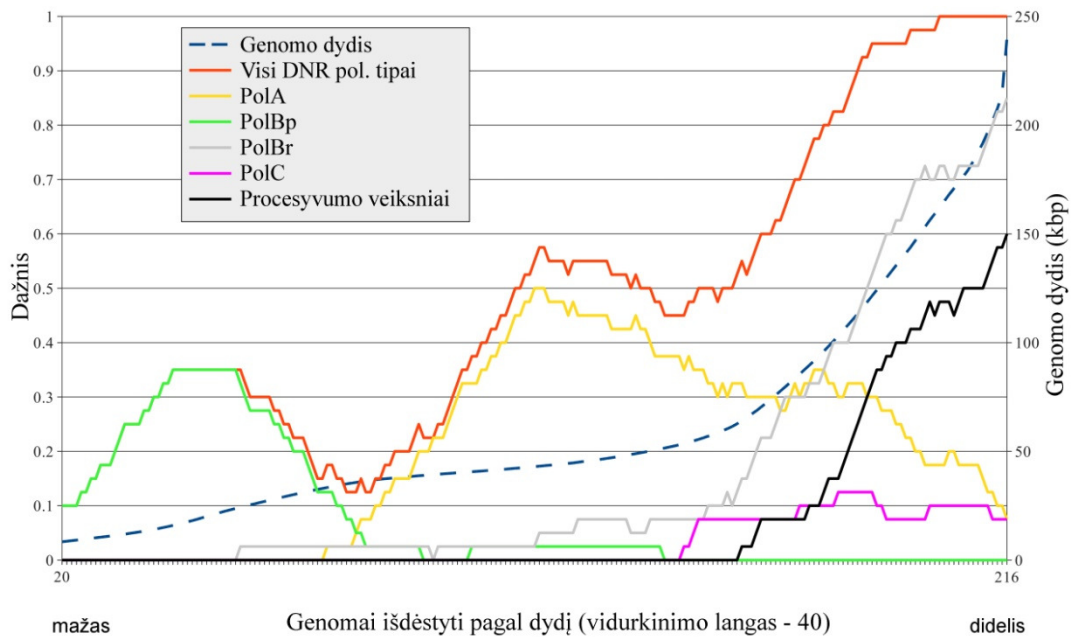
Tyrimo rezultatai atskleidė didelę dgDNR virusų replikazės komponentų ir jų kombinacijų įvairovę. DgDNR virusų replikazių įvairovė yra didesnė nei visuose ląstelinuose organizmuose kartu sudėjus ir iš pirmo žvilgsnio atrodo atsitiktinė. Nepaisant didelės virusų įvairovės, pavyko rasti aiškius dėsningumus (3.1 pav.). Išdėsčius virusų taksonomines grupes pagal jų vidutinį genomo dydį nustatyta, kad mažiausių genomų (<40 kbp) virusai neturi DNR polimerazių arba turi DNR polimerazes, pradmeniu naudojančias baltymą (PolBp). Didesni virusai (40–140 kbp) dažniau turi nuosavas DNR polimerazes. Jos dažniausiai priklauso A, rečiau – B ir C šeimoms. Didžiausių genomų (>140 kbp) virusai visada turi DNR polimerazes (dažniausiai B šeimos), dažnai koduoja procesyvumo veiksnius ir kartais netgi žiedo užkėlėjus.



3.1 pav. DgDNR virusų replikazių komponentai. Virusų taksonominės grupės išdėstytos pagal jų genomo vidurkį. DNR pol., DNR polimerazės tipas; PolA, A šeima; PolBr, B šeimos DNR polimerazės, pradmeniu naudojančios RNR; PolBp, B šeimos DNR polimerazės, pradmeniu naudojančios baltymą; PolC, C šeima. Spalvinimo schema: balta, nerasta DNR polimerazių; žalia, PolBp; geltona, PolA; pilka PolBr; rožinė, PolC. Naujai rasti replikazės komponentai vaizduojami raudonu šriftu. Netipiški procesyvo veiksniai – pabraukti. Minus ženklas žymi tai, kad ne visi grupės nariai turi tam tikrą komponentą. Paklaidų rėžiai žymi standartinį nuokrypį nuo genomo dydžio vidurkio.

Virusų taksonominių grupių reprezentatyvumas gali stipriai skirtis. Be to, kai kurie taksonai labai varijuoja savo genomo dydžiu. Todėl tolimesniame tyrimo etape buvo nuspręsta patikrinti ar nustatytos tendencijos priklauso nuo taksonominės klasifikacijos ar yra bendresnės. Virusų genomai buvo išdėstyti pagal dydį. Replikazės komponentų dažnius ir genomo dydį suvidurkinus 40

genomų judančiu langu, nubraižytas grafikas (3.2 pav.). Siekiant sumažinti imties paklaidą, analizei naudoti genomai tarpusavyje palyginti ir atrinkti tik mažiau nei 70% tarpusavyje panašūs. Rezultatai parodė jau anksčiau minėtą geno dydžio ir replikazės komponentų ryšį. Taigi, paaiškėjo, kad stebimi dėsniniai yra bendri, o ne atsiradę dėl taksonominės virusų klasifikacijos.

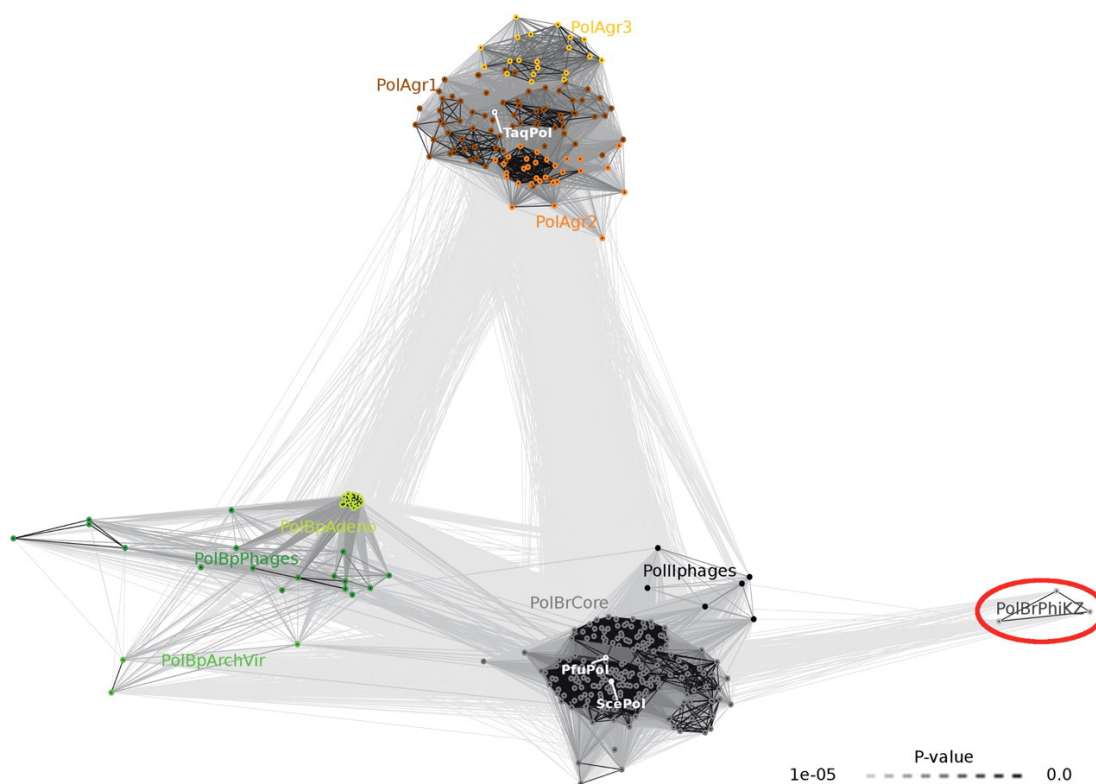


3.2 pav. DgDNR virusų DNR replikazės komponentų dažnumas ir geno dydis. X ašis – genomai išdėstyti pagal dydį; pagrindinė Y ašis (kairėje) – nustatyti DNR replikazės komponentų dažniai genomuose. Genomo dydis ir replikazės komponentų dažniai suvidurkinti naudojant 40 genomų langą, judantį vieno geno žingsniu. Santrumpos kaip ir 3.1 pav.

Nustatytas geno dydžio ir DNR replikazių buvimo jame ir jų tipo dėsninumas (3.1 ir 3.2 pav.) turėjo išimčių. Įdomu tai, kad didelių genomų virusai naudoja tik savo DNR polimerazes, tuo tarpu, ne visi jų turi procesyvumo veiksnius ir jų užkėlėjus. Kaip užtikrinamas replikazės procesyvumas minėtais atvejais? Norint atsakyti į šį klausimą bei apžvelgti dgDNR virusų replikazių baltymus buvo atlikta virusų DNR polimerazių, procesyvumo veiksmių ir jų užkėlimo kompleksų sekų ir struktūrų analizė.

3.1.2. DNR polimerazės

Pagrindinės DNR polimerazių grupės. Išanalizavus 808 virusų genomus, DNR polimerazės buvo rastos 415 atvejais. Nustatyta, kad dauguma (255) DNR polimerazių priklauso B šeimai, rečiau – A (132), ir labai retai (28) – C šeimai. Archėjoms būdingų D šeimos polimerazių virusuose nerasta. B šeimos atstovai yra būdingi visų trijų gyvybės domenų virusams, o A ir C atstovus turi tik bakteriofagai. Didžiausia baltymų įvairovė yra B šeimoje, o jai tolimai giminingoje A šeimoje baltymai yra tarpusavyje panašesni (3.3 pav.). Pastarosioms negiminingoje C šeimoje baltymai įvairove nepasižymėjo.



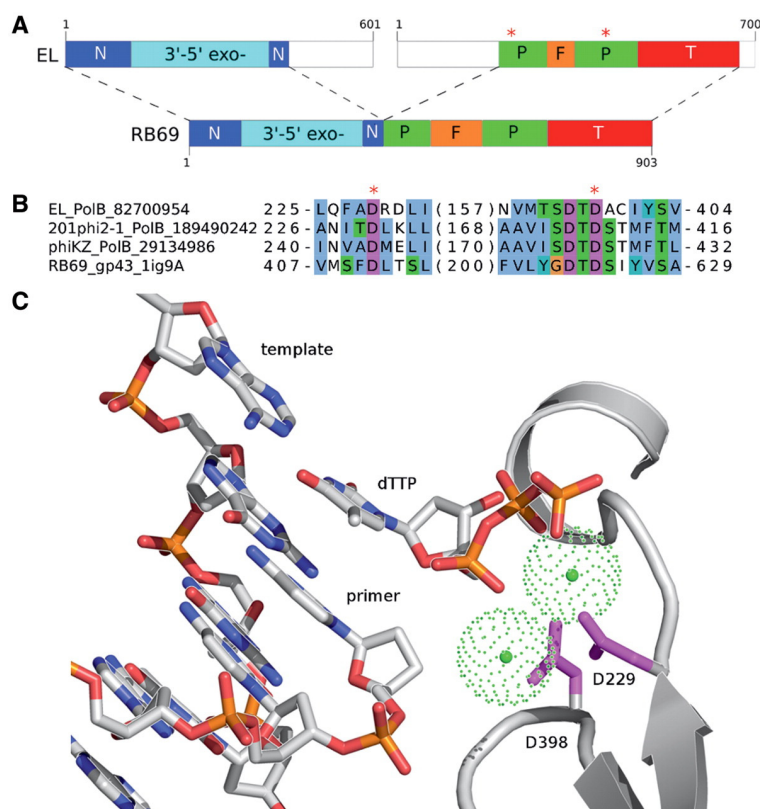
3.3 pav. A ir B šeimų DNR polimerazės sugrupuotos pagal sekų tarpusavio panašumą. Taškai žymi sekas. Linijomis sujungtos sekos, kurių reikšmingumo įvertis (P-value) $\leq 1e-05$. Linijų spalvų skalė pateikta apatiniame dešiniajame kampe (šviesios ir ilgos linijos jungia tolimai giminingas sekas). A šeimos polimerazės nuspalvintos oranžiniais atspalviais; PolBp, žaliais atspalviais; PolBr, pilkais atspalviais; ląstelių organizmų DNR polimerazės pateiktos baltu šriftu. Naujai rastos DNR polimerazės apvestos raudona elipse. ArchVir, archėjų virusai; Adeno, *Adenoviridae*; gr., grupė; PhiKZ, phiKZ-like fagai; Pfu, *Pyrococcus furiosus*; Sce, *Saccharomyces cerevisiae*; Taq, *Thermus aquaticus*.

Pagal sekų panašumą galima išskirti tris B šeimos polimerazių grupes. Pirmąją grupę sudaro DNR polimerazės, pradmeniu naudojančios baltymą (3.3

pav., PolBp). Kitoms dviems grupėms priklauso polimerazės, pradmeniu naudojančios RNR (PolBr). PolBp klasterį sudaro tarpusavyje panašūs adeno virusų (PolBpAdeno) ir joms giminingi, bet įvairesni fagų (PolBpPhages) ir archėjų virusų (PolBpArchVir) baltymai. Didžiausiam PolBr klasteriui priklauso dauguma virusų B šeimos DNR polimerazių (PolBrCore) ir nedidelė, bet specifinė fagų PolBr grupelė (PolIIPHages). PolBrCore grupei priklauso detaliai iširtos T4 fago ir herpes virusų polimerazės. Jos rodo didelį panašumą į eukariotų ir archėjų B šeimos atstovus (3.3 pav., ScePol, PfuPol). PolIIPHages grupė iš kitų B šeimos polimerazių išsiskiria jos aktyviajame centre esančiu „NTDG“ motyvu bei panašumu į *E. coli* PolIII. Mažas antrasis PolBr klasteris (PolBrPhiKZ) sudarytas iš naujai šiame tyrime rastų B šeimos DNR polimerazių.

PhiKZ-like fagų genomai yra beveik du kartus didesni nei T4 fago (palyginimui, 201phi2-1 – 317 kbp, T4 – 169 kbp), tačiau iki šiol DNR polimerazių juose nebuvo rasta (Mesyanzhinov, Robben et al., 2002; Hertveldt, Lavigne et al., 2005; Thomas, Rolando et al., 2008). Replikazės komponentų ir geno dydžio priklausomybės tyrimas parodė, kad didžiausi virusai turi nuosavas DNR polimerazes. Vienintelė išimtis buvo PhiKZ-like fagai. Todėl būtent jų genomai buvo kruopščiai išanalizuoti. Taikant standartinius homologų paieškos metodus (BLAST, RPS-BLAST ir PSI-BLAST) nepavyko rasti reikšmingo panašumo tarp phiKZ-like fagų baltymų ir charakterizuotų DNR polimerazių. Panašumas nustatytas tik pritaikius jautrius profilių lyginimo metodus. HHsearch programa (Söding, 2005) nustatė *Pseudomonas phage EL* baltymo (gi: 82700954) panašumą į RB69 (T4-like) fago DNR polimerazę (gp43) su aukšta 89% tikimybe. COMA serveris rado to pačio baltymo panašumą į *Thermococcus sp.* B šeimos polimerazę su geriausiu įverčiu (E-value=4e-07). Numanoma fago EL DNR polimerazė bei jos homologai iš PhiKZ-like fagų turi visus B šeimos DNR polimerazėms būdingus domenus, išskyrus 3'-5' egzonukleazę. Įdomu, kad baltymas, turintis 3'-5' egzonukleazės aktyvumą, jau buvo rastas šiuose virusuose (Thomas, Rolando et al., 2008). Taigi, panašu, kad minėti du aktyvumai PhiKZ-like

faguose koduojami skirtinguose polipeptiduose (3.4 pav.). PhiKZ-like virusų numanomų DNR polimerazių konservatyvių motyvų analizė parodė, kad joms būdingos katalizei būtinos liekanos (3.4 pav.). Kaip ir RB69 gp43 DNR polimerazei, phiKZ polimerazėms yra būdingi konservatyvūs aspartatai, kurie naudojami metalo jonų surišimui (3.4 pav., B ir C).



3.4 pav. RB69 ir PhiKZ fagų DNR polimerazių palyginimas. (A) *Pseudomonas phage EL* 3'-5' egz nukleazės (gi: 82700984) bei DNR polimerazės ir RB69 fago DNR polimerazės domeninės organizacijos palyginimas. N, N-galas; P, delnas (palm); F, pirštai (fingers); T, nykštys (thumb). Žvaigždutė žymi polimerazės aktyvųjį centrą (D229 ir D398). (B) DNR polimerazių aktyviojo centro palyginys. Sekos pavadinimas sudarytas iš fago akronimo, baltymo pavadinimo ir jo gi numerio (PDB kodo RB69 fago atveju). (C) *Pseudomonas phage EL* DNR polimerazės aktyviojo centro modelis su DNR ir įeinančiu dTTP sudarytas šablonu naudojant RB69 DNR polimerazės kristalinę struktūrą (PDB kodas: 1ig9). Aktyviojo centro aspartatų liekanos vaizduojamos rožinėmis lazdelėmis, metalo jonai – žaliomis sferomis.

Dažnai DNR polimerazės su procesyvumo veiksniais sąveikauja per trumpą C-gale esantį motyvą (Dalrymple, Kongsuwan et al., 2001). Panašų motyvą (pvz., 201phi2-1 fago atveju – „TRLISDFY“ (pagrindinės hidrofobinės pozicijos pabrauktos)) savo C-gale turėjo ir phiKZ DNR polimerazės. Tačiau

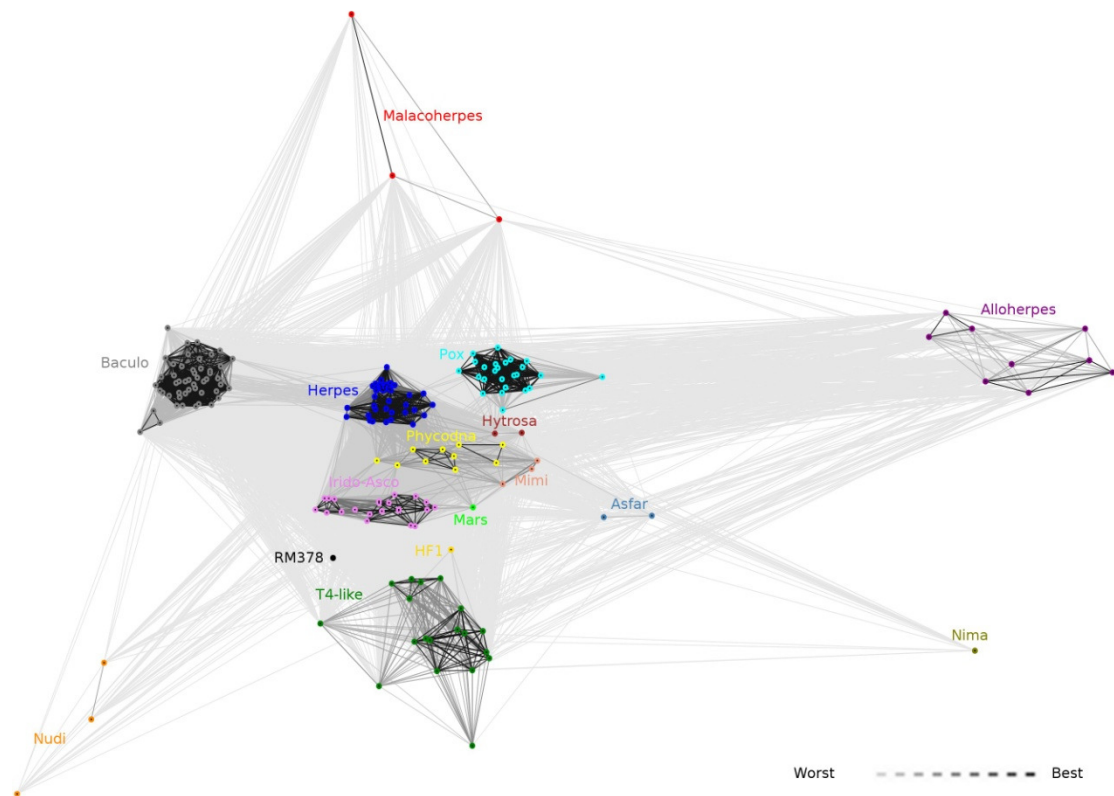
funkcinė šio motyvo svarba nėra aiški, nes viename iš phiKZ-like virusų jo aromatinės a-r. yra pakeistos į hidrofilingas. Be to procesyvumo veiksniai šiuose virusuose nebuvo rasti. Tačiau gali būti, kad jie yra stipriai divergavę.

A šeimos polimerazės sudaro tris grupes. Didžiausia įvairovė stebima PolAgr1 grupėje, į kurią patenka phiKMV, L5, N4, T5, SPO1, RSL1 ir Ma-LMM01 fagų baltymai. Įdomu tai, kad SPO1 DNR polimerazė savo N-gale turi papildomą uracil-DNR glikozilazės (UDG) domeną. Manoma, kad jis galėtų veikti kaip procesyvumo veiksnys (Weigel ir Seitz, 2006). Analizė rodo, kad T5 fago DNR polimerazė, kuri yra labai procesyvi (Andraos, Tabor et al., 2004), taip pat turi į UDG panašų domeną savo N-gale. Tikimybė, kad UDG domenai gali atlikti vidinio procesyvumo veiksnio funkciją, sustiprina raupų virusų pavyzdys. Žinoma, kad raupų virusų procesyvumo veiksnys sudarytas iš UDG (D4) ir A20 baltymo (Druck Shudofsky, Silverman et al., 2010). Antra ir trečia A šeimos baltymų grupės yra sudarytos atitinkamai iš T7-like ir Bcep1-like virusų baltymų.

Virusų C šeimos DNR polimerazėms yra būdinga *E. coli* polIII α domeninė organizacija (Lamers, Georgescu et al., 2006). Virusų PolC sekų konservatyvumas prasideda nuo N-galo PHP domeno, apima polimerazės aktyvų centrą ir baigiasi už „pirštų“ subdomeno. Tikėtina, kad virusų polIII α homologai savo C-gale turi kitokių nei polIII α struktūrinių domenų. Geriausiai išsilygino *E. coli* polIII α ir *Bacillus phage 0305phi8-36* DNR polimerazės (gi: 154622917) C-galai. Be to ši polimerazė turi sekos motyvą (1131-EEDLL-1135), kuris išsilygina su *E. coli* polIII α motyvu (920-QADMF-924) sąveikaujančiu su *E. coli* procesyvumo veiksniumi (polIII β). *Bacillus phage 0305phi8-36* yra vienintelis virusas, kuriame rasta C šeimos DNR polimerazė ir polIII β procesyvumo veiksnys (gi: 154622720), todėl tikėtina, kad šie du komponentai veikia drauge.

PolBrCore polimerazių pogrupiai. Naudojant griežtesnę sekų klasterizavimo procedūrą (CLANS programa atliekant BLAST, o ne PSI-BLAST sekų paieškas) buvo išskirti PolBrCore polimerazių pogrupiai (pav. 3.5). Kadangi PolBrCore daugiausia sudaro didelių genomų virusų PolBr

sekos, analizuojant necharakterizuotų PolBrCore pograpių atstovus, tikėtasi geriau suprasti kaip šie virusai užtikrina savo DNR replikacijos procesyvumą. Žinoma, kad T4-like fagų ir herpes virusų DNR polimerazės sąveikai su procesyvumo veiksniumi naudoja savo C-galą (Dalrymple, Kongsuwan et al., 2001). Todėl buvo nuspręsta paieškoti sąveikos su procesyvumo veiksniumi motyvų PolBrCore pograpiuose. Su PCNA sąveikaujantis motyvas buvo rastas phycodna virusų polimerazių C-galuose (konsensuso seka: QxxIxxFF, kur x yra bet kokia a-r.). Kitų DNR polimerazių subgrupių atvejais minėti sąveikos motyvai nebuvo rasti. Ištyrus PolBrCore subgrupių domeninę organizaciją, trijose divergavusiose grupėse (pav. 3.5, *Malacoherpesviridae*, *Alloherpesviridae* ir *Nimaviridae* šeimos) nustatyti papildomi sekos regionai, nebūdingi charakterizuotiems PolB atstovams. Deja, nepavyko šiuose regionuose identifikuoti domenų, tačiau pats jų buvimas rodo, kad šios polimerazės galbūt „išrado“ naujų būdų didelių genomų replikacijos procesyvumui užtikrinti.

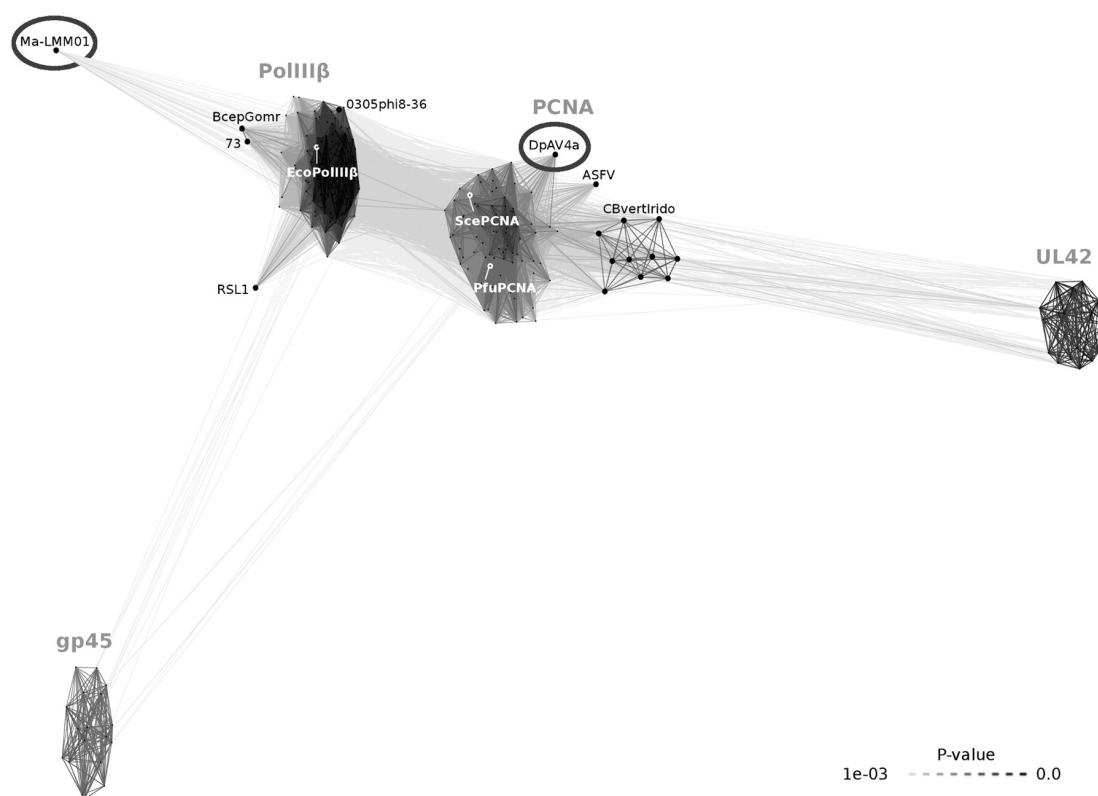


3.5 pav. PolBrCore klasterio pogrupiai. Sekos, priklausančios tam pačiam pogrupiui, nuspalvintos vienodai. Pogrupių santrumpos: Alloherpes – *Alloherpesviridae*, Asco – *Ascoviridae*, Asfar – *Asfarviridae*, Baculo – *Baculoviridae*, Herpes – *Herpesviridae*, HF1 – Halovirus HF1, Hytrosa – *Hytrosaviridae*, Irido – *Iridoviridae*, Malacoherpes – *Malacoherpesviridae*, Mars – *Marseillevirus*, Mimi – *Mimiviridae*, Nima – *Nimaviridae*, Nudi – *Nudiviruses*, Phycodna – *Phycodnaviridae*, Pox – *Poxviridae*, RM378 – *Rhodothermus phage RM378*, T4-like – T4-like fagai.

3.1.1. Procesyvumo veiksniai

Įvairovė ir taksonominis pasiskirstymas. Virusų genomuose buvo ieškoma charakterizuotų ląstelių (PCNA ir polIII β) bei virusinių (gp45, UL42, UL44 ir BMRF1) procesyvumo veiksnių (PV) homologų. Analizės metu naujai rasti į procesyvumo veiksnius reikšmingai panašūs baltymai: PCNA homologas *Ascoviridae* šeimoje ir polIII β Ma-LMM01 fage. Visi rasti baltymai buvo papildyti charakterizuotais ląstelių organizmų procesyvumo veiksniais ir suklastertizuoti (3.6 pav.). Kaip ir DNR polimerazių atveju, virusų procesyvumo veiksnių įvairovė yra didesnė nei ląstelių organizmų atitinkamų baltymų. PolIII β homologai rasti tik faguose, o visi PCNA homologai (išskyrus PhiCh1 ir bakulo virusų baltymus) nustatyti NCLD virusų grupėje (3.1 pav.).

Irido virusų, infekuojančių šaltakraujus stuburinius, PCNA formuoja atskirą grupę (3.6 pav., CBvertIrido). PCNA ir polIIIβ tolimi homologai – gp45 ir UL42, sudaro kompaktiškas grupes. Gp45 grupei priklauso T4-like fagų, o UL42 – herpes virusų procesyvumo veiksniai. Pastarosios grupės turi struktūriškai charakterizuotų atstovų (Moarefi, Jeruzalmi et al., 2000; Zuccola, Filman et al., 2000). Trys divergavusios procesyvumo veiksnių grupės (UL44, BMRF1 ir G8R) nepavaizduotos klasteryje, nes grupavimo metu nebuvo rasta panašumo tarp jų ir kitų baltymų grupių. Vis dėlto, yra žinoma, kad UL44 ir BMRF1 yra struktūriškai panašūs į UL42 ir kitus DNR procesyvumo veiksnius (Appleton, Loregian et al., 2004; Murayama, Nakayama et al., 2009). G8R yra tolimas PCNA homologas (Iyer, Aravind et al., 2001), randamas vaccinia ir kituose *Chordopoxvirinae* šeimos virusuose. Žinoma, kad replikacijos metu jis neatlieka procesyvumo veiksnio funkcijos (Boyle ir Traktman, 2009).



3.6 pav. Virusų ir reprezentatyvūs ląstelių organizmų procesyvumo veiksniai sugrupuoti pagal tarpusavio sekų panašumą. Naujai rasti procesyvumo veiksniai pažymėti elipsėmis. Santrumpos: Ma-LMM01 – *Microcystis phage Ma-LMM01*; RSL1 – *Ralstonia phage RSL1*; 73 – *Pseudomonas phage 73*; BcepGomr – *Burkholderia phage BcepGomr*; 0305phi8-36 – *Bacillus phage 0305phi8-36*; Eco – *Escherichia coli*; ASFV – *African swine fever virus*; DpAV4a – *Diadromus pulchellus ascovirus 4a*; CBvertIrido – šaltakraujų stuburinių gyvūnų irido virusai.

Atlikus asko viruso DpAV-4a genomo transliaciją jame rastas PCNA homologas (3.6 pav., DpAV4a). PCNA homologai buvo rasti ir kituose asko virusuose: HvAV-3e (gi: 134287330) ir SfAV-1a (gi: 11932043 ir 11932044). Tačiau šie baltymai blogai išsilygina su kitais PCNA ir yra nepilno ilgio. Be to, SfAV-1a viruse PCNA homologo seka suskaldyta į du atviro skaitymo rėmelius. Taigi, tikėtina, kad šių dviejų virusų minėti baltymai yra nefunkcionalūs. Kai kuriuose virusuose buvo rasta keletas PCNA kopijų. Ehv-86 ir PBCV-1 virusai iš *Phycodnaviridae* šeimos turi du, o mimi virusas – net tris PCNA homologus (Lentelė 3.1). MimiPCNA1 ir PBCV1_PCNA1 yra panašesni į PCNA virusų (*Ostreococcus virus OsV5* ir CroV), turinčių vieną šio baltymo kopiją. Todėl tikėtina, kad mimi ir PBCV-1 PCNA1 yra ortologai, būtini šių virusų DNR replikacijai. Kita vertus, PBCV1_PCNA2 ir EhV86_PCNA2 rodo panašumą į dumblių PCNA, todėl gali būti, kad jie yra perimti iš šeimininko. MimiPCNA2 ir MimiPCNA3 panašiausi į MimiPCNA1, todėl tikriausiai jie atsirado po genų ir genomo duplikacijų, įvykusių mimi viruso evoliucijos eigoje (Suhre, 2005).

Lentelė 3.1 *Paramecium bursaria Chlorella virus 1* (PBCV-1), *Emiliana huxleyi virus 86* (EhV-86) ir mimi viruso PCNA homologų savybės.

PCNA pavadinimas	PBCV-1		EhV-86		Mimi virusas		
	PBCV1_PCNA1	PBCV1_PCNA2	EhV86_PCNA1	EhV86_PCNA2	MimiPCNA1	MimiPCNA2	MimiPCNA3
PCNA gi	9631761	9632127	73852914	73852490	55664866	55417433	55416733
PCNA ilgis (a-r.)	262	264	288	259	464	323	273
PCNA pI*	5	4.28	4.21	4.66	5.46	7.34	9.35

* izoelektrinis taškas

PolIIIβ homologai buvo rasti tik dvylikoje fagų. Penki iš jų yra trumpesni ir apima tik antrą ir trečią polIIIβ domenų. Pirmą kartą Ma-LMM01 fage buvo

rastas (HHsearch tikimybė – 96%) pilno ilgio polIIIβ homologas. Šis baltymas (MaLMM01_gp176) yra koduojamas greta Ma-LMM01 fago replikacinių baltymų (Yoshida, Nagasaki et al., 2008), todėl, tikėtina, kad jis atlieka procesyvumo veiksnio funkciją.

Dalis virusų procesyvumo veiksmių galbūt buvo įgyti horizontalios genų pernašos (HGP) būdu. Pastariesiems baltymams būdinga fragmentiška taksonominė sklaida, didelis panašumas į atitinkamus šeimininko baltymus bei DNR polimerazės nebuvimas genome. Pavyzdžiui, tik devyni iš 53 bakulo virusų, infekuojančių vabzdžius, turi PCNA homologus, o septyni iš jų, rodo panašumą į uodų ir kandžių PCNA (3.7 pav.). Beje, *Autographa californica nucleopolyhedrovirus* (AcMNPV) tyrimai rodo, kad šio viruso PCNA nereikalingas DNR geno replikacijai (Kool, Ahrens et al., 1994). Kadangi PCNA ir polIIIβ, įgyti HGP būdu, galbūt tiesiogiai nedalyvauja DNR replikacijoje, todėl jie nebuvo įtraukti į 3.1 ir 3.2 paveiksluose pateiktą analizę.

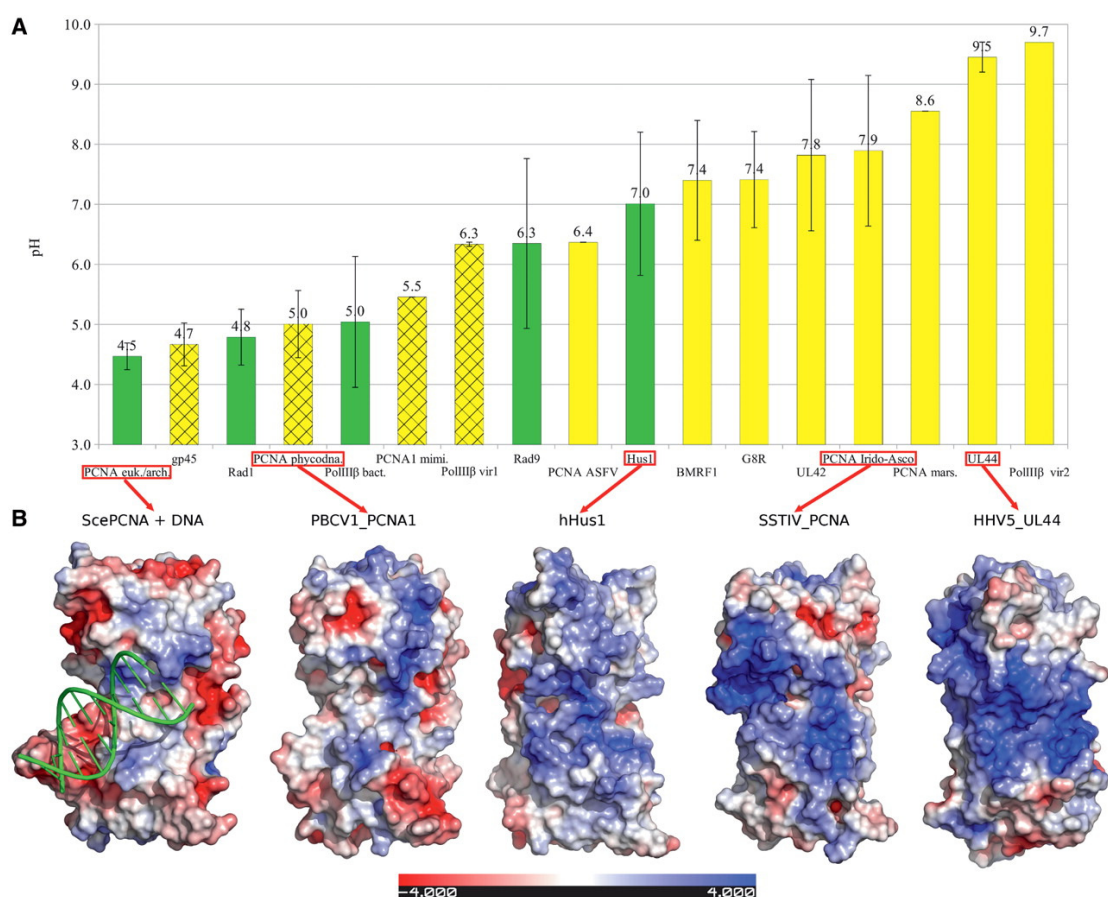
PCNA_Scer_536356	1 MLEAKFEEASLFRRIIDGFKDCVOLVNFCKEGLIITADAVDDSRVLLVLSLEIGVAFQYRCDHPVTLGMDLTLSSKILRCGNNTDILLIIA92
PCNA_Hsap_4505641	1 MFEARLVQGSILKXVLEALKDLINFAWIDISSGVNLSMDSSHVSLVQLTRSEGFDYRCDRNLAMGVNLTSMKILKACAGNEDITLIRA92
PCNA_Dsec_195336048	1 MFEARLVQATILKXVLEALKDLINFAWIDISSGVNLSMDSSHVSLVQLTRSEGFDYRCDRNLAMGVNLTSMKILKACAGNEDITLIRA92
PCNA_Aaeg_157167857	1 MFEARLVQATILKXVLEALKDLINFAWIDISSGVNLSMDSSHVSLVQLTRSEGFDYRCDRNLAMGVNLTSMKILKACAGNEDITLIRA92
PCNA_Aalb_67866971	1 MFEARLVASTTLKXVLEALKDLINFAWIDISSGVNLSMDSSHVSLVQLTRSEGFDYRCDRNLAMGVNLTSMKILKACAGNEDITLIRA92
PCNA_Cqui_170046021	1 MFEARLVASTTLKXVLEALKDLINFAWIDISSGVNLSMDSSHVSLVQLTRSEGFDYRCDRNLAMGVNLTSMKILKACAGNEDITLIRA92
PCNA_Bmor_112984050	1 MFEARLLRSSILKXVLEALKDLINFAWIDISSGVNLSMDSSHVSLVQLTRSEGFDYRCDRNLAMGVNLTSMKILKACAGNEDITLIRA92
PCNA_Hcun_21717396	1 MFEARLLRSSILKXVLEALKDLINFAWIDISSGVNLSMDSSHVSLVQLTRSEGFDYRCDRNLAMGVNLTSMKILKACAGNEDITLIRA92
PCNA_Sfru_21717394	1 MFEARLLRSSILKXVLEALKDLINFAWIDISSGVNLSMDSSHVSLVQLTRSEGFDYRCDRNLAMGVNLTSMKILKACAGNEDITLIRA92
PCNA_RoMNPV_23577875	1 MFEAEFKTGAVALRLEVFKDLLPHATFDQDDGGVSMVMDTSHVALVSLQVLAHAGFKKRCRNLVNLVNSLSKILKACAGNEDITLIRA92
PCNA_PixyMNPV_114680103	1 MFEAEFKTGAVALRLEVFKDLLPHATFDQDDGGVSMVMDTSHVALVSLQVLAHAGFKKRCRNLVNLVNSLSKILKACAGNEDITLIRA92
PCNA_AcMNPV_9627791	1 MFEAEFKTGAVALRLEVFKDLLPHATFDQDDGGVSMVMDTSHVALVSLQVLAHAGFKKRCRNLVNLVNSLSKILKACAGNEDITLIRA92
PCNA_AnpeNPV_96979862	1 MVEATFATAAELGVDAVRLGSLHATFDQDDGGVSMVMDTSHVALVSLQVLAHAGFKKRCRNLVNLVNSLSKILKACAGNEDITLIRA92
PCNA_CfMNPV_30387279	1 -MEATFATAGALNIVDALRGLISYATFDQDDGGVSMVMDTSHVALVSLQVLAHAGFKKRCRNLVNLVNSLSKILKACAGNEDITLIRA92
PCNA_HycuNPV_86355623	1 -MEVTFATAGSLNIVDALRGLLTHATFDQDDGGVSMVMDTSHVALVSLQVLAHAGFKKRCRNLVNLVNSLSKILKACAGNEDITLIRA92
PCNA_OpMNPV_9629991	1 -MEATFATAAAAFOSIVGALRGLLTHATFDQDDGGVSMVMDTSHVALVSLQVLAHAGFKKRCRNLVNLVNSLSKILKACAGNEDITLIRA92
PCNA_TnSNPV_74229749	1 MFEAEFKNVI AFKXAMEVI ESVVDVY TLEITSKGISMOCSDTRISFIFKTMKKEAFKSFETT KNL SLSVKMSGFCIKLKTADTSMLTITWA92
PCNA_ChcNPV_68304202	1 MFEAEFKNVI AFKXAMEVI ESVVDVY TLEITSKGISMOCSDTRISFIFKTMKKEAFKSFETT KNL SLSVKMSGFCIKLKTADTSMLTITWA92
PCNA_Scer_536356	93 DN - - - - - T P S T I L L F E D T K K D I A E Y S I K L M D I D A D F K I E - F L Q D S T L S L P S E F S K I V R D L S Q L S D - S I N I M I L K E T I R F V A D G D I G - S G S 179
PCNA_Hsap_4505641	93 ED - - N A D T L A L V F E A R N Q E K V S D Y E M K L M D L D V E Q L G P - E Q E Y S C V V K M P S G E F A R I C R D L S H I G D - A V V I S C A K O G V F S A S G E L G - N G N 179
PCNA_Dsec_195336048	93 OD - - N A D T V T I M F E A N Q E K V S D Y E M K L M D L D E H L G P - E T D F S C V V R M P A M E F A R I C R D L A F S E - S V V I C C T K E G V K F S A S G D A G - I A N 179
PCNA_Aaeg_157167857	93 OD - - N A D T V T F M F E S O N Q E K V S D Y E M K L M D L D E H L G P - E T D Y A C V V R M P A M E F A R I C R D L S Q F G E - S V V I S C T K E G V K F S A S G D A G - S A N 179
PCNA_Aalb_67866971	93 OD - - N A D T V T F M F E S O N Q E K V S D Y E M K L M D L D E H L G P - E T D Y A C I V R M P A I E F A R I C R D L S Q F G E - S V V I S C T K E G V K F S A S G D A G - S A N 179
PCNA_Cqui_170046021	93 OD - - N A D T V T F M F E S O N Q E K V S D Y E M K L M D L D E H L G P - E T D Y A C I V R M P A I E F A R I C R D L S Q F G E - S V V I S C T K E G V K F S A S G D A G - S A N 179
PCNA_Bmor_112984050	93 OD - - N A D N V T F V F E S P N Q E K V S D Y E M K L M D L D E H L G P - E T E S C T I R L P S E F A R I C R D L S Q F G E - S M V I S C T K E G V K F S A T G D I G - S A N 179
PCNA_Hcun_21717396	93 OD - - N A D T V T F V F E S P N Q E K V S D Y E M K L M D L D E H L G P - E T E S C T I R L P S E F A R I C R D L S Q F G E - S M V I S C T K E G V K F S A T G D I G - S A N 179
PCNA_Sfru_21717394	93 OD - - N A D T V T F V F E S P N Q E K V S D Y E M K L M D L D E H L G P - E T E S C T I R L P S E F A R I C R D L S Q F G E - S M V I S C T K E G V K F S A T G D I G - S A N 179
PCNA_RoMNPV_23577875	93 ED - - O G D V M A F V F N N D - - N R I C T Y T L K L I C I D V E H L G P - D S D Y D C V V H M S S V E F A Q V C K M T Q F D H - D I I V S C S K K G L F R A N G D I G - S A D 177
PCNA_PixyMNPV_114680103	93 ED - - O G D V M A F V F N N D - - N R I C T Y T L K L I C I D V E H L G P - D S D Y D C V V H M S S V E F A Q V C K M T Q F D H - D I I V S C S K K G L F R A N G D I G - S A D 177
PCNA_AcMNPV_9627791	93 ED - - O G D V M A F V F N N D - - N R I C T Y T L K L I C I D V E H L G P - D S D Y D C V V H M S S V E F A Q V C K M T Q F D H - D I I V S C S K K G L F R A N G D I G - S A D 177
PCNA_AnpeNPV_96979862	93 DA - - R D H V H F F - - - - - R A R A L T L A Q V S L D R V L G P - T A K P C A L T V M S E L A R A C D L Q R L G A T M L A A G S A G L L A A D A G G V R A 175
PCNA_CfMNPV_30387279	92 SA - - R D R L H F K F K A - - R A V A G A L A O S L D V E R L G M L - E N D C V L A V M D E A V K C R D L M L R L A G A V E V S S G A A L C F A A D A D G V R A S 177
PCNA_HycuNPV_86355623	92 SA - - R D D V H F A F K A - - E R I V T C A L S O I S L D V E R L G V P D D D E D C V L A V A S D A W A V G S D U A L D A T V V E L S S G A A L C F A A D A D G V R A S 179
PCNA_OpMNPV_9629991	92 AA - - R D R V R L A Y E A - - R R A V S C T L A O I S L D A D R L G P - D K E T C V L A V P S A E L R T A C A D I A R L G A T V E M S S G T T G L S F A A H A D D G V R V 178
PCNA_TnSNPV_74229749	93 GE - - K K D H L N F K I K I S - - N S L R L Y E Y K L F N F D S O D Y G M S D D L E V T A R A V M R S A E L R L C R H L D L L G N L N V G I K I D I B K I N I S R G D E V - N V N 179
PCNA_ChcNPV_68304202	93 G E N D S H Y L N F K I K I G - - K S L K S Y E Y K L Y H F D S O D Y G M S D L Q V T R A M M H S V E L K L C R H L D L L G N L N V G I K I D I B K I I F N S K G D E I - N V N 181
PCNA_Scer_536356	180 V I I K F V D M E P E T S I K L E M D P V D - L T F G A K Y L L D I I R G S S L D R V G I R L S E A P L F O F D L K - - - - S G F L O F F L A P I F N D E E - - - - 258
PCNA_Hsap_4505641	180 I K L S O T S V B K E E E A V I I E M Q E P V T - L T F A L R Y L F P F F A K A P L S T V T I L S M S A D V P L V V E Y K I A - - - - D M G K L K Y L A P R I E D E E S - - - - 261
PCNA_Dsec_195336048	180 I K L A O T S V D K E E E A V I I E M Q E P V T - L T F A C R Y L N F S F T K A P L S S Q V O L S M S A D V P L V V E Y K I A - - - - D L G H I R Y Y L A P R I E D E E T - - - - 260
PCNA_Aaeg_157167857	180 I K L A Q T S V D K E E E A V I I E M Q E P V T - L T F A C R Y L N F S F T K A P L S S Q V O L S M S A D V P L V V E Y K I O - - - - D L G H I R Y Y L A P R I E D D E N - - - - 260
PCNA_Aalb_67866971	180 I K L A Q T S V D K E E E A V I I E M Q E P V T - L T F A C R Y L N F S F T K A P L S S Q V O L S M S A D V P L V V E Y K I O - - - - D L G H I R Y Y L A P R I E D D E N - - - - 260
PCNA_Cqui_170046021	180 I K L A Q T S V D K E E E A V I I E M Q E P V T - L T F A C R Y L N F S F T K A P L S S Q V O L S M S A D V P L V V E Y K I O - - - - D L G H I R Y Y L A P R I E D D E N - - - - 260
PCNA_Bmor_112984050	180 V K L A Q T A S I D K E E E A V I I E M E E P V T - L T F A C O Y L N Y F T K A P L S S Q V O L S M S A D V P L V V E R I P - - - - D I G H I R Y Y L A P R I E E E D S - - - - 260
PCNA_Hcun_21717396	180 I K L A Q T A S I D K E E E A V I I E M E E P V T - L T F A C O Y L N Y F T K A P L S S Q V O L S M S A D V P L V V E R I P - - - - D I G H I R Y Y L A P R I E E E D S - - - - 260
PCNA_Sfru_21717394	180 I K L A Q T A S I D K E E E A V I I E M E E P V T - L T F A C O Y L N Y F T K A P L S S Q V O L S M S A D V P L V V E R I P - - - - D I G H I R Y Y L A P R I E E E D S - - - - 260
PCNA_RoMNPV_23577875	178 V O M S A D S - - - - - E N F S V L K K A K O I V T - H T F A G D Y L C H F A A A S L A P T V T I Y M S E E P P K L E Y C I K - - - - D V G L A C F L A P R I V N N D E E I F 256
PCNA_PixyMNPV_114680103	178 V O M S A D N - - - - - E N F S V L K K A K O I V T - H T F A G D Y L C H F A A A S L A P T V T I Y M S E E P P K L E Y C I K - - - - D V G L A C F L A P R I V N N D E E I F 256
PCNA_AcMNPV_9627791	178 V O M S A D N - - - - - E N F S V L K K A K O I V T - H T F A G D Y L C H F A A A S L A P T V T I Y M S E E P P K L E Y C I K - - - - D V G L A C F L A P R I V N N D E E I F 256
PCNA_AnpeNPV_96979862	176 V I L R - - - - - - - - - - - T P A R P L A - Q E F A V L Y L T L A R A A P L A P T V T I Y M S E E L P P K L E Y C I K - - - - S H G A L D Y L A P L A R S D R D K A Q 246
PCNA_CfMNPV_30387279	178 V L L R - - - - - - - - - - - V A P R H L T - Q V F A C S Y L N T F A R A C I L S K D V D C M T A D A P L R L R F C I G - - - - Q L G L D L Y L A P M R I D V - - - - 244
PCNA_HycuNPV_86355623	180 V L L R - - - - - - - - - - - A A P R R P L T - Q A F A C R Y L N F A A C A P L S K F V N V C S M A N A P L R F C L E - - - - R L G L D L Y L A P R V S S A K D B O 250
PCNA_OpMNPV_9629991	179 V L L R - - - - - - - - - - - S P R R P L L - Q A F A C R Y L N F A A A L S E T V D V C H D A S M P L R L R F R L E - - - - P L G L D L Y L A P R V S S A K D B O 249
PCNA_TnSNPV_74229749	180 Y A I K - - - - - E G E S E E N C I E N A T E P T D A I G E N K Y L Y N Y K - - N P L N S L V E I S I I E K O L L K L K Y Y L D N E N K D D G Y I E L Y I A S Y V N E D M - - - - 261
PCNA_ChcNPV_68304202	182 Y V V K - - - - - E G E S E E N C A I D S A T E P T D M I E N N K K F Y N Y N Y K - - N P I N S F V E I S I I E K L L L K K Y F L D N E N R D D G Y M E L Y I A S Y V N E D I - - - - 263

3.7 pav. Bakulo virusų (raudonas fonas) ir eukariotų (žalias fonas) PCNA sekų palyginys. PCNA motyvas būdingas abiem grupėms pateiktas geltoname fone.

Nuostabą kėlė tai, kad procesyvumo veiksniai nerasti kai kuriose dideliu genomu turinčių virusų grupėse (*Nimaviridae*, *Alloherpesviridae*, ir *Malacoherpesviridae* šeimos bei phiKZ-like ir *Clostridium phage c-st* fagas). PCNA nebuvimą minėtose eukariotų virusų šeimose bei *Clostridium phage c-st* fage galima paaiškinti jų netipiškais DNR polimerazėmis, o phiKZ fagai galbūt turi stipriai divergavusį procesyvumo veiksnį, kurio nepavyko rasti.

Elektrostatinės savybės. Procesyvumo veiksmų taksonominis paplitimas (3.1 pav.) rodo, kad *Bacillus phage 0305phi8-36* ir kelios eukariotų virusų šeimos turi ląstelinių procesyvumo veiksmų homologus, bet žiedo užkėlimo kompleksai juose nerasti. Kaip galėtų veikti minėti PCNA/polIIIβ homologai? Galbūt jie naudoja šeimininko žiedo užkėlėjus? Kita vertus jie galėtų neformuoti žiedo ir su DNR sąveikauti kaip herpes virusų UL42 ir UL44. Vienas iš žiedinių ir nežiedinių (UL42 ir UL44) procesyvumo veiksmų skirtumų yra pastaruosiuose padidėjęs teigiamas krūvis sąveikos su DNR pusėje (Loregian, Sinigalia et al., 2007; Komazin-Meredith, Santos et al., 2008). Siekiant iširti identifikuotų procesyvumo veiksmų elektrostatines savybes, jiems visiems buvo paskaičiuotas teorinis izoelektrinis taškas (pI). Be to, reprezentatyviems PCNA homologams buvo sudaryti jų struktūros modeliai ir iširta jų paviršiaus elektrostatika. Gauti duomenys palyginti su charakterizuotais ląstelinių organizmų ir virusų procesyvumo veiksniais (3.8 pav.). Paaiškėjo, kad virusų procesyvumo veiksmų pI stipriai koreliuoja su žiedo užkėlėjo subvienetų (ne)buvimu. Taigi, mimi ir phycodna virusų ortologiškai PCNA turi elektrostatines savybes artimas žiediniams PV. Tuo tarpu G8R ir PCNA homologai iš irido, asko, asfar bei marselio virusų savo elektrostatika yra panašesni į nežiedinius herpes virusų PV. Phycodna ir mimi virusai turi žiedo užkėlėjo (RFC) homologus, o nežiedinių procesyvumo veiksmų virusų grupėms šie baltymai nebūdingi. Bakteriofaguose taip pat aptikta minėta koreliacija. Ma-LMM01 ir RSL1 fagu, turinčių žiedo užkėlėjus, polIIIβ (3.8 pav., PolIIIβ vir1) pI yra mažesnis nei pI polIIIβ (PolIIIβ vir2) fago (0305phi8-36), kuris žiedo užkėlėjo neturėjo. Taigi, elektrostatinės savybės rodo, kad phycodna ir mimi virusų PV, tikėtina, formuoja žiedus, o

PCNA homologai iš kitų virusų šeimų bei fago 0305phi8-36 polIIIβ - neformuoja. Pagal savo pI įverčius Ma-LMM01 ir RSL1 polIIIβ atsiduria tarpinėje padėtyje tarp žiedus formuojančių ir nežiedinių PV. Tačiau pastaruose faguose yra rasti žiedo užkėlėjo subvienetai (polIIIγ), todėl šiems virusams labiau tikėtini žiediniai PV.



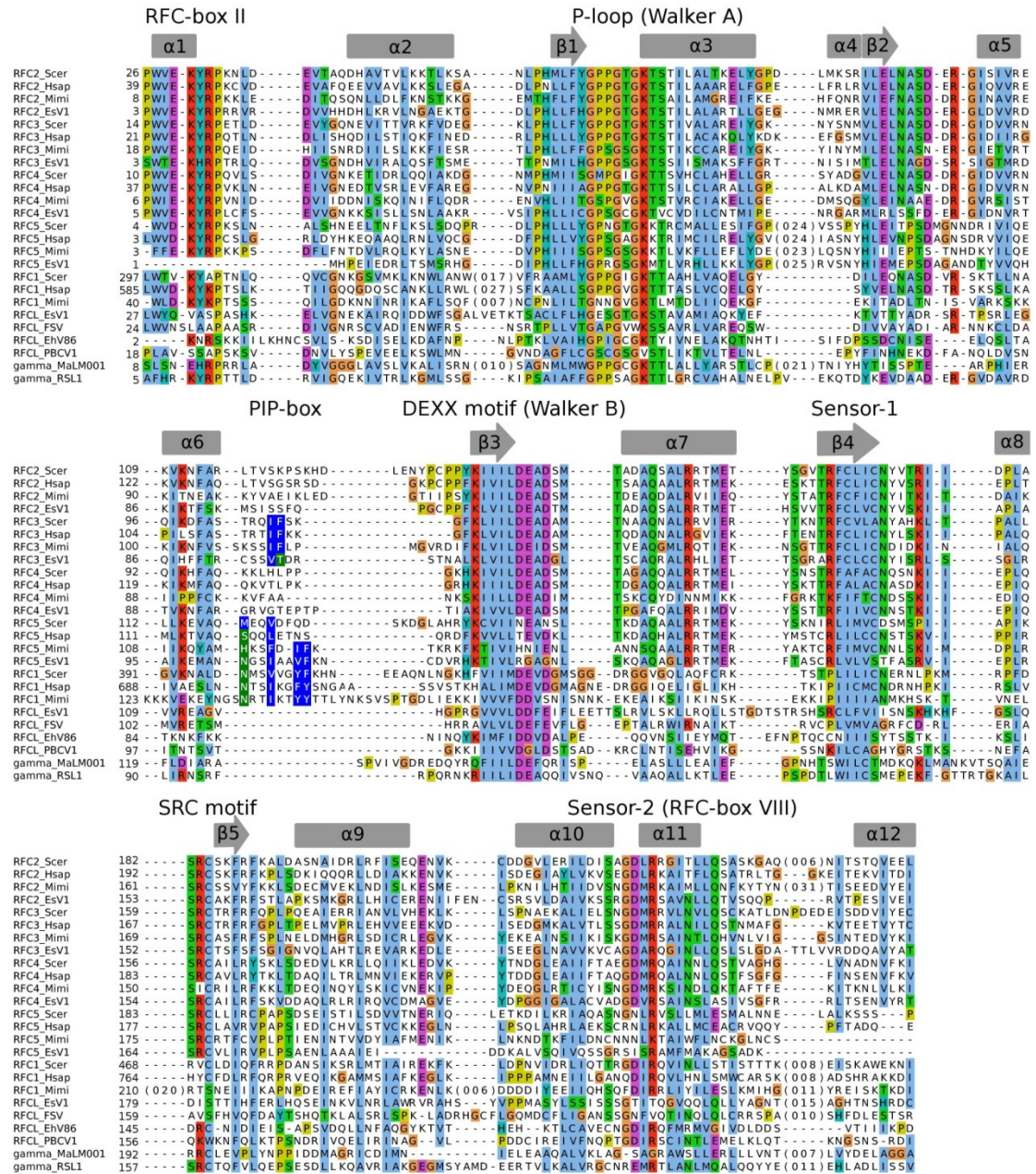
3.8 pav. Procesyvumo veiksmų elektrosstatinės savybės. (A) Virusų (geltoni stulpeliai) ir ląstelinių organizmų (žali stulpeliai) procesyvumo veiksmų teorinio pI vidurkiai. Grotelėmis pažymėti virusų, kurie turi žiedo užkėlimo subvienetų, procesyvumo veiksmiai. (B) Tirpikliui prieinamo paviršiaus elektrosstatinio potencialo žemėlapis (raudona spalva, neigiamas krūvis; mėlyna, teigiamas; skalės vienetai, K_bT/e_c). Visos struktūros yra toje pačioje orientacijoje kaip ScePCNA ir DNR kompleksas (PDB kodas: 3k4x). Santrumpos: arch. - *Archaea*, asco. - *Ascoviridae*, ASFV - *African swine fever virus*, euk. - *Eukarya*, hHus1 - *Homo sapiens* Hus1 (PDB kodas: 3g65) HHV5_UL44 - *Human Herpesvirus 5* UL44 (PDB kodas: 1t6l), irido. - *Iridoviridae*, PCNA mars. - *Marseillevirus* PCNA (gi: 284504238), PCNA mim. - *Mimivirus* PCNA (gi: 55664866), PBCV1_PCNA1 - *Paramecium bursaria Chlorella Virus-1* PCNA1 (gi: 9631761), Sce - *S. cerevisiae*, SSTIV_PCNA - *Soft-shelled turtle iridovirus* PCNA (gi: 228861299), PolIIIβvir1 - polIIIβ iš *Microcystis phage Ma-LMM01* ir *Ralstonia phage RSL1* (atitinkamai gi: 117530347, 189233246), PolIIIβvir2 - polIIIβ iš *Bacillus phage 0305phi8-36* (gi: 154622720).

Virusų ir ląstelinių PV elektrostatiinių savybių palyginimas atskleidė įdomų faktą. Žinoma, kad žmogaus DNR reparacijoje dalyvauja Rad9-Hus1-Rad1 (9-1-1) kompleksas, sudarytas iš trijų PCNA homologų. Įdomu tai, kad Hus1 elektrostatinės savybės yra artimesnės nežiediniams, o Rad1 – žiediniams PV (3.8 pav.). Nustatyta, kad 9-1-1 komplekso subvienetai neformuoja homotrimerų (Burtelow, Roos-Mattjus et al., 2001). Be to, nustatyta, kad Rad9-Hus1-Rad1 subvienetai gali sudaryti poras (heterodimerus) (Burtelow, Roos-Mattjus et al., 2001). Šio tyrimo bei eksperimentiniai duomenys leidžia manyti, kad Hus1 ir galbūt Rad9 galėtų sąveikauti su DNR kaip monomerai arba kaip heterodimerinių subkompleksų komponentai. Deja, kol kas nėra Rad9, Hus1 ir Rad1 sąveikos su DNR eksperimentinių duomenų.

3.1.2. Žiedo užkėlimo kompleksai

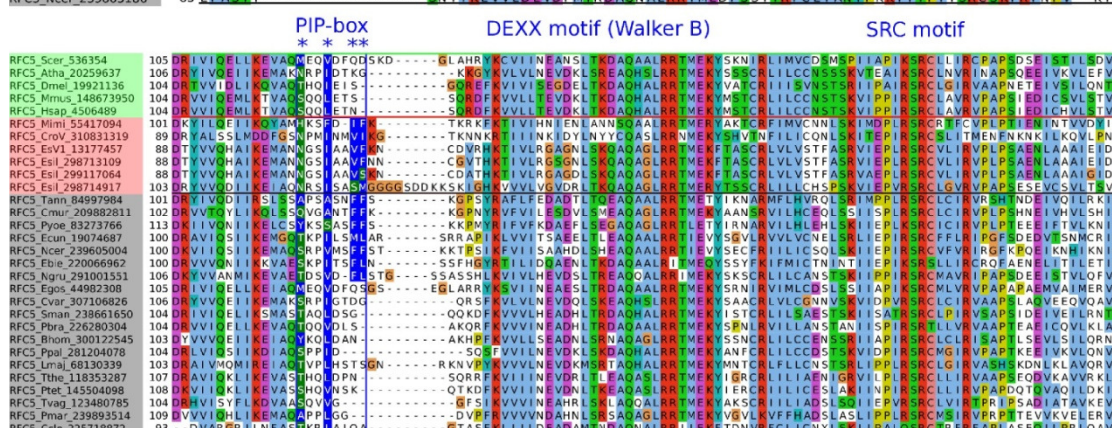
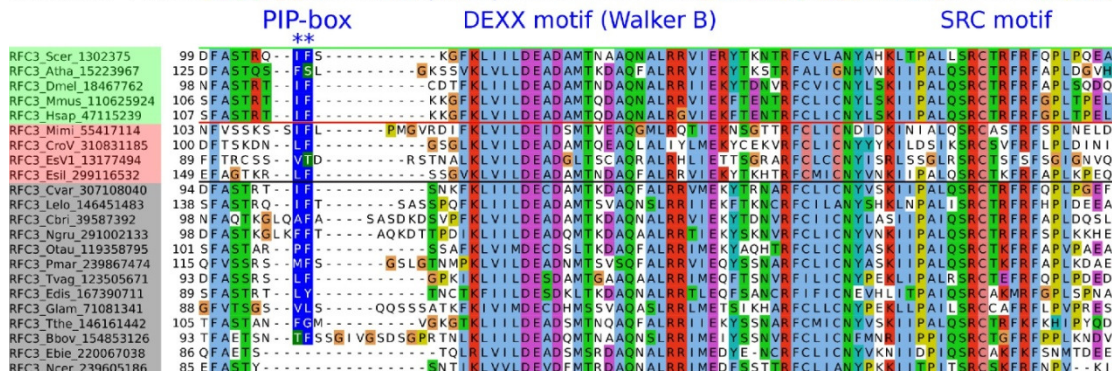
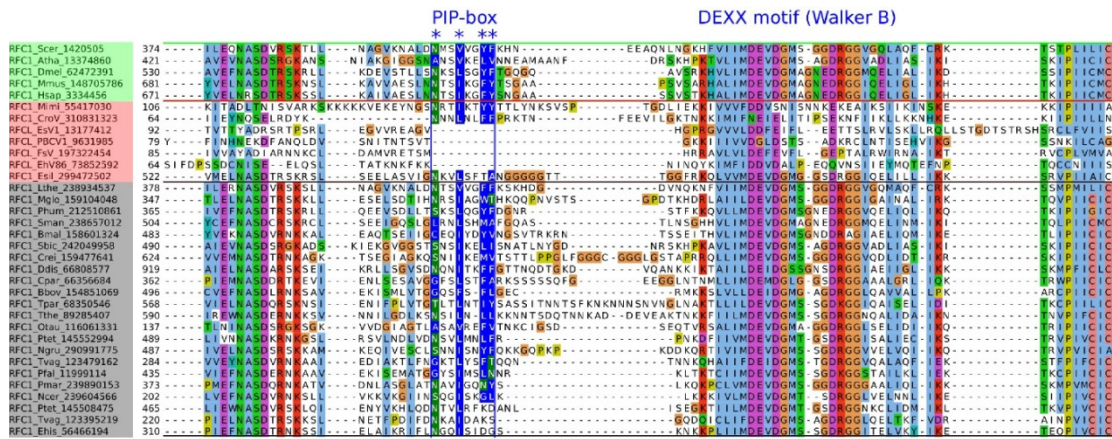
Lyginant su kitais replikazės komponentais, žiedo užkėlimo komplekso (ŽUK) subvienetų virusų genomuose buvo rasta mažiausiai. Tačiau šių baltymų taksonominis paplitimas nėra atsitiktinis. Jie randami tik didžiausių virusų genomuose ir tik kartu su procesyvumo veiksniais (3.1 pav.). Beje, kaip jau minėta aukščiau, ŽUK (ne)buvimas koreliuoja su PV elektrostatinėmis savybėmis (3.8 pav.). Žmogaus RFC subvienetų homologai rasti mimi ir phycodna virusuose (3.1 pav.). Mimi ir jam giminingas CroV virusas turi visus penkis RFC subvienetus. *Phycodnaviridae* šeimos atstovai (išskyrus EsV-1, OsV5 ir *Ostreococcus tauri virus 1*) turi tik vieną RFC subvienetą, giminą didžiajam archėjų RFC subvienetui (RFCL). EsV-1 virusui yra būdingas pilnas RFC kompleksas, o OsV5 ir *Ostreococcus tauri virus 1* virusai ŽUK baltymų neturi. Įdomu tai, kad pastarieji du virusai yra vieni mažiausių savo šeimoje. Bakterijų ŽUK homologai rasti Ma-LMM01 ir RSL1 faguose. Jie turi tik polIIIγ subvienetą. Tikėtina, kad šie baltymai yra aktyvios NTP'azės, nes jiems yra būdinga P-kilpa (3.9 pav., P-loop), DEXX ir SRC motyvai. Kaip jau minėta, pastarųjų fagų procesyvumo veiksmų pI yra mažesnis nei ŽUK subvieneto neturinčio 0305phi8-36 fago polIIIβ atveju (3.8 pav.). T4-like

fagams būdingas ŽUK, sudarytas iš gp44 ir gp62 baltymų, buvo rastas tik T4-like faguose.



3.9 pav. Eukariotų ir virusų žiedo užkėlimo komplekso subvienetų palyginys. Palyginys sudarytas MUSTANG programa (Konagurthu, Whisstock et al., 2006) užklojus eukariotų RFC kristalines struktūras ir virusų baltymų struktūrų modelius. Virš palyginio pateikta mielių RFC3 (PDB kodas: 1sxj) antrinė struktūra. PBCV1, *Paramecium bursaria Chlorella Virus-1*; EsV1, *Ectocarpus siliculosus virus 1*; EhV86, *Emiliana huxleyi virus 86*; FSV, *Feldmania species virus*; Hsap, *Homo sapiens*; MaLM001, *Microcystis phage Ma-LM001*; Mimi, Mimivirus; RSL1, *Ralstonia phage RSL1*; Scer, *Saccharomyces cerevisiae*.

Visi penki mimi ir EsV-1 RFC homologai yra panašūs į atitinkamus žmogaus ir mielių baltymus (3.9 pav.) bei turi sąveikai su ATP (P-loop) ir hidrolizei (DEXX) būtinus motyvus. Vis dėlto, virusiniai ir eukariotiniai RFC turi keletą skirtumų. Struktūriniai mielių RFC ir PCNA komplekso (Bowman, O'Donnell et al., 2004) bei biocheminiai tyrimai (Yao, Coryell et al., 2003; Yao, Johnson et al., 2006) rodo, kad RFC1, RFC3 ir RFC5 sąveikauja su PCNA monomerų hidrofobinėmis kišenėmis. Žmogaus ir mielių atvejais sąveikos su PCNA motyvai (3.9 pav., PIP-box) silpnėja sekančia tvarka: RFC1, RFC3 ir RFC5 (Yao, Coryell et al., 2003; Bowman, O'Donnell et al., 2004; Yao, Johnson et al., 2006). Mimi viruso RFC1 ir RFC3 PIP-motyvai panašūs į eukariotų atitinkamus baltymų regionus, o RFC5 PIP-motyvas panašesnis į RFC1. Įdomu tai, kad EsV-1 turi „stipriausią“ sąveikos su PCNA motyvą RFC5 subvienete, „silpnesnį“ – RFC3, o RFC1 šio motyvo neturi išvis. Kai kuriems eukariotų RFC taip pat būdingas šis netipiškas sąveikos su PCNA „stiprumo“ pasiskirstymas (3.10 pav.). PIP-motyvas nenustatytas phycodna virusuose (FSV, EhV-86 ir *Chlorella*) turinčiuose tik vieną RFC subvienetą. Tikėtina, kad *Chlorella* virusų RFCL yra neaktyvūs, nes jiems būdingos konservatyvių a-r. (Iyer, Leipe et al., 2004) mutacijos P-kilpos ir DEXX motyvuose.



3.10 pav. Eukariotų ir virusų RFC1, RFC3 ir RFC5 sąveikos su žiedu motyvai (PIP-box). PIP motyvo pagrindinės aminorūgšties pažymėtos žvaigždutėmis ir nuspalvintos taip: mėlyna spalva – hidrofobinės ir aromatinės; žalia – hidrofilinės.

3.1.3. DNR replikazių tyrimo rezultatų apibendrinimas ir sąsajos su ląsteliniais organizmais

Tyrimo rezultatai rodo, kad DNR replikazės komponentų (ne)buvimas ir jų tipai yra susiję su viruso genomo dydžiu. Šis ryšys pasireiškia tuo, kad didėjant genomo dydžiui DNR polimerazės ir jų procesyvo komponentai genomuose koduojami dažniau.

Mažų genomų (<40 kbp) virusai dažnai neturi savų DNR polimerazių. Jei šie baltymai randami, jie dažniausiai priklauso PolBp grupei. Įdomu tai, kad tai būdinga labai įvairiems virusams, infekuojantiems visų trijų gyvybės domenų ląsteles. Be to, PolBp polimerazės visiškai nebūdingos didesnių genomų virusams (3.1 ir 3.2 pav.). Taigi, panašu, kad B-šeimoms DNR polimerazės, pradmeniu naudojančios baltymą, yra optimalus sprendimas minėto dydžio genomų replikacijai.

Didėjant genomui (~40–140 kbp) populiariausiomis tampa A šeimos polimerazės. Tačiau neaišku ar šių baltymų dominavimas šiame geno dydžio ruože yra reikšmingas. Dvejonų kelia tai, kad A-šeimoms polimerazės yra randamos tik bakteriofaguose bei tai, kad ~40–140 kbp geno dydžio imtį sudaro daugiausia bakteriofagai. Šio dydžio virusai dar pasižymi tuo, kad dažniausiai neturi procesyvumo veiksnių. Jų nebuvimo negalima paaiškinti tuo, kad PolA polimerazės nenaudoja polIIIβ, nes tyrimai rodo, kad polIIIβ stimuliuoja *E. coli* PolI (Lopez de Saro ir O'Donnell, 2001). Be to, du didelių genomų (>150 kbp) fagai (Ma-LMM01 ir RSL1) turi, tikėtina funkcionalią replikazę, sudarytą iš PolA, polIIIβ ir jo užkėlėjo (polIIIγ). Kita vertus, yra virusų, kurie išvystė alternatyvų būdą procesyvumui užtikrinti. Pavyzdžiui, T7 fago A šeimos DNR polimerazė, kaip procesyvumo veiksnį naudoja *E. coli* tioredoksiną (Bedford, Tabor et al., 1997). SPO1-like ir T5-like fagų PolA polimerazių į UDG panašus domenai galėtų būti dar vienas procesyvumo veiksnio kandidatas, tačiau jį dar reiktų iširti eksperimentiškai.

Didesnių nei 140 kbp genomų grupei priklauso eukariotų virusai ir fagai. Jie visi turi nuosavas, dažniausiai B-šeimoms, DNR polimerazes. Vienintelė šios taisyklės išimtis buvo pašalinta radus stipriai divergavusias phiKZ fagų DNR polimerazes. Šio geno dydžio virusų replikazės dažnai turi procesyvumo veiksnius ir kartais – jų užkėlėjus. Pastarųjų baltymų pasiskirstymas genomuose iš pradžių atrodė atsitiktinis (3.1 pav.). Tačiau DNR polimerazių ir procesyvumo veiksnių savybių analizė atskleidė žiedo užkėlimo komplekso (ne)buvimo genome dėsningumus.

Procesyvumo veiksniai nerasti keliose didelio genomo virusų šeimose. Pastarųjų DNR polimerazėms yra būdingi papildomi necharakterizuoti domenai arba nehomologiški regionai. Tikėtina, kad minėti priedai galbūt galėtų užtikrinti polimerazės procesyvumą arba šie virusai turi stipriai divergavusius arba netipiškus procesyvumo veiksnius, kurių nepavyko rasti. Pastaroji situacija gali būti phiKZ faguose. Jų DNR polimerazės yra stipriai divergavusios, tačiau išlaikiusios architektūrą, tipiską B-šeimai. Be to, dvi iš trijų polimerazių savo C-gale turi numanomą sąveikos su procesyvumo veiksmu motyvą. Šie faktai leidžia manyti, kad phiKZ fagai koduoja savus, stipriai divergavusius procesyvumo veiksnius, kurių dabartiniai metodai neranda.

Analizuojant procesyvumo veiksnius turinčius virusus, stebino tai, kad dalis genomų nekodavo žiedo užkėlėjo. Šį reiškinį padėjo paaiškinti procesyvumo veiksmių elektrostatiinių savybių analizė. Ji atskleidė, kad irido, asco, asfar ir marselio virusų bei 0305phi8-36 fago minėtiems baltymams yra būdingas padidėjęs pI (3.8 pav., A). Reprerentatyvių PV modeliai parodė, kad išaugęs teigiamas krūvis yra lokalizuotas sąveikos su DNR pusėje (3.8 pav., B). Ši savybė yra būdinga gerai charakterizuotiems herpes virusų procesyvumo veiksmiams. Pastarieji neformuoja žiedų, o su DNR sąveikauja kaip monomerai (UL42) arba dimerai (UL44). Tikėtina, kad minėtu būdu veikia ir šiame darbe tirti PV, turintys padidėjusį pI ir nekoduojantys žiedo užkėlėjų. Įdomu tai, kad panašias savybes turi du (Hus1 ir Rad9) iš trijų žmogaus 9-1-1 komplekso baltymų. Trečiojo baltymo (Rad1) elektrostatinės savybės yra panašesnės į ląstelinių PCNA. Šis pastebėjimas leidžia manyti, kad Hus1 ir galbūt Rad9, galėtų sąveikauti su DNR kaip monomerai arba formuoti heterodimerinius kompleksus. Įdomu, tai kad yra eksperimentinių duomenų apie skirtingą 9-1-1 komplekso komponentų vaidmenį. Nustatyta, kad *Schizosaccharomyces pombe* Rad1 mutantams buvo būdingi telomerų defektai, tuo tarpu Hus1 ir Rad9 mutantai turėjo normalias telomeras (Dahlen, Olsson et al., 1998). Be to, nustatyta, kad Rad1 atlieka telomerų priežiūros funkciją tik esant Hus1 arba

Rad9 (Khair, Chang et al., 2010). Tikėtina, kad minėtus skirtingus mutantų fenotipus galbūt lemia Rad9, Hus1 ir Rad1 elektrostatinės savybės.

Žiedo užkėlimo komplekso baltymų tyrimas kelia daugiausia neaiškumų. Tik trys eukariotų virusai koduoja visus RFC komplekso baltymus. Tikėtina, kad virusų RFC yra funkcionalus, nes jo subvienetai (RFC1-4) turi P-kilpos ir DEXX motyvus, o RFC(1/3/5) subvienetuose nustatyti sąveikos su PCNA (PIP-box) motyvai. Įdomu tai, kad pastarieji skiriasi nuo žmogaus ir mielių RFC PIP-box'ų (3.9 pav.). Panašu, kad sąveikos su PCNA „stiprumas“ RFC(1/3/5) subvienetuose gali evoliucionuoti skirtingai. Šią idėją palaiko tai, kad, priešingai nei mielių ir žmogaus atvejais, kai kuriuose eukariotuose stebimas tipiškas PIP-box'as RFC5 subvienete, o RFC1 – jis redukuotas (3.10 pav.). Keli *Phycodnaviridae* šeimos nariai turi tik RFC didžiojo subvieneto homologą. Mielių ir žmogaus RFC tyrimai rodo, kad minėtas subvienetas lemia komplekso specifiškumą žiedui (Indiani ir O'Donnell, 2006). Pavyzdžiui, RFC1 sąlygoja specifiškumą PCNA, o Rad17 – 9-1-1 kompleksui. Taigi, tikėtina, kad viruso RFC didysis subvienetas jungiasi su šeimininko mažaisiais subvienetais, sudarydamas viruso PCNA specifišką kompleksą. Faguose Ma-LMM01 ir RSL1 nustatyti bakterinio žiedo užkėlėjo homologai – polIII γ . Šiuose faguose taip pat randamos A-šeimoms DNR polimerazės, bei polIII β procesyvumo veiksnys. Įdomu, kokia šių fagų funkcionalios replikazės sudėtis? Galbūt polIII β užkeliamas viruso polIII γ , kuris papildomai prisijungia šeimininko žiedo užkėlėjo baltymus? Kitą vertus, polIII γ galėtų sudaryti funkcionalų kompleksą jungdamasis su tam tikru viruso baltymu, kaip tai vyksta T4 fage. Keturi šio viruso polIII γ homologai (gp44) jungiasi su vienu T4-like fagams specifiniu baltymu (gp62), suformuodami funkcionalų žiedo užkėlėją. Į šiuos klausimus tiksliausiai gali atsakyti tik eksperimentiniai tyrimai.

Taigi, nustatytas ryšys tarp viruso genomo dydžio ir replikazės komponentų gali turėti praktinę naudą. Pavyzdžiui, sekvenuojant genomą, pagal jo dydį galima prognozuoti replikazės tipą ir pilnumą. Be to, šis pastebėtas dėsningumas galbūt yra bendresnis bei tinka ląsteliniams

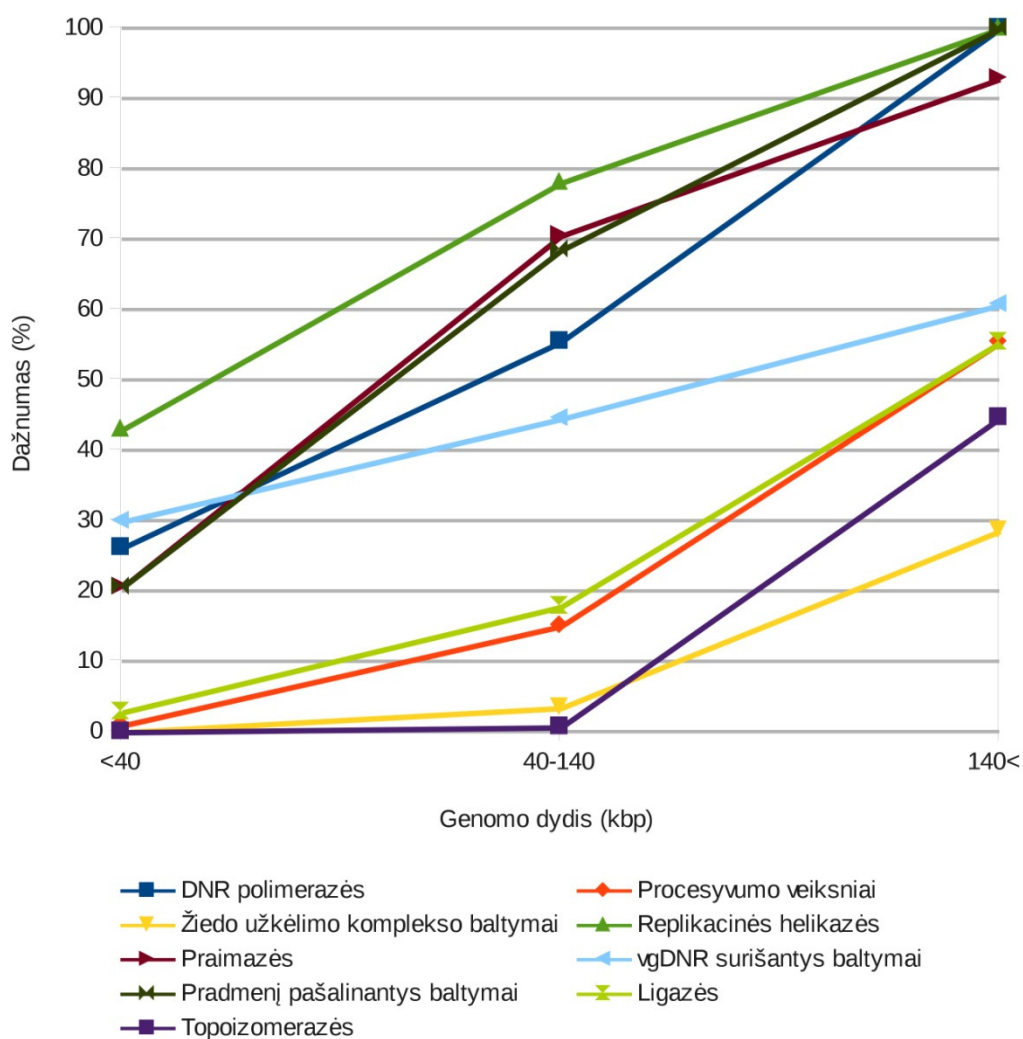
organizmams. Pavyzdžiui, simbiotinės bakterijos, priklausančios *Hodgkinia* ir *Carsonella* gentims, turi vienus iš mažiausių genomų (atitinkamai 144 ir 160 kbp) (McCutcheon, 2010). Pastarosiose bakterijose nerastas procesyvumo veiksnys ir jo užkėlėjas, o didesni ląsteliniai organizmai (*Sulcia cicada* (277 kb), *Buchnera Cc* (416 kb) ir *Nanoarchaeum equitans* (491 kb)) jau turi pilną replikazę. Atsirandant vis daugiau didelių virusų bei mažų ląstelių genomų bus įdomu stebėti, kiek universalus yra nustatytas genomo dydžio ir replikacijos komponentų ryšys.

3.2. DNR replikacijos baltymų ir viruso genomo dydžio bei šeimininko taksonomijos ryšys

Replikazių tyrimo metu nustatyta jos komponentų ir genomo dydžio sąsaja. Pagal replikazių sandarą ir replikacinių DNR polimerazių tipus virusus galima suskirstyti į tris genomo dydžio grupes (<40 kbp, 40–140 kbp ir >140 kbp). Ištyrus virusų replikazes, kilo idėja išanalizuoti kitus virusų replikacijos baltymus ir patikrinti, ar jiems taip pat galioja minėta priklausomybė nuo genomo dydžio. Taigi, atnaujinus virusų genomų ir baltymų duomenų bazes bei naudojant jautriausius sekų paieškos metodus identifikuoti dgDNR virusų replikacijos baltymai: DNR polimerazės ir jų pagalbiniai baltymai (žiedai ir jų užkėlėjai), viengrandę DNR surišantys baltymai, DNR praimazės, helikazės, pradmens pašalinime dalyvaujančios nukleazės, DNR ligazės bei topoizomerazės. Gauti rezultatai išanalizuoti ir tirta replikacijos baltymų buvimo bei jų tipų priklausomybė nuo genomo dydžio ir šeimininko taksonomijos.

Nustatyta, kad net tik replikazių, bet ir kitų replikacijos baltymų buvimo genome dažnumas priklauso nuo jo dydžio: replikacinius baltymus didelių genomų virusai turi dažniau nei maži (3.11 pav.). Didėjant genomui ryškiausiai išauga nuosavų topoizomerazių, DNR polimerazių, procesyvumo veiksmų ir jų užkėlėjų, pradmenį pašalinančių baltymų bei ligazių poreikis (3.11 pav.). Įdomu tai, kad lyginant 40–140 kbp ir >140 kbp grupes, tikimybė

pastaruosiuose rasti SSB baltymus padidėja mažiau nei kitų replikacinių baltymų atveju (3.11 pav.). Panašu, kad dideli virusai turi SSB baltymus, tačiau jie yra labai divergavę arba SSB funkciją atlieka kiti baltymai. Šią hipotezę palaiko tyrimo metu naujai surasti SSB baltymai didžiausiuose eukariotų virusuose (NCLDV) bei nustatytas raupų viruso SSB (I3) panašumas į bakterijų SmpB baltymą.

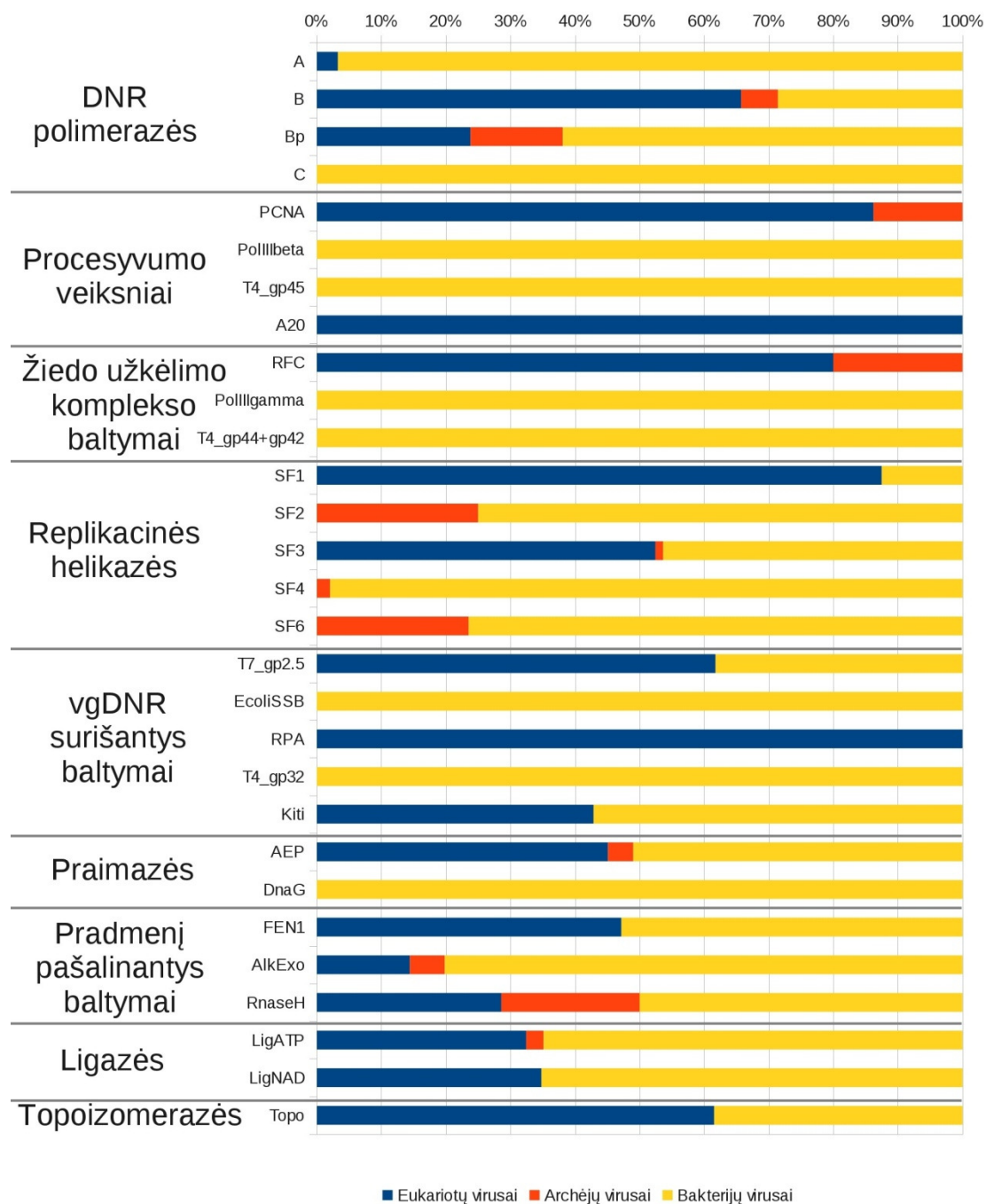


3.11 pav. DgDNR virusų genomo dydis ir replikacijos baltymų dažnumas.

Analizės metu paaiškėjo, kad dažniausias virusų replikacinis baltymas yra helikazė (3.11 pav.). Ją turėjo net 70% visų atstovų. Faguose ir archėjų virusuose naujai rasti archėjų-eukariotų replikacinės helikazės (MCM)

homologai bei nustatyta, kad herpes virusų helikazės-praimazės UL8 subvienetas yra inaktyvuota B-šeimoms DNR polimerazė.

Žinoma, kad bakterijų ir archėjų-eukariotų DNR replikacija skiriasi. Manoma, kad jų replikacijos baltymai evoliucionavo nepriklausomai (Leipe, Aravind et al., 1999). Buvo nuspręsta patikrinti ar virusų replikacinių baltymų (ne)buvimas ir jų tipai priklauso nuo šeimininko priskyrimo tam tikram gyvybės domeniui. Tiriamojoje imtyje bakterijų, archėjų ir eukariotų virusų dalys sudarė atitinkamai 71, 9 ir 20 procentų. Tik faguose buvo rasti bakterijų replikacijai specifiniai baltymai: A ir C šeimos polimerazės, PolIII β subvienetas ir jo užkėlimo kompleksas, DnaB helikazė (SF4), DnaG praimazė, SSB baltymas (3.12 pav.). Eukariotams ir archėjoms specifiniai replikacijos baltymai (PCNA ir RFC) buvo rasti tik šių organizmų virusuose. Archėjų virusuose santykinai dažnai buvo randamos SF6 ir SF2 šeimų helikazės (jas turi atitinkamai 24% ir 25% visų atstovų). Kiti replikaciniai baltymai yra būdingi virusams parazituojančioms visų trijų gyvybės domenų atstovus. Pavyzdžiui, SF3 helikazės, AEP šeimos praimazės ir B-šeimoms polimerazės yra randamos bakterijų ir archėjų-eukariotų virusuose. Ląsteliniai organizmai neturi SF3 superšeimos replikacinių helikazių. Tad tikėtina, kad minėti trys replikacijos baltymai nebuvo perimti iš šeimininko, o juos jau turėjo virusai, parazitavę paskutinįjį bendrą ląstelinių organizmų protėvį (LUCA – Last Universal Common Ancestor). Taigi, kaip ir ląstelinių organizmų atveju, bakterijų ir archėjų-eukariotų virusų replikacijos aparatai turi skirtumų, kurie tikriausiai atsirado dėl neortologinio baltymų pakeitimo adaptuojantis prie šeimininko. Tačiau, šie skirtumai nėra reikšmingi ir neužgožia ryškios genomo dydžio ir replikacijos baltymų priklausomybės.



3.12 pav. Replikacijos baltymų pasiskirstymas pagal viruso šeimininką.

3.3. Replikacinių helikazių ir praimazių analizė




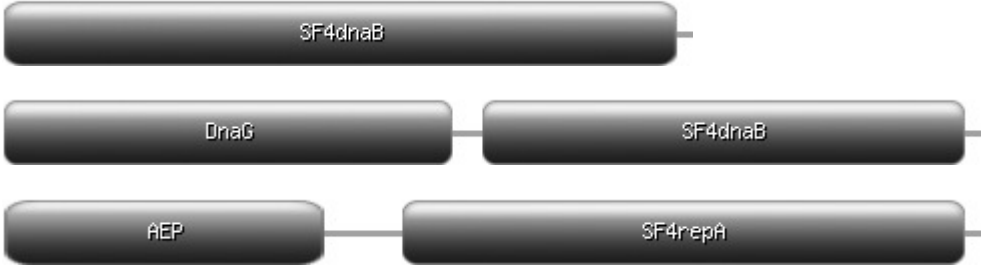

Tyrimo metu identifikuotos dgDNR virusų replikacinės helikazės, jos priskirtos helikazių superšeimoms ir nustatyta jų domeninė organizacija (Lentelė 3.2). Paaikškėjo, kad jos dažnai turi praimazės domeną. Kiekvienai superšeimai specifiniai bruožai aptarti tolimesniame tekste. Detaliau išanalizuota herpes virusų helikazė-praimazė.

3.3.1. Replikacinių helikazių ir praimazių identifikavimas, įvairovė ir domeninė organizacija

NTP surišantys baltymai yra funkciškai įvairūs (Snider ir Houry, 2008). Replikacinės helikazės sudaro tik nedidelę šių baltymų dalį (McGlynn, 2013). Dažnai virusų genomuose buvo rasti keli NTP surišantys baltymai. Replikacinės helikazės atrinktos ir priskirtos superšeimoms pagal tam tikrus kriterijus (žr. metodai, replikacinių baltymų identifikavimas). Svarbiausias jų – baltymo panašumas į jau charakterizuotas replikacines helikazes. Kitas reikšmingas požymis – kartu su NTP surišimo domenu buvo koduojama praimazė. Šį domeną turėjo 41% visų replikacinių helikazių (Lentelė 3.2). Ląsteliniam organizmams būdingos dviejų tipų praimazės: DnaG ir AEP. Bakterijų praimazės (DnaG) homologai rasti tik bakteriofaguose, o archėjoms ir eukariotams būdinga AEP nustatyta visų trijų gyvybės domenų virusuose. AEP buvo aptikta dažniau (58% visų atvejų) nei DnaG.

Nustatyta, kad dgDNR virusai dažniausiai turi SF4 šeimos helikazes (bakterijų DnaB homologus) (Lentelė 3.2). To buvo tikėtasi, nes didžiąją dalį dgDNR virusų (71%) sudaro bakteriofagai. SF3 šeimos atstovų buvo rasta mažiau (38% visų atvejų). Šiuos baltymus turi labai įvairūs virusai. Jų šeiminkai apima visus tris gyvybės domenus. Rečiausi yra SF1, SF2 ir SF6 superšeimų baltymai (Lentelė 3.2). Kiekviena replikacinių helikazių grupė detaliau aptariama tolimesniuose skyreliuose.

Lentelė 3.2 DgDNR virusų replikacinių helikazių pasiskirstymas pagal superšeimas ir dažniausia domeninė organizacija.

Superšeima*	Dažniausiai grupėje rasta domeninė organizacija
SF1 (7%)	
SF2 (2%)	
SF3 (38%)	
SF4 (45%)	
SF6 (8%)	

* – skliaustuose pateikta kokią dalį visų rastų replikacinių helikazių sudaro konkreti superšeima.

SF1 superšeima. Tyrimas rodo, kad dgDNR virusai retai (7% visų atvejų) naudoja SF1 šeimos helikazes DNR replikacijai (Lentelė 3.2). Priešingai nei kitose superšeimose, SF1 atstovai yra nesujungti su praimazės domenu. Dauguma rastų SF1 atstovų priklauso herpes virusams. Šių virusų helikazė-praimazė yra heterotrimerinis kompleksas, sudarytas iš UL5 (helikazės), UL52 (archėjų-eukariotų praimazė) ir UL8 (Muylaert, Tang et al., 2011). UL8 baltymo struktūra ir domeninė organizacija nebuvo žinoma (Muylaert, Zhao et al., 2012). Šio tyrimo metu, naudojant jautriausius sekų paieškos metodus,

buvo nustatytas reikšmingas UL8 panašumas į B šeimos DNR polimerazes, sudarytas ir išanalizuotas UL8 struktūros modelis (žr. skyrių: „Herpes virusų helikazės-praimazės UL8 subvienetas yra inaktyvuota B šeimos polimerazė“).

SF2 superšeima. SF2 superšeimai priklausė mažiausiai helikazių (2%), kurios, tikėtina, dalyvauja replikacijoje. Fagai N15 ir PY54 turi helikazes (atitinkamai gi: 9630494 ir 33770544) kurių N-galas yra sujungtas su DnaG praimazės domenais, o PAU fago helikazė (gi: 435844571) turi AEP praimazės domeną. *Acidianus filamentous virus 6* helikazė (gi: 162329366) neturi praimazės domeno. Ji yra gimininga archėjų-eukariotų HEL308 helikazėms. Tyrimai rodo, kad šie baltymai išvynioja motinines grandines ir pritraukia replikacijos baltymus, taip padėdami tęsti DNR replikaciją esant sustojusiai (stalled) replikacinei šakutei (Tafel, Wu et al., 2011). Todėl tikėtina, kad *Acidianus filamentous virus 6* HEL308 homologas dalyvauja viruso DNR replikacijoje, pritraukdamas šeimininko replikacijos baltymus, nes šiame viruse kitų replikacijos baltymų nerasta (išskyrus PD-(D/E)XK šeimos nukleazę, kuri galėtų dalyvauti pradmens pašalinime).

SF3 superšeima. Šiai superšeimai priklauso gerai žinomos poliomos ir papilomos virusų bei mažiau tyrinėtos NCLDV replikacinės helikazės (pvz. raupų virusų D5) (Singleton, Dillingham et al., 2007). SF3 atstovai naujai rasti didelio genomo eukariotų viruse SWSSV. Nustatyta, kad grupė fagų baltymų yra reikšmingai panašūs į SF3 helikazes bei į Pfam profilį DUF927. Pastarajam būdingi konservatyvūs helikazių motyvai ir reikšmingas panašumas į SF3 atstovų profilius. Pfam duomenų bazės duomenimis, DUF927 domenas dažnai yra randamas kartu su Toprim arba AEP praimazės domenais. Todėl tikėtina, kad šiuos domenus turintys virusų baltymai yra replikacinės helikazės. Dauguma rastų SF3 atstovų C-gale turi AEP praimazės domeną, o DnaG praimazės domenas rastas daug rečiau. Įdomu tai, kad *Phaeocystis globosa* viruso ir Sputnik virofaguose prie SF3 helikazės domeno C-galo yra prijungtas A šeimos DNR polimerazės domenas (gi atitinkamai: 509141013 ir 195982544).

SF4 superšeima. Daugiausia replikacinių helikazių (45%) priklauso SF4 superšeimai. Jos dažniausiai yra panašios į *E. coli* DnaB ir būdingos bakteriofagams. Tik dviejuose archėjų virusuose (HCTV-1 ir HVTV-1) buvo rasti SF4 atstovai. Pastarieji baltymai rodė panašumą į ląstelių organizmų DNR rekombinacijoje dalyvaujančius Rad51/RecA/radA baltymus. Minėti archėjų virusai koduoja daug nuosavų replikacinių baltymų, tačiau tipišku replikacinių helikazių juose nebuvo rasta, todėl, tikėtina, kad pastarosios helikazės galbūt galėtų dalyvauti šių virusų replikacijoje. Eukariotų virusuose SF4 replikacinių helikazių nerasta. Dauguma SF4 atstovų neturi praimazės domeno (pvz., T4 fago gp41) (Lentelė 3.2). Kitos virusų replikacinės helikazės savo N-gale turi DnaG (pvz., T7 fago gp4) arba AEP praimazės domenų. Pastarajai grupei priklauso BFK20 fago gp43 baltymas, kurio N-galui nustatyti praimazės ir polimerazės aktyvumai (Halgasova, Mesarosova et al., 2012).

SF6 superšeima. SF6 superšeimai priklauso MCM baltymai – eukariotų ir archėjų replikacinės helikazės. Iki šiol šios grupės atstovai buvo žinomi tik archėjų BJ1 viruse ir *Thermococcus prieurii virus 1* (Krupovic, Gribaldo et al., 2010; Gorlas, Koonin et al., 2012). Šios analizės metu, naudojant jautrius profilių lyginimo metodus pirmą kartą buvo rasti MCM homologai faguose. Pavyzdžiui, *Mycobacterium phage Corndog* baltymui (gi: 29566306) HHpred rado panašumą į MCM profilį (PF00493) su 98% tikimybe. Fagų MCM homologai priklauso Pfam profiliui (DUF3987), kurio atstovų funkcija nėra žinoma. Tikėtina, kad šie baltymai dalyvauja plazmidžių/fagų replikacijoje, nes jiems, kaip ir DUF927 atveju, būdingi praimazės domenai. Faguose DUF3987 domenas randamas vienas arba kartu su AEP/DnaG praimazės domenais (Lentelė 3.2).

3.3.2. Replikacinės helikazės – dažniausias dgDNR virusų replikacijos baltymas

Intuityviai buvo tikėtasi, kad dažniausias replikacijos baltymas bus DNR polimerazė, nes būtent ji atlieka DNR sintezę. Identifikavus replikacines

helikazes, netikėtai paaiškėjo, kad jos yra dažniausiai randamas virusų replikacijos baltymas (3.11 pav.). Replikacinę helikazę turi net 70% visų dgDNR virusų (3.11 pav.). Kokia galėtų būti šio reiškimo priežastis? Dažną replikacinės helikazės buvimą didesnių genomų (40 ir daugiau kbp) virusuose galima paaiškinti jau nustatytu dėsningumu, teigiančiu, kad didėjant genomui koduojama daugiau nuosavų replikacinių baltymų. Tačiau nuostabą kelia santykinai dažnos replikacinės helikazės mažų genomų virusuose (<40 kbp). Kilo idėja, kad galbūt mažų genomų virusai naudoja helikazes replikacijos iniciacijai ir šeimininko replikacijos baltymų pritraukimui. Pavyzdžiui, taip elgiasi papildomos ir poliomos virusai (Mansky, Batiza et al., 1997). Išanalizavus virusų proteomas ir literatūros duomenis, nustatyti virusų, neturinčių DNR helikazių, bruožai. Dalį šių atvejų sudaro virusai, naudojančys pradmeniu baltymą. Yra žinoma, kad jų replikacijai DNR helikazės nereikalingos (Salas, 1991). Kiti virusai, neturintys savo helikazės, dažnai koduoja helikazės užkėlėjus, kurie tikriausiai naudojami šeimininko helikazei užkelti. Likusieji turi integrazes ir(arba) baltymus, kurie gali būti panaudoti šeimininko replikacijos aparato pasitelkimui savo DNR replikacijai (PCNA/PolIII β , SSB). Žinoma, kad kai kurie archėjų virusai aktyvuoja šeimininko replikacinį aparatą, savais baltymais, pvz., PCNA arba su juo sąveikaujančiais (Gardner, Bell et al., 2014).

Neįtraukus pradmens-baltymo replikaciją naudojančių virusų ir pridėjus tuos, kurie turi helikazės užkėlėjus (DnaC_{Eco} DnaI_{Bsu}, Cdc6/Orc1, lambda fago P baltymo, SPP1 fago G39P baltymo, P2 fago B baltymo homologai), virusų dalis, kurie potencialiai gali naudoti savo arba šeimininko helikazes mažų genomų virusuose (<40 kbp) išauga nuo 43% iki 78%, o skaičiuojant visų geno dydžių virusus – nuo 70% iki 90%. Taigi, panašu, kad helikazių naudojimas geno replikacijai inicijuoti yra universalesnė strategija, būdinga ne tik mažiesiems eukariotų, bet ir kitiems virusams.

3.3.3. Herpes virusų helikazės-praimazės UL8 subvienetas yra inaktyvuota B šeimos polimerazė

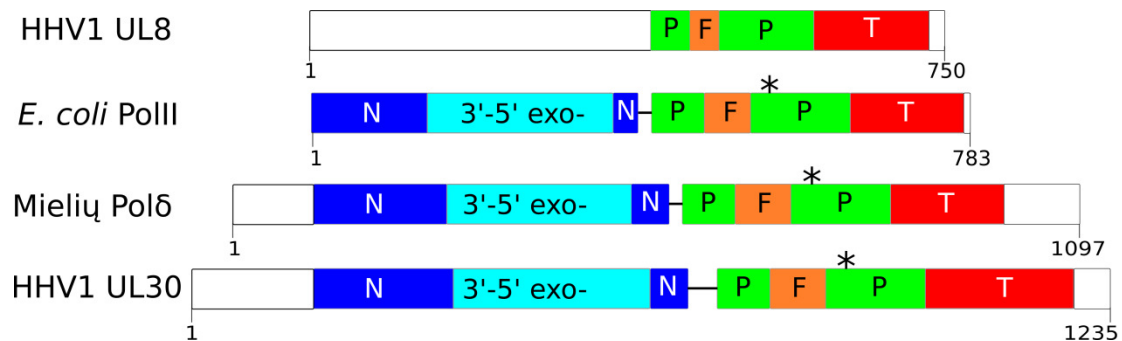
Herpes virusų replikacijos aparatas yra vienas geriausiai ištirtų virusuose (Weller ir Coen, 2012). Nepaisant to, helikazės-praimazės kompleksas vis dar išlieka mįslingiausias minėtos sistemos komponentas. Įdomu tai, kad herpesų helikazė-praimazė sudaryta iš trijų baltymų: UL5 (helikazės), UL52 (praimazės) ir UL8. Nepaisant intensyvių tyrinėjimų UL8 struktūra ir domeninė organizacija nebuvo žinoma (Muylaert, Zhao et al., 2012). Netikėtai, šiame darbe, nustatyta, kad UL8 yra inaktyvuota B-šeimos DNR polimerazė. Iteratyvios paieškos jackhammer programa davė pirmąją užuominą, kad UL8 galėtų būti panašus į B šeimos DNR polimerazes. Pavyzdžiui, atliekant paiešką su BoHV-2 UL8 (gi: 14161473), po keturių iteracijų randama *Methanococcus aeolicus* B šeimos polimerazė (gi: 150401083) su statistiškai patikimu įverčiu (E-value=0.001). Šio rezultato patikimumas buvo patikrintas naudojant dar jautresnius metodus, paremtus profilių lyginimu. HHpred programa nustatė lūpų pūslelinės viruso (HHV-1) UL8 ir archėjos *Thermococcus gorgonarius* PolB (SCOP: d1tgoa2) panašumą su 95% tikimybe. Kiti jautrūs metodai (žr. metodai, replikacijos baltymų paieška) taip pat rodė UL8 ir B šeimos polimerazių panašumą. Nustatytas homologiškumas apėmė UL8 C-galinę dalį (~393-727 a-r. HHV-1 atveju). Ryšys tarp minėtų baltymų yra tolimas, nes identiškumas tarp išlygintų UL8 ir polimerazės regionų yra tik 9-12%. UL8 N-galas nerodė panašumo į jau charakterizuotus baltymus. Tačiau tai nebūtinai reiškia, kad jam būdinga nauja struktūrinė sankloda. Galbūt, UL8 N-galas yra stipriai divergavęs egzonukleazės domenas, nes B-šeimos polimerazės dažniausiai jį turi (Rothwell ir Waksman, 2005).

HHV-1 UL8 C-galo (UL8c) detalesnei analizei buvo sudarytas jo struktūros modelis (prieinamas adresu: http://www.ibt.lt/bioinformatics/models/hhv1_ul8/). Šablonu naudota *E. coli* PolIII (PDB: 3k57) (Wang ir Yang, 2009). Esant mažam taikinio ir šablono identiškumui svarbu objektyviai įvertinti modelio kokybę. Šiam tikslui pasiekti buvo atliktas kontrolinis eksperimentas: naudojant *E. coli* PolIII struktūrą kaip

šabloną, sumodeliuotos DNR polimerazės su jau žinomomis struktūromis, o modelių kokybės įverčiai palyginti su UL8c. Norint sukurti situaciją kuo panašesnę į PolIII ir UL8c atvejį, buvo pasirinktos DNR polimerazės, tolimai giminingos PolIII (identiškumas <30%) (Lentelė 3.3). Išlyginus PolIII ir RB69 fago, mielių Polδ, *P. furiosus* PolB bei HHV-1 DNR polimerazės (UL30) struktūras buvo gauti „idealūs“ taikinio-šablono palyginiai, pagal kuriuos sudaryti modeliai. Pastarieji įvertinti naudojant ProSA-web (mažesnis Z-įvertis rodo energetiškai palankesnę struktūrą) (Lentelė 3.3). HHV-1 UL8 C-galo ir jo homologų iš *Simplex* virusų genties įverčiai buvo blogesni nei naudoto šablono (Lentelė 3.3, HHV-1 UL8c, -7.48; PolIII, -8.86). Tačiau HHV-1 UL8c įvertis buvo geresnis už visų B šeimos polimerazių, išskyrus *P. furiosus* PolB (Lentelė 3.3). Šie rezultatai rodo, kad UL8c ir B šeimos polimerazių struktūros yra panašios, o UL8c modelis tikriausiai neturi didelių trūkumų.

Lentelė 3.3 UL8c ir B šeimos DNR polimerazių struktūrų modelių kokybės įverčiai. UL8c iš Simplex virusų genties pavaizduoti pilkam fone.

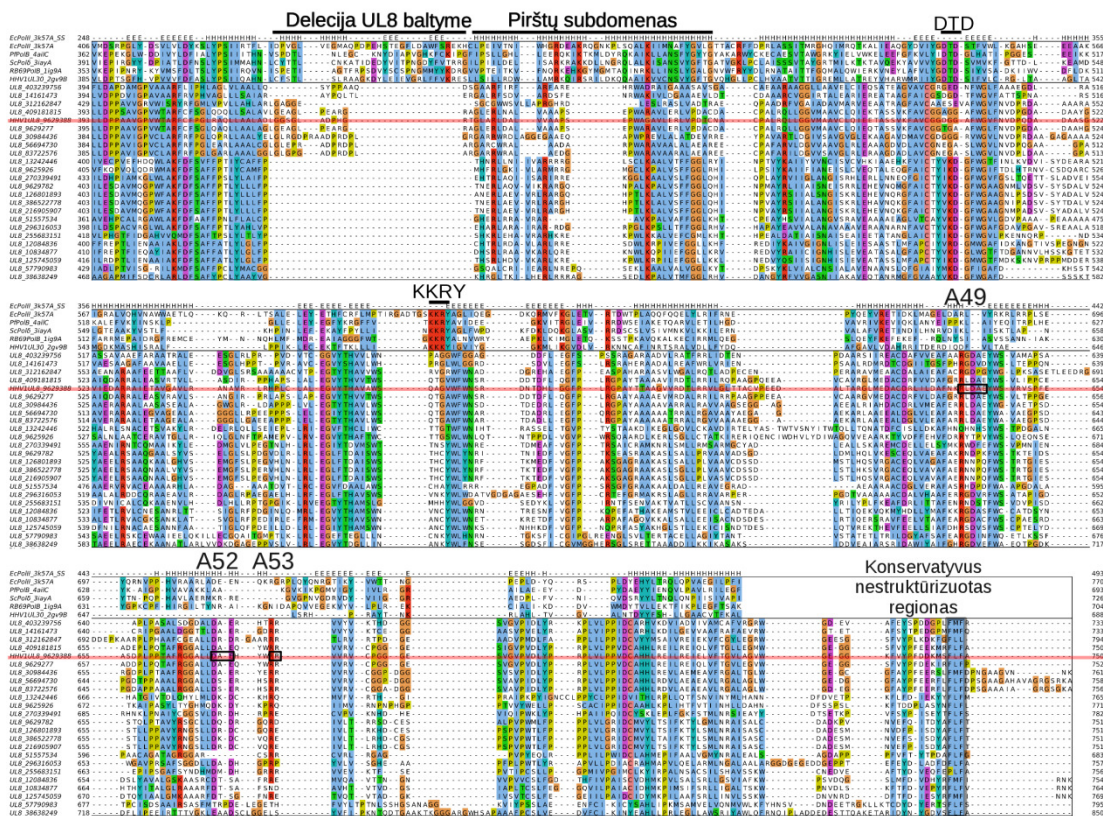
Modelis	Modelio sekos GI numeris	Šablonas (PDB kodas ir grandinė)	Modelio ProSA-web Z-įvertis	Šablono ProSA-web Z-įvertis	Taikinio-šablono sekų identiškumas (%)
UL8c_HHV1	9629388	3k57_A	-7.48	-8.86	12
RB69PolB_lig9A	14719800	3k57_A	-6.01	-8.86	20
ScPolδ_3iayA	259090299	3k57_A	-6.9	-8.86	25
PfPolB_4ailC	379318587	3k57_A	-8.92	-8.86	26
HHV1UL30_2gv9B	109158093	3k57_A	-4.98	-8.86	19
403239756	403239756	3k57_A	-6.12	-8.86	9
14161473	14161473	3k57_A	-7.11	-8.86	11
312162847	312162847	3k57_A	-6.81	-8.86	9
409181815	409181815	3k57_A	-7.8	-8.86	12
9629277	9629277	3k57_A	-8.31	-8.86	12
30984436	30984436	3k57_A	-7.51	-8.86	11
56694730	56694730	3k57_A	-6.89	-8.86	10
83722576	83722576	3k57_A	-7.16	-8.86	10



3.13 pav. HHV-1 UL8 ir B šeimos DNR polimerazių domeninė organizacija. N, N-galas; P, delnas (palm); F, pirštai (fingers); T, nykštys (thumb). Žvaigždutė žymi polimerazės aktyvų centrą („DTD“ motyvas).

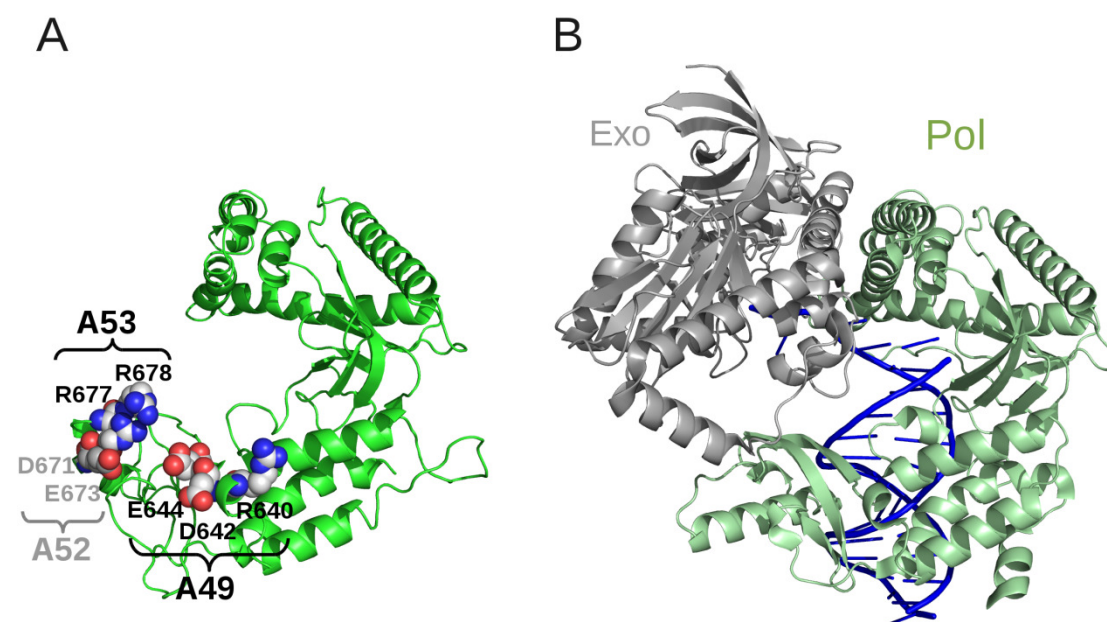
UL8c ir B šeimos polimerazių skirtumai nustatyti atliekant detalią UL8c sekos ir struktūros motyvų analizę. UL8c neturi polimerazėms būdingo aktyviojo centro („DTD“) (3.13 ir 3.14 pav.). Tik antrasis aspartatas buvo aptiktas keliuose α herpes virusuose (3.14 pav.). Pirštų subdomenas, svarbus įeinančio nukleotido atpažinimui ir surišimui (Rothwell ir Waksman, 2005), UL8c atveju yra redukuotas. UL8c regionui, esančiam prieš pirštų subdomeną, būdinga delecija (3.14 pav.). UL8c taip pat neturi „KKRY“ motyvo, kuris

reikalingas B-formos DNR stabilizavimui (Franklin, Wang et al., 2001). Taigi, minėti požymiai ir eksperimentiniai UL8 tyrimai (Dodson ir Lehman, 1991) rodo, kad šis baltymas neturi katalizinio aktyvumo. UL8c sąveikos su DNR motyvų pakitimai turėtų susilpninti šią sąveiką. Tyrimai rodo, kad UL8 vienas nesijungia su viengrande ar dvigrande DNR (Parry, Stow et al., 1993). Tačiau žinoma, kad UL8 helikazės komplekse moduliuoja jo sąveiką su DNR (Chen, Bai et al., 2011). Apibendrinant, sekų ir struktūrų analizė rodo, kad UL8c yra B-šėimos polimerazė, praradusi katalizinį aktyvumą ir turinti redukuotus sąveikos su DNR motyvus.



3.14 pav. UL8c ir DNR polimerazių sekų palyginys. HHV-1 UL8c seka pažymėta raudona linija. Juoda linija rodo sekų motyvų ilgį. HHV-1 UL8 mutuoti regionai yra juoduose stačiakampiuose, mutantų pavadinimai pateikti virš palyginio. Pirmoje palyginio eilutėje pavaizduota *E. coli* PolII (PDB: 3k57) antrinė struktūra, nustatyta DSSP programa (Joosten, te Beek et al., 2011); E - β-juosta, H-spiralė. Santrumpos: Ec, *Escherichia coli*; Pf, *Pyrococcus furiosus*; Sc, *Saccharomyces cerevisiae*; RB69, *Enterobacteria phage RB69*.

UL8 sąveikos su baltymais vietos. UL8c struktūros modelis leido pažvelgti į žinomas ir tikėtinas sąveikos su baltymais vietas. Neseniai HHV-1 UL8 baltymui buvo atlikta krūvių pakeitimo į alaninus mutagenezė ir nustatyti keturi mutantai su replikacijos defektais (Muylaert, Zhao et al., 2012). Trys iš jų patenka į UL8c modelio ribas (3.14 ir 3.15 pav.). Mutantai A49 (R640A, D642A ir E644A) ir A53 (R677A ir R678A) nesąveikavo su praimaze (UL52), o A52 (D671A ir E673A) mutanto atveju, sąveika buvo silpnesnė. A52 mututos liekanos UL8c modelyje atsiduria greta A53 mutacijų. UL8c modelis dera su eksperimentiniais duomenimis, nes visų minėtų mutantų a-r. formuoja vieną sąveikos paviršių.

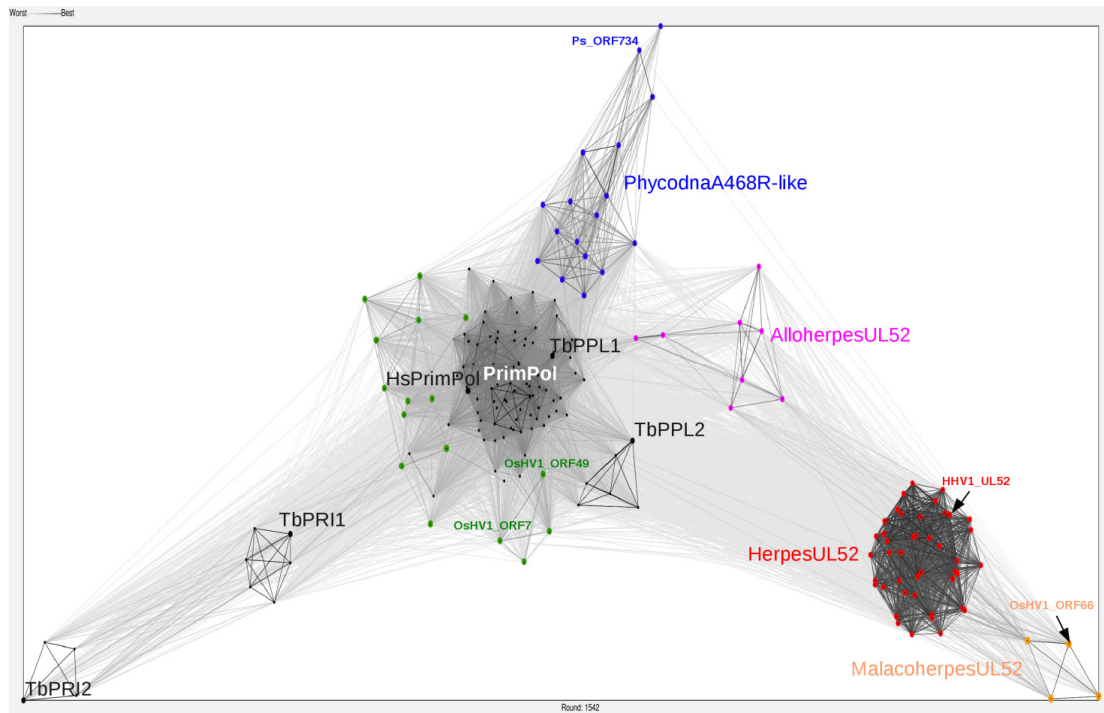


3.15 pav. HHV-1 UL8c modelis palygintas su *E. coli* PolIII. (A) UL8c ir jo a-r., kurių mutacijos turi įtakos sąveikai su UL52. Mutuotos liekanos pavaizduotos erdvės užpildymo reprezentacija (anglies atomai, pilki/balti; azoto, mėlyni; deguonies, raudoni). (B) *E. coli* PolIII ir DNR kompleksas. Polimerazės (Pol) ir egzonukleazės (Exo) domenai nuspalvinti atitinkamai žaliai ir pilkai.

Tikėtina, kad UL8 C-galas yra dar viena sąveikos su baltymais vieta (3.14 pav.). Šiam regionui būdingas trumpas hidrofobinis konservatyvus motyvas (HHV-1 UL8 747-FLF-749). Programos PrDOS (Ishida ir Kinoshita, 2007) ir MetaDisorder3D (Kozłowski ir Bujnicki, 2012) prognozuoja, kad jis yra nestruktūrizuotame regione. Žinoma, kad minėti požymiai yra būdingi sąveikos su baltymais vietoms. Eksperimentiškai nustatyta, kad HHV-1 UL8 C-galas

jungiasi prie DNR polimerazės (Marsden, McLean et al., 1997). Pastarasis tyrimas identifikavo su DNR polimeraze sąveikaujančias a-r. šalia konservatyvaus C-galo motyvo. Tai leidžia manyti, kad konservatyvus hidrofobinis motyvas, esantis nestruktūrizuotame regione, galbūt yra antrinė sąveikos su DNR polimeraze vieta arba dalyvauja sąveikoje su kitais replikacijos baltymais.

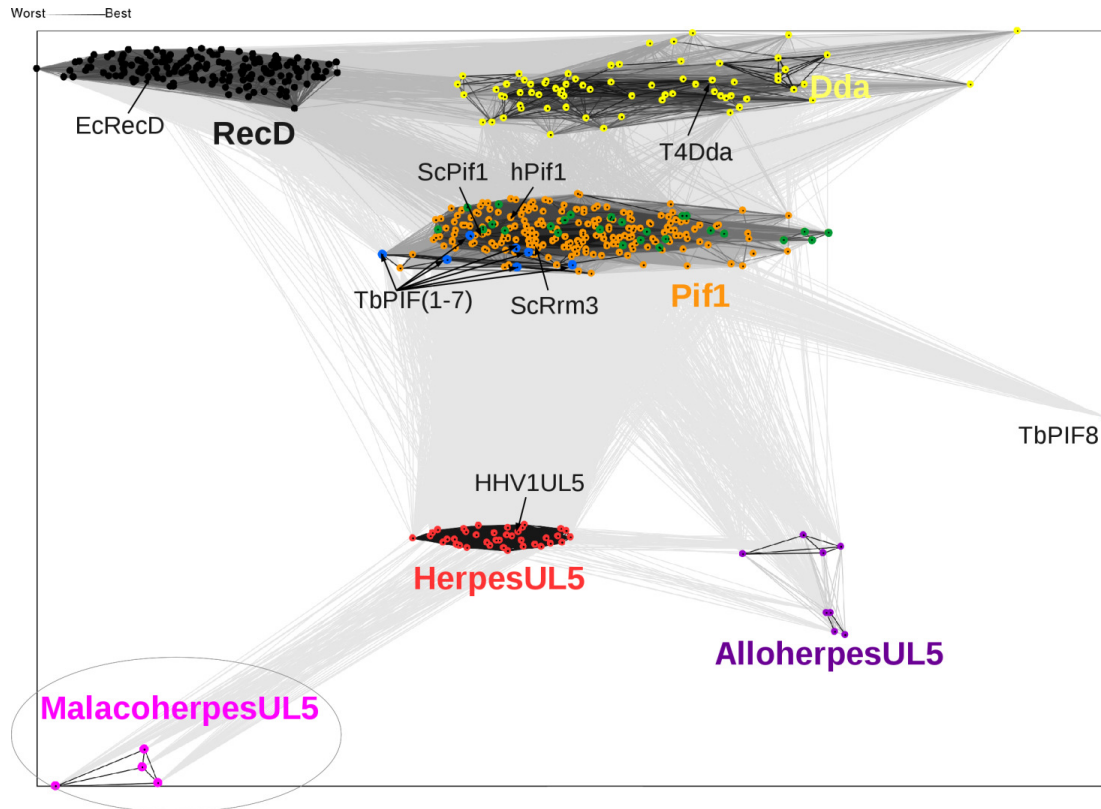
UL52 ir UL5 yra giminingi, atitinkamai, eukariotų PrimPol ir Pif1 helikazei. UL8 kartu su UL52 ir UL5 sudaro helikazės-praimazės kompleksą. Tikintis geriau pažinti UL8 buvo ištirtas jo partnerių UL52 ir UL5 homologiškumas kitiems baltymams. HHV-1 UL52 (gi: 9629434) yra 1058 a-r. ilgio baltymas, kurio C-galinė dalis gimininga archėjų-eukariotų praimazei (AEP) (Iyer, Koonin et al., 2005). UL52 homologai, rasti sekų paieškos metu, sudaro keletą grupių (3.16 pav.). Kartu grupuojasi HHV-1 ir kitos sekos iš *Herpesviridae* šeimos. Pastarosioms giminingos moliuskų (*Malacoherpesviridae*) bei šaltakraujų gyvūnų (*Alloherpesviridae*) herpesvirusų UL52. UL52 homologų buvo rasta NCLD virusuose ir eukariotuose. Pastarajai grupei priklauso neseniai charakterizuota žmogaus praimazė-polimerazė (PrimPol (CCDC111)), kuri svarbi DNR replikacijai, esant jos pažaidoms (Garcia-Gomez, Reyes et al., 2013). Žmogaus PrimPol yra pirmas eukariotinio baltymo, turinčio ir praimazės, ir polimerazės aktyvumus pavyzdys. Įdomu tai, kad kinetoplastidai koduoja net keletą PrimPol kopijų. Pavyzdžiui, *Trypanosoma brucei* turi keturis PrimPol. Nustatyta, kad du iš jų (3.16 pav., TbPRI1 ir TbPRI2) turi praimazės aktyvumą ir yra būtinos ląstelės augimui ir kinetoplasto DNR replikacijai (Bezalel-Buch, Yaffe et al., 2013). Kiti du baltymai yra DNR polimerazės, o TbPPL1 dar turi ir praimazės aktyvumą (Rudd, Glover et al., 2013). PrimPol grupei taip pat priklauso virusų baltymai iš *Marseilleviridae*, *Mimiviridae*, *Asfarviridae* šeimų bei *Ostreid herpesvirus 1*. Pastarasis *Malacoherpesviridae* atstovas turi net tris UL52 homologus. Du iš jų (OsHV1_ORF7 ir OsHV1_ORF49) grupuojasi su PrimPol, o trečiasis (OsHV1_ORF66) yra panašesnis į herpesų UL52.



3.16 pav. UL52 ir jo homologai sugrupuoti pagal tarpusavio sekų panašumą. Linijomis sujungtos sekos, kurių CLANS panašumo įvertis (P-įvertis) $\leq 1e-03$. Žaliai pažymėtos sekos priklauso *Malacoherpesviridae* šeimai ir NCLD virusams. Santrumpos: Ps, *Pandoravirus salinus*; Tb, *Trypanosoma brucei*; Hs, *Homo sapiens*; HHV1, *Human herpesvirus 1*; OsHV1, *Ostreid herpesvirus 1*; Alloherpes, *Alloherpesviridae*; Herpes, *Herpesviridae*; Malacoherpes, *Malacoherpesviridae*.

UL5 priklauso SF1 helikazių superšeimai. Atliekant sekų paieškas su HHV-1 UL5 (gi: 9629385), pirmose iteracijose buvo rastos eukariotų ir virusų Pif1 helikazės. Tolimesni UL5 homologai (bakterijų RecD ir T4 Dda) buvo nustatyti vėlesnėse iteracijose. *Malacoherpesviridae* šeimoje UL5 homologai buvo rasti pirmą kartą (3.17 pav.). Grupuoiant UL5 homologus, kaip ir UL52 atveju, išsiskiria trys herpes virusų grupės. Joms artimai giminingi yra eukariotų ir virusų Pif1. Pif1 būdingas daugeliui eukariotų. Žmogus turi vieną Pif1, o *Saccharomyces cerevisiae* – du (ScPif1 ir ScRrm3). Įdomu, kad *T. brucei* koduoja net aštuonis Pif1 paralogus. Pif1 šeimoje geriausiai yra charakterizuoti mielių baltymai. ScPif1 dalyvauja telomerinės, ribosominės ir mitochondrinės DNR replikacijoje, Okazaki fragmentų brendime (Bochman, Sabouri et al., 2010). Neseniai buvo nustatyta, kad ScPif1, kartu su Polδ, dalyvauja su rekombinacija susijusioje (recombination-coupled) DNR sintezėje (Wilson, Kwon et al., 2013). ScRrm3, judėdama kartu su replikacine šakute,

padeda įveikti sunkiai replikuojamus genomo regionus (Azvolinsky, Dunaway et al., 2006). Nustatyta, kad ScRrm3 sąveikauja su Pole, todėl tikėtina, kad ScRrm3 gali būti stabilus replisomos komponentas (Azvolinsky, Dunaway et al., 2006).



3.17 pav. UL5 ir jo homologų grupės. Linijomis sujungtos sekos, kurių CLANS panašumo įvertis (P-įvertis) $\leq 1e-07$. Naujai rasti UL5 homologai *Malacoherpesviridae* šeimoje apibraukti elipse. Pif1 homologai iš NCLDV, bakulo virusų ir fagų pažymėti žaliai. Santrumpos: HHV1, *Human herpesvirus 1*; Sc, *Saccharomyces cerevisiae*; Ec, *Escherichia coli*; h, žmogaus (human); T4, T4 fagas; Tb, *Trypanosoma brucei*; Alloherpes, *Alloherpesviridae*; Herpes, *Herpesviridae*; Malacoherpes, *Malacoherpesviridae*.

Herpes virusų helikazės-praimazės kompleksas: sąsajos su eukariotų DNR replikacija. UL8 kompiuterinė analizė atskleidė, kad šis baltymas yra neaktyvi B-šeimės polimerazė. Tai paaiškina, kodėl UL8 nebuvo nustatytas katalizinis aktyvumas. Kokia galėtų būti UL8 kilmė? Galbūt ji atsirado įvykus herpesų DNR polimerazės (UL30) duplikacijai. Kita vertus, UL8 galėjo kilti iš B šeimos polimerazių protėvių. Deja, mažas sekų identiškumas neleidžia

tiksliai nustatyti UL8 baltymo kilmės. UL52 ir UL5 homologų analizė atskleidė jų panašumus į eukariotinius baltymus. Taigi, visi trys helikazės-praimazės komplekso komponentai turi struktūrinius ir galbūt funkcinius panašumus į atitinkamus eukariotų baltymus. Pavyzdžiui, UL8 gali būti lyginama su Polε polimerazės C-galu, kuris yra neaktyvus egzonukleazės-polimerazės modulis naudojamas kaip sąveikos su baltymais platforma (Tahirov, Makarova et al., 2009). Neseniai nustatyta, kad UL5 sąveikauja su DNR polimeraze (UL30) (Weller ir Coen, 2012), o Pif1 jungiasi su Polδ (Wilson, Kwon et al., 2013). UL52-UL5 formuoja subkompleksą. Kyla klausimas, ar šių baltymų atitikmenys eukariotuose taip pat galėtų sąveikauti esant tam tikroms sąlygoms? Tiesioginių sąveikų įrodymų nėra, bet užuominų tyrimuose galima rasti. Pavyzdžiui, PrimPol ir Pif1 baltymai padeda įveikti sunkiai replikuojamus genomo regionus (Azvolinsky, Dunaway et al., 2006; Garcia-Gomez, Reyes et al., 2013). Koreliuotas PrimPol ir Pif1 baltymų išplitimas *Trypanosoma brucei* genome yra dar viena užuomina, kad jie gali būti funkciškai ir galbūt fiziškai susiję. Žinoma, kad bent dvi tripanosomų PrimPol ir Pif1 homologų poros dalyvauja tame pačiame procese. TbPRI1 ir TbPIF2 funkcionuoja maksižiedų DNR replikacijoje, o TbPRI2 ir TbPIF1 dalyvauja minižiedų replikacijoje ir segregacijoje (Bezalel-Buch, Yaffe et al., 2013). Natūralu, kad yra skirtumų tarp herpes virusų ir eukariotų minėtų kompleksų. Vis dėlto jų panašumų analizė galėtų pagilinti žinias apie abi sistemas.

3.4. Viengrandę DNR surišantys baltymai

Atlikus preliminarią replikacijos baltymų pasiskirstymo dgDNR virusų genomuose analizę, nustebino SSB baltymų nebuvimas didžiausiuose eukariotų virusuose. Šie virusai buvo išanalizuoti detaliau. Tyrimo metu virusų vgDNR surišantys baltymai naujai rasti NCLDV grupėje bei nustatyta, kad raupų SSB (I3) yra giminingas bakterijų SmpB (Small protein B).

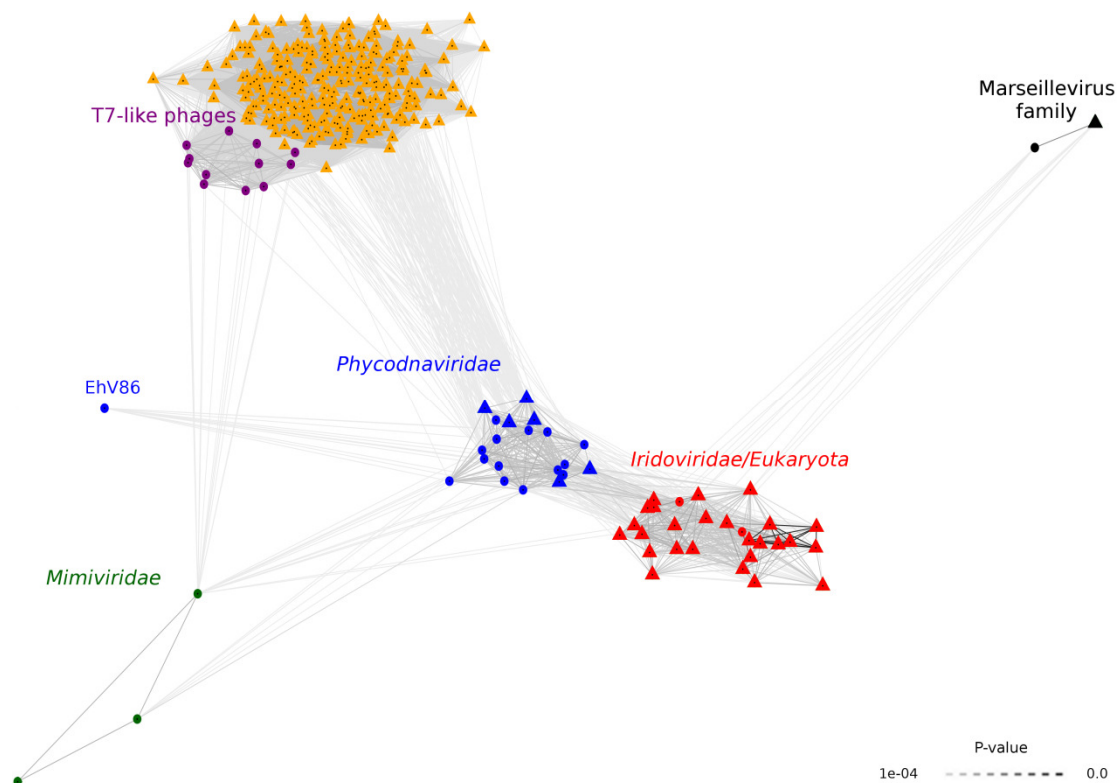
Apskritai, virusuose dažniausiai (43% atvejų) nustatyti *E. coli* SSB homologai. T7 fago gp2.5 buvo rasti 26% atvejų. Šie SSB naujai nustatyti ir charakterizuoti NCLD virusuose. T4-fagų SSB (gp32) turi 18% reprezentatyvių atstovų. SSB baltymai, kuriuose nerasta OB-domeno, sudaro 11% atvejų. Jie yra labai įvairūs ir dažniausiai specifiški taksonui: lambda fago ea10, phi29 fago gp5, PRD1 fago P12 ir P19, N4 fago SSB, adeno virusų E2A, raupų virusų I3. Pastarojo analizė, pateikta žemiau, atskleidė jo panašumą į bakterijų SmpB. RpA homologas su reikšmingu įverčiu buvo rastas tik EhV-86 viruse. Didelis jo panašumas į ląstelinius organizmus ir šio baltymo nebuvimas virusuose, gimininguose EhV-86, leidžia manyti, kad jis įgytas horizontalios genų pernašos būdu. Bakulo virusų SSB (LEF3) yra būtinas viruso replikacijai, tačiau jo struktūra ir panašumas į charakterizuotus SSB nenustatyti (Yu ir Carstens, 2010). Visiems LEF3 homologams buvo atliktos jautrios paieškos HHsearch programa. Daugeliu atvejų buvo rastas nereikšmingas (HHsearch tikimybė siekė iki 57% (gi: 285002415)) panašumas į archėjų-eukariotų RpA. Minėtų rezultatų palyginiai parodė, kad LEF3 galėtų turėti RpA baltymo OB domenų. Tikintis pagerinti rezultatų reikšmingumą, baltymai suskaldyti į tris tikėtinus domenų ir paieškos pakartotos. Pirmasis domenų rodo nereikšmingą panašumą į įvairius domenų, o antram ir trečiam domenams panašumo į RpA OB reikšmingumas išaugo (atitinkamai iki 90% (gi: 29567141) ir 77% (gi: 118197537)). Tikėtina, kad bakulo virusų LEF3 turi stipriai divergavusius OB-domenų, giminingus RpA. Dėl mažo jų panašumo į ląstelių organizmų RpA nepavyko sudaryti LEF3 modelių ir atlikti detalesnės jų analizės.

SSB baltymai būtini ląstelių organizmų DNR replikacijai. Juos taip pat turi virusai, kurių replikacijos aparatas gerai charakterizuotas (pvz., T4, T7 ir phi29 fagai, herpes, bakulo ir raupų). Tai žinant, stebėtina, kad santykinai mažai SSB baltymų buvo rasta didelių genomų virusuose (3.11 pav.). SSB nerasta *Nimaviridae*, *Alloherpesviridae*, *Nudivirus*, *Malacoherpesviridae*, *Hytrosaviridae*, *Asfarviridae* šeimose bei phiKZ faguose (žr. priedas). SSB baltymų svarba DNR replikacijai leidžia manyti, kad minėtos šeimos turi šiuos baltymus, tačiau jų nepavyko rasti. NCLDV, bakulo ir herpes virusų SSB

pavyzdžiai rodo, kad paieškas gali apsunkinti stipriai divergavę OB-domenai. Kompiuterinę analizę dar labiau komplikuoja tai, kad SSB funkcijai gali būti panaudoti netipiški baltymai, kuriems nebūdinga OB-sankloda. Tai iliustruoja raupų virusų ir kai kurių fagų SSB pavyzdžiai.

3.4.1. Keturiuose NCLDV šeimose rasti OB-sanklodos SSB baltymai

SSB baltymų paieška NCLD virusuose pradėta tikintis juose rasti tipiškus, bet stipriai divergavusius OB-sanklodos SSB. Šiam tikslui pasiekti buvo atliktos iteratyvios homologų paieškos PSI-BLAST ir jackhmmmer programomis. Sekų ieškota nr70 duomenų bazėje (NCBI „nr“ duomenų bazė nufiltruota iki 70% identiškumo) naudojant įtraukimo ribą E-value=0.001. Užklausoms pasirinkti gerai charakterizuoti ląstelinių organizmų ir virusų baltymai. Netikėtai, užklausiai naudojant T7 fago SSB baltymą (gp2.5) (gi: 9627442), buvo statistiškai reikšmingai (E-value<0.001) nustatyti jo homologai keturiuose NCLDV šeimose (*Phycodnaviridae*, *Mimiviridae*, *Iridoviridae* ir *Marseilleviridae*). Šis rezultatas buvo patikrintas vykdant atvirkštinę paiešką. HHsearch programa sugeneruoti naujai rastų SSB sekų profiliai rodė reikšmingą (tikimybė virš 95%) panašumą į T7 gp2.5 ir du profilius su necharakterizuotų domenų (domains of unknown function (DUFs)) sekomis iš Pfam duomenų bazės. Viename iš jų (DUF2738) yra dalis NCLDV SSB sekų, o kitą (DUF2815) sudaro daugiausiai profagų SSB baltymai (3.18 pav.). Taigi, bendrai paėmus, atlikti tyrimai patikimai rodo, kad keturioms NCLDV šeimoms būdingi T7 fago SSB homologai. Minėti rezultatai yra ne tik patikimi, bet ir specifiniai, nes T7 gp2.5 OB-sankloda turi išskirtinių bruožų. Pavyzdžiui, T7 gp2.5, statinaitę dengianti α spiralė yra įterpta ne tarp $\beta 3$ ir $\beta 4$ juostų, kaip kituose OB baltymuose, o tarp $\beta 2$ ir $\beta 3$.



3.18 pav. NCLDV SSB ir jų homologai sugrupuoti pagal tarpusavio sekų panašumą. Pfam profilių sekos pavaizduotos trikampaiais (DUF2815, oranžinė; DUF2738, kitos spalvos). Taksonominių grupių pavadinimai pateikti virš klasterių (EhV86, *Emiliana huxleyi virus 86*).

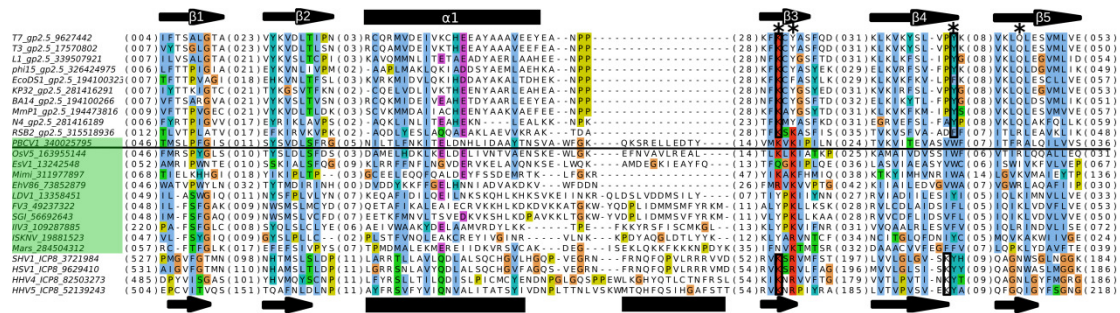
Kai kuriuose NCLDV genomuose buvo rasta daugiau nei vienas T7 SSB homologas. NCLDV SSB ortologai buvo identifikuoti CLANS programa klasterizuojant juos pagal tarpusavio sekų panašumą. Dažniausiai ortologams yra būdingas daug rūgštinių liekanų turintis C-galas. Žinoma, kad T7 gp2.5 C-galas yra naudojamas sąveikai su kitais baltymais (Marintcheva, Hamdan et al., 2006). Paralogų funkcinis vaidmuo nėra aiškus. Tikėtina, kad jie dalyvauja formuojant oligomerinius SSB. Pavyzdžiui, T7 fago, *E. coli* ir žmogaus SSB tyrimai rodo, kad pastarieji veikia kaip oligomerai (Kim ir Richardson, 1994; Raghunathan, Ricard et al., 1997; Bochkareva, Korolev et al., 2002).

Phycodnaviridae šeimos atstovas *Emiliana huxleyi virus 86* (EhV86) turi ne tik stipriai divergavusį T7 SSB homologą (3.18 pav.), bet ir du baltymus, giminingus archėjų-eukariotų SSB (Replication Protein A, RPA). Tikėtina, kad EhV86 gp2.5 homologas stipriai divergavo tuomet, kai SSB vaidmenį šiame viruse perėmė horizontaliu genų pernašos būdu įgyti minėti RPA baltymai.

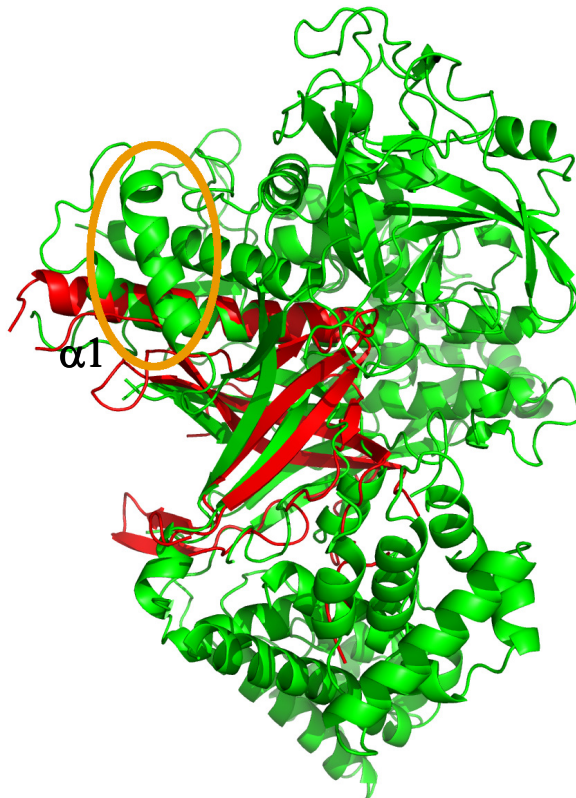
Nustatyta homologija tarp NCLDV SSB ir T7 gp2.5 leidžia manyti, kad jie atlieka panašias funkcijas DNR replikacijos ir rekombinacijos procesuose. Ši teiginį palaiko NCLD virusų genomino konteksto analizė. Nustatyta, kad NCLDV SSB dažnai koduojami greta kitų DNR replikacijos ir rekombinacijos baltymų, pavyzdžiui, DNR polimerazės, helikazės, topoizomerazės, žiedai ir jų užkėlėjai. Be to, transkriptomikos duomenys rodo, kad naujai rasti SSB iš *Frog virus 1 (Iridoviridae)*, PBCV-1 (*Phycodnaviridae*) ir mimi viruso yra aktyviai ekspresuojami ankstyvosiose infekcijos stadijose kartu su kitais DNR replikacijos baltymais (Majji, Thodima et al., 2009; Yanai-Balser, Duncan et al., 2010; Legendre, Audic et al., 2010).

3.4.2. NCLDV SSB struktūra tinkama sąveikai su vgDNR

Siekiant geriau struktūriškai ir funkciškai pažinti NCLDV SSB baltymus, vieno iš jų atstovų struktūra (phycodna viruso PBCV-1 SSB (gi: 340025795)) buvo sumodeliuota. Modeliavimas vyko iteratyviai optimizuojant taikinio-šablono palyginį bei bandant įvairias šablonų kombinacijas (Venclovas ir Margelevičius, 2009). Pradinis PBCV-1 SSB modelis buvo sudarytas šablonu naudojant tik T7 gp2.5 kristalinę struktūrą (PDB: 1je5). Vėlesnė analizė parodė, kad PBCV-1 SSB $\alpha 1$ spirale ir toliau esantis regionas (3.19 pav.) galėtų būti panašesnis į herpes virusų SSB (ICP8, PDB: 1urj; 3.20 pav.) atitinkamas vietas. Kaip ir galima buvo tikėtis, modeliavimui kartu naudojant T7 ir herpes virusų SSB šablonus, modelio kokybė žymiai pagerėjo. Galutinio modelio (prieinamas adresu: http://www.ibt.lt/bioinformatics/models/pbcv1_ssb/) ProSA-web įvertis buvo geresnis nei modeliavimui naudoto pagrindinio šablono (atitinkamai: PBCV-1 SSB, -7,34 ir T7 gp2.5, -6,96), kuris įvertintas jam įmodeliavus trūkstamas kilpas. Taigi, geras modelio kokybės įvertinimas parodė, kad jis yra pakankamai tikslus ir tinkamas tolimesnei struktūrinei-funkcinei analizei.



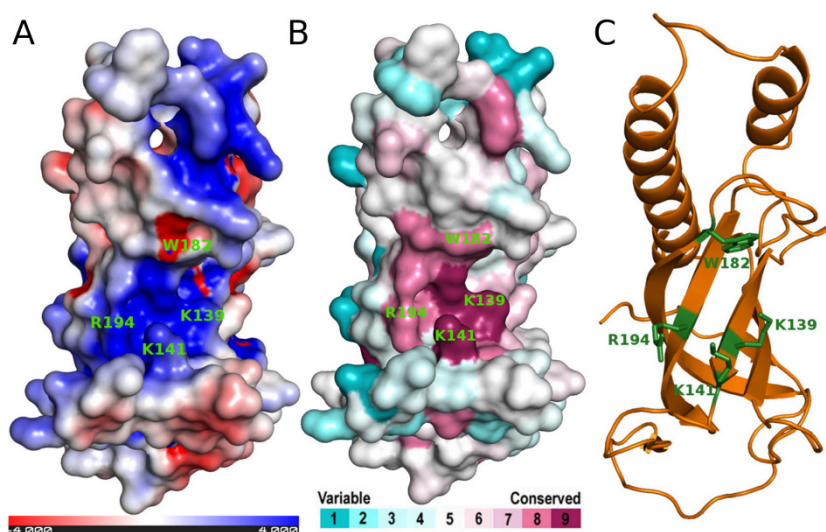
3.19 pav. T7 ir į jį panašių fagų, NCLDV ir herpes virusų SSB baltymų OB-sanklodos regiono palyginys. Sekos pavadinimą sudaro viruso akronimas, baltymo pavadinimas (jei yra) ir gi numeris. T7 gp2.5 ir HHV-1 ICP8 baltymų antrinė struktūra pateikta atitinkamai virš ir žemiau palyginio (β juostos vaizduojamos kaip strėlės, α spiralės – stačiakampiai). NCLDV SSB liekanos, kurios manomai sąveikauja su vgdNR, pažymėtos žvaigždutėmis. Juoduose stačiakampiuose pateikiamos a-r., kurių svarba sąveikai su vgdNR nustatyta eksperimentiškai (T7 gp2.5, K109 ir Y158; SHV1 ICP8, K756 ir K970). NCLDV SSB sekų pavadinimai pavaizduoti žaliame fone, o PBCV-1 SSB – pabrauktas.



3.20 pav. T7 fago gp2.5 (PDB kodas: 1je5; raudona) ir HHV-1 ICP8 (PDB kodas: 1urj; žalia) struktūrų išlyginimas naudojant Dali serverį. ICP8 turi papildomą α -spiralę (apvesta oranžiniu ovalu) už α 1 spiralės. Panašus struktūrinis motyvas nustatytas ir PBCV-1 atveju, todėl modeliuojant kartu su pagrindiniu šablonu, buvo naudotas minėtas ICP8 regionas.

Norint nustatyti ar NCLDV SSB yra tinkamas sąveikai su vgdNR, buvo iširta jo struktūros modelio paviršiaus savybės (3.21 pav.). Dažnai baltymai,

sąveikaujantys su DNR, sąveikos vietoje turi padidėjusį teigiamą krūvį. Nustatyta, kad PBCV-1 SSB yra būdingas išaugęs teigiamas krūvis regione, kuris T7 gp2.5 baltyme, manoma, jungiasi su vgDNR (Hyland, Rezende et al., 2003). Be to, tam pačiam PBCV-1 SSB regionui yra būdingos pačios konservatyviausios aminorūgštys (3.21 pav., B). Remiantis konservatyvumu ir(arba) elektrosstatinėmis savybėmis buvo nustatytos keturios liekanos, kurios tikėtina, sąveikauja su vgDNR (3.19 ir 3.21 pav.)



3.21 pav. PBCV-1 SSB modelis ir jo tikėtinas sąveikos su vgDNR regionas. Liekanos, manomai svarbios sąveikai su vgDNR, pažymėtos žaliai. (A) Paviršiaus elektrosstatinio potencialo žemėlapis (neigiamas krūvis, raudona; teigiamas, mėlyna; skalės vienetai, K_bT/e_c). (B) NCLDV SSB baltymų a-r. konservatyvumas pavaizduotas ant PBCV-1 SSB modelio paviršiaus (nekonservatyvios liekanos, šviesiai mėlyna; konservatyvios, tamsiai raudona). (C) PBCV-1 SSB modelis, kuriame sąveikai su vgDNR svarbios a-r. pavaizduotos lazdelėmis.

Eksperimentai rodo, kad dviejų iš jų atitikmenys T7 fago ir kiaulių herpes viruso SSB baltymuose sąveikauja su vgDNR. PBCV-1 SSB K139 atitinka SHV1 SSB K756, kuris yra būtinas sąveikai su vgDNR (Wu, Li et al., 2009), o tos pačios pozicijos mutacija (K109I) T7 gp2.5 baltyme keičia sąveikos su vgDNR pobūdį (Hyland, Rezende et al., 2003). T7 SSB Y158C mutantai nesijungia su vgDNR (Hyland, Rezende et al., 2003), tikriausiai dėl to, kad suardoma stekingo sąveika su nukleotido baze. Šį teiginį palaiko tas faktas, kad NCLDV SSB šioje pozicijoje (PBCV-1, W182) aptinkamos vien tik aromatinės liekanos (3.19 pav.). SHV1 SSB minėtoje pozicijoje yra tirozinas

(Y971). Tiesioginių įrodymų, kad pastarasis sąveikauja su vgDNR nėra, tačiau nustatyta, kad jo kaimyno (K970) mutacija reikšmingai įtakoja jungimąsi prie vgDNR (Wu, Li et al., 2009). Eksperimentinių duomenų apie likusių dviejų NCLDV SSB pozicijų (PBCV-1, K141 ir R194) svarbą sąveikai su DNR nėra. Tačiau jų buvimas sąveikos su vgDNR griovyje ir didelis konservatyvumas leidžia manyti, kad jos yra svarbios jungimuisi prie vgDNR.

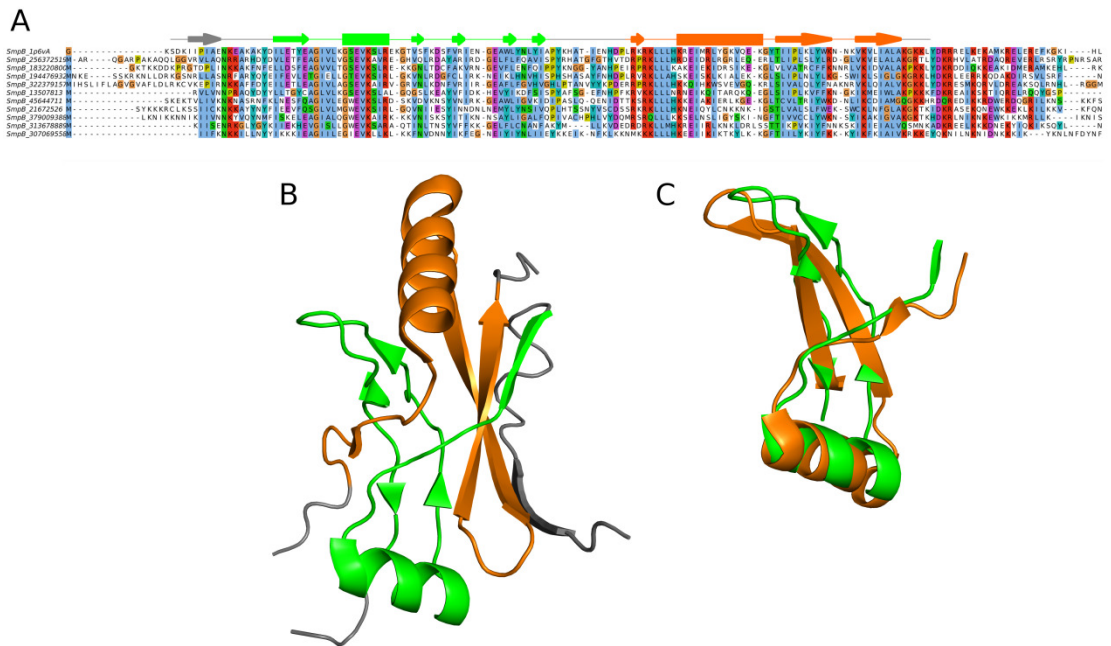
3.4.3. Raupų SSB baltymai (I3) negiminingi NCLDV SSB

Raupų virusuose nebuvo rasta nei T7 gp2.5, nei kitų OB-sanklodos SSB baltymų homologų. Tai nenuostabu, nes yra žinoma, kad geriausiai ištyrinėtam karvių raupų virusui (vaccinia) yra būdingas vgDNR surišantis baltymas (I3) (Rochester ir Traktman, 1998). Tyrimai rodo, kad I3 baltymas dalyvauja DNR replikacijoje bei yra svarbus viruso DNR rekombinacijai (Gammon ir Evans, 2009; Greseth, Boyle et al., 2012). Tačiau, nepaisant didelių pastangų, I3 panašumas į kitas baltymų šeimas nebuvo nustatytas (Greseth, Boyle et al., 2012).

Pradinei I3 homologų paieškai buvo naudotos PSI-BLAST ir jackhmmer programos atliekant iki penkių iteracijų su 0,001 įtraukimo riba. Ieškota per nr70 duomenų bazę, o užklausai pasirinktos reprezentatyvios I3 šeimos sekos. Deja, homologų su reikšmingais įverčiais nepavyko rasti. Tuomet viskas pakartota naudojant „švelnesnę“ įtraukimo ribą (0,01). Šiomis sąlygomis, jackhmmer programa nustatė panašumą į bakterijų SmpB (Small protein B) baltymą. Papildoma analizė buvo atlikta, norint įsitikinti šio rezultato patikimumu. Buvo pastebėta, kad SmpB baltymas išsilygina tik su centrine I3 baltymo dalimi. Tyrimai rodo, kad šalia esantys, nesusiję domenai gali trukdyti homologų paieškai. Taigi, iškelta prielaida, kad jei SmpB ir I3 homologija yra reikšminga, tai pašalinus aplinkinius domenų reikšmingumo įvertis turėtų padidėti. Ši idėja pasitvirtino. Viena iš I3 sekų (gi: 115531731) su sutrumpintais N- ir C-galais, naudojant jackhmmer programą ir nustačius griežtą įtraukimo ribą (E-value=0,001), rado SmpB baltymus. Tolimesnei

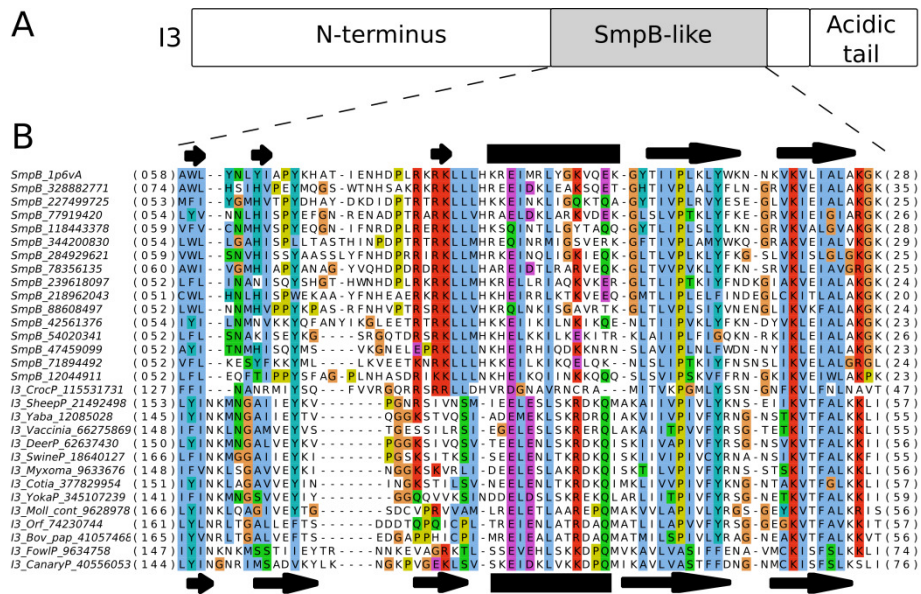
analizei buvo naudotos jautrios paieškos HHsearch programa. Iš pradžių buvo sudaryti profiliai visoms reprezentatyvioms pilno ilgio I3 sekoms. Naudojant juos, atliktos paieškos per sekų su žinomomis struktūromis (PDB) profilių duomenų bazę. HHsearch su geriausiais įverčiais rasdavo *Aquifex aeolicus* SmpB (PDB: 1p6v). Reikšmingiausias rezultatas (tikimybė=89%) gautas užklausai naudojant Yoka raupų viruso I3 (gi: 345107239). Šį rezultatą sustiprino tas faktas, kad HHsearch ir jackhmmer atvejais buvo gauti identiški I3 ir SmpB sekų palyginiai (3.22 pav.). Be to, atliekant HHsearch piešką su I3 (gi: 41057468), kurio N- ir C-galai buvo sutrumpinti, SmpB rastas su statistiškai reikšmingu įverčiu (tikimybė=95%). Kaip papildomi SmpB ir I3 giminytės testai buvo atliktos atvirkštinės paieškos. Naudojant SmpB profilius buvo ieškota homologų per Pfam duomenų bazę. HHsearch programa su geriausiu įverčiu (neskaitant SmpB profilio suradimo) rado I3 profilį. Paieškai naudojant tik SmpB (gi: 240047684) regioną, išsilyginusį su I3, buvo gautas statistiškai reikšmingas (tikimybė=95%) HHsearch įvertis. Taigi, remiantis minėtais rezultatais, galima teigti, kad SmpB ir I3 baltymai tikriausiai yra bendros kilmės.

SmpB yra bakterijų baltymas surišantis tmRNR. Pastaroji RNR molekulė atlieka tiek tRNR, tiek mRNR funkcijas. SmpB baltymas struktūriškai nepanašus į kitas baltymų šeimas. Nepaisant to, pasiūlyta, kad jam yra būdinga sankloda panaši į OB (Dong, Nowakowski et al., 2002). Atlikta SmpB struktūros analizė rodo, kad šis teiginys yra klaidinantis. Nustatyta, kad SmpB pasižymi $\beta\alpha\beta\beta$ struktūrinio motyvų duplikacija (3.22 pav.). Taigi, pseudosimetriška SmpB struktūra sudaryta iš dviejų pasikartojančių dalių, nėra panaši į pasikartojimų neturinčią ir asimetrišką OB-sanklodą.



3.22 pav. SmpB turi du struktūriškai panašius $\beta\beta\beta$ pasikartojimus. (A) Reprezentatyvių SmpB atstovų palyginys. *Aquifex aeolicus* SmpB (PDB: 1p6v) antrinė struktūra, nustatyta DSSP programa, pateikta virš palyginio. Du pasikartojimai pateikti atitinkamai žalia ir oranžine spalvomis. (B) *A. aeolicus* SmpB struktūra su nuspalvintais pasikartojimais (C) *A. aeolicus* SmpB pasikartojimai užkloti Dali programa.

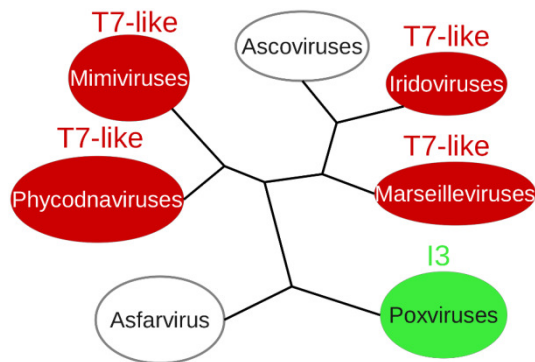
Įdomu tai, kad tik antrasis SmpB pasikartojimas pilnai išsilygina su I3 (3.23 pav.). Tikėtina, kad šis pasikartojimas yra konservatyvesnis. Be to, išsilyginusiam regionui būdingos I3 ir SmpB funkcijai svarbios liekanos (Gutmann, Haebel et al., 2003; Greseth, Boyle et al., 2012). I3 C-galas yra praturtintas aspartato ir glutamato liekanomis bei, panašu, kad yra nestruktūrizuotas. Šiomis savybėmis jis primena rūgštinės fagų ir bakterijų SSB baltymų uodegas, naudojamas sąveikai su baltymais.



3.23 pav. Raupų SSB (I3) ir SmpB šeimų panašumas. (A) I3 sekos regionai. (B) SmpB ir I3 atstovų sekų palyginys. Iš A. *aeolicus* SmpB kristalinės struktūros (PDB: 1p6v) nustatyta ir HHsearch nuspėta vaccinia I3 antrinė struktūra pateikta atitinkamai virš ir po palyginiu.

3.4.4. NCLDV vgDNR surišančių baltymų taksonominis pasiskirstymas ir evoliucija

NCLD virusų filogenetiniame medyje buvo pažymėtas jų SSB taksonominis pasiskirstymas (3.24 pav.). Manoma, kad NCLDV yra monofiletinės kilmės (Iyer, Aravind et al., 2001), o fagai vaidino svarbų vaidmenį ankstyvoje NCLDV evoliucijoje (Iyer, Aravind et al., 2001; Iyer, Balaji et al., 2006; Koonin ir Yutin, 2010). Tikėtina, kad NCLDV protėviai jau turėjo į T7 fago gp2.5 panašų SSB. Iš tikrųjų, daugelis NCLDV genų, būtinų replikacijai, rodo faginę kilmę (pvz., DNR praimzė-helikazė, NAD-ligazė ir resolvazė) (Koonin ir Yutin, 2010). Šiame tyrime rasti T7 gp2.5 homologai papildo minėtą sąrašą. I3 baltymas nėra giminingas minėtiems SSB. Tikėtina, kad šis baltymas raupų virusuose atsirado įvykus neortologiniam geno pakeitimui.



3.24 pav. NCLDV kilmės medis (adaptuota iš (Koonin ir Yutin, 2010) ir jų SSB taksonominis pasiskirstymas (T7-like, naujai rasti NCLDV SSB baltymai).

3.5. Pradmens pašalinimo baltymai

DgDNR virusuose rastos FEN, RNaseH ir PD-(D/E)XK šeimų 5'-3' egzonukleazės. Žinoma, kad ląstelinių organizmų pradmens pašalinime dalyvauja FEN ir RNaseH šeimų atstovai, tuo tarpu PD-(D/E)XK šeimos baltymai atlieka įvairias nukleorūgščių metabolizmo funkcijas (Yang, 2011). Nustatyta, kad pastarosios šeimos atstovas – žmogaus Dna2 – dalyvauja ilgų pradmenų pašalinime (Yang, 2011), o herpes UL12 šarminė egzonukleazė (λ egzonukleazės homologas) turi 5'-3' RNazės aktyvumą (Buisson, Geoui et al., 2009).

Baltymai, galimai atliekantys pradmens pašalinimo funkciją, virusų genomuose buvo rasti santykinai dažnai (3.11 pav.). Mažiausią jų dalį (8%) sudaro RNaseH grupė. Šie baltymai pasižymi plačiu taksonominiu paplitimu. Juos turi eukariotų virusai (šeima *Phycodnaviridae*), fagai (ϕ KZ) ir archėjų virusai (Halovirus HCTV-1 ir HVTV-1). Daugumą rastų baltymų sudaro PD-(D/E)XK šeimos 5'-3' egzonukleazės (62%). Joms priklauso baltymai, homologiški Dna2, UL12 ir λ egzonukleazei. Tikėtina, kad dalis šios grupės baltymų yra specifiški DNR ir dalyvauja tik viruso homologinėje rekombinacijoje. Dėl šios priežasties pradmens pašalinimo funkciją atliekančių baltymų galbūt yra mažiau nei rastų baltymų.

3.6. Ligazės

DNR ligazės, kaip ir procesyvumo veiksniai, virusuose buvo randamos retai ir dažniausiai didesniuose genomuose (3.11 pav.). ATP kaip kofaktorių naudojančių ligazių buvo rasta daugiau (61% visų atvejų). Pastebėta, kad ligazių tipas (ATP/NAD) nėra specifiškas taksonui. Pavyzdžiui, T4-like fagų grupėje rastos tiek ATP, tiek NAD ligazės (atitinkamai 63 ir 35 atvejai). Žinoma, kad NCLD virusų ligazės evoliucijos eigoje būdavo prarandamos, neortologiškai pakeičiamos arba įgyjamos (Yutin ir Koonin, 2009). NCLDV grupei priklauso didžiausi žinomi Pandora virusai. Juose DNR ligazės buvo nerastos. Tai stebina, nes šie baltymai didesniuose genomuose koduojami dažniau (3.11 pav.) . Tai gali būti dėl to, kad virusas galbūt naudoja šeimininko ligazę. Pavyzdžiui, nustatyta, kad vacinia virusas su išmutuota savo ligaze sėkmingai dauginasi, o nuslopinus šeimininko ligazės raišką, mutantas neišgyvena (Paran, De Silva et al., 2009). Taigi, panašu, kad DNR ligazės yra įgyjamos, prarandamos arba neortologiškai pakeičiamos lengviau nei kiti replikacijos baltymai.

3.7. Topoizomerazės

DNR topoizomerazių, kaip ir žiedo užkėlimo kompleksų, buvo rasta mažiausiai (3.11 pav.). Jos yra būdingos tik didelių genomų (virš 130 kbp) virusams. Archėjų virusuose topoizomerazių nerasta (3.12 pav.). Nustatyta, kad kartais virusai turi kelių tipų topoizomerazes. Žinoma, kad visų tipų (išskyrus IA) atstovai dalyvauja mažinant teigiamą superspiralizaciją, susidarančią DNR replikacijos metu (Vos, Tretter et al., 2011). Virusuose IA tipas buvo rastas retai (13% atvejų) ir tik kartu su kitomis topoizomerazėmis. Įdomu tai, kad virusai, turintys topoizomerazes, turi ir ligazę (išskyrus S-TIM5 fagą). Galbūt pastaroji padeda topoizomerazei liguoti DNR gandinę(es) po DNR superspiralizacijos pakeitimo.

IŠVADOS

1. Replikazės komponentų (ne)buvimas genome ir jų tipai priklauso nuo viruso genomo dydžio: maži virusai (<40 kbp) dažnai neturi DNR polimerazių arba turi B-šeimoms polimerazę pradmeniu naudojančią baltymą. Didėjant genomui virusai vis dažniau koduoja nuosavą DNR polimerazę, o patys didžiausi atstovai (>140 kbp) dažnai papildomai turi procesyvumo veiksnį ir kartais jo užkėlėją.
2. PhiKZ fagai turi stipriai divergavusias B-šeimoms DNR polimerazes.
3. Didėjant viruso genomui jame dažniau koduojami nuosavi replikacijos baltymai, o pats dažniausias virusų replikacinis baltymas yra helikazė.
4. Herpes virusų helikazės-praimazės subvienetas (UL8) yra inaktyvuota B šeimoms DNR polimerazė.
5. Didžiausi eukariotų virusai (NCLDV) turi bent dvi evoliuciškai skirtingas SSB šeimas, negiminingas eukariotų baltymams.

MOKSLINIŲ DARBŲ SĄRAŠAS

Disertacijos rezultatai paskelbti straipsniuose:

1. **Kazlauskas D.**, Venclovas Č. (2014). Herpesviral helicase-primase subunit UL8 is inactivated B-family polymerase. *Bioinformatics*, 30:2093-2097.
2. **Kazlauskas D.**, Venclovas Č. (2012) Two distinct SSB protein families in nucleo-cytoplasmic large DNA viruses. *Bioinformatics*, 28:3186-3190.
3. **Kazlauskas D.**, Venclovas Č. (2011) Computational analysis of DNA replicases in double-stranded DNA viruses: relationship with the genome size. *Nucleic Acids Res*, 39:8291-305.

Disertacijos rezultatai pristatyti tarptautinėse konferencijose:

1. **Kazlauskas D.**, Venclovas Č. Viral DNA replication: new insights and discoveries from large scale computational analysis. *European Conference on Computational Biology*, Strasbūras, Prancūzija, 2014 09 06-10 d. (žodinis ir stendinis pranešimai).
2. **Kazlauskas D.**, Venclovas Č. Viral DNA replication: new insights and discoveries from large scale computational analysis. *XIIIth International Conference of Lithuanian Biochemical Society*, Birštonas, Lietuva, 2014 06 18-20 d. (stendinis pranešimas).
3. **Kazlauskas D.**, Venclovas Č. Computational identification and analysis of Single-Stranded DNA Binding Proteins from Nucleo-Cytoplasmic Large DNA Viruses. „*Viruses of Microbes*“ Meeting of International Society for Viruses of Microorganisms. Briuselis, Belgija, 2012 07 16-20 d. (stendinis pranešimas).
4. **Kazlauskas D.**, Venclovas Č. Computational analysis of DNA replicases in double-stranded DNA viruses: Relationship with the genome size. *International Society for Computational Biology / European Conference on Computational Biology*, Viena, Austrija, 2011 07 14-20 d. (stendinis pranešimas).

FINANSINĖ PARAMA

Disertacijos darbą finansiškai parėmė Lietuvos Švietimo ir Mokslo ministerija, Howard Hughes medicinos institutas (55005627), Europos bendrijos septintosios bendrosios programos (FP7-REGPOT-2009-1) „Mobili“ projektas (245721) bei Lietuvos mokslo taryba (Visuotinės dotacijos projektas VP1-3.1-ŠMM-07-K-03-004).

GYVENIMO APRAŠYMAS

Vardas, Pavardė	Darius Kazlauskas
Gimimo data	1985 12 20
Darbo adresas	Bioinformatikos skyrius, Biotechnologijos institutas, Vilniaus universitetas V. A. Graičiūno g. 8 Vilnius LT-02241
Telefonas	+37060191333
E. paštas	dariausk@gmail.com
Išsilavinimas	
2004 – 2008	Molekulinės biologijos bakalauras Vilniaus universitetas
2008 – 2010	Biochemijos magistras Vilniaus universitetas
2010 – 2014	Biochemijos doktorantūros studijos Biotechnologijos institutas, Vilniaus universitetas
Darbo patirtis	Vilniaus universiteto Biotechnologijos instituto Bioinformatikos skyriaus:
2006 – 2014	Bioinžinierius
2013 – dabar	Jaunesnysis mokslo darbuotojas

PADĖKA

Esu dėkingas Česlovui Venclovui už vadovavimą bei pagalbą rengiant straipsnius. Mindaugui Margelevičiui, Albertui Timinskui, Ievai Mitašiūnaitei-Besson ir Ryčiui Dičiūnui už techninę pagalbą. Justui Dapkūnui, Kęstučiui Timinskui, Kliment Olechnovič, Mart Krupovič ir Elenai Bakienei už naudingus patarimus. Algirdui Grybauskui – už pagalbą analizuojant virusų DNR helikazes. Giedriui Sasnauskui, Editai Sužiedėlienei bei Anai Venclovienei už komentarus ir patarimus rengiant straipsnius. Peter Kerpedjiev dėkingas už straipsnius, kurie man nebuvo prieinami.

Galiausiai, nuoširdžiai dėkoju žmonai Natalijai ir savo tėvams už supratimą ir palaikymą.

LITERATŪROS SĄRAŠAS

1. Akabayov, B., S. R. Akabayov, et al. (2010). "Conformational dynamics of bacteriophage T7 DNA polymerase and its processivity factor, Escherichia coli thioredoxin." Proc Natl Acad Sci U S A **107**(34): 15033-15038.
2. Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-3402.
3. Andraos, N., S. Tabor, et al. (2004). "The highly processive DNA polymerase of bacteriophage T5. Role of the unique N and C termini." J Biol Chem **279**(48): 50609-50618.
4. Appleton, B. A., J. Brooks, et al. (2006). "Crystal structure of the cytomegalovirus DNA polymerase subunit UL44 in complex with the C terminus from the catalytic subunit. Differences in structure and function relative to unliganded UL44." J Biol Chem **281**(8): 5224-5232.
5. Appleton, B. A., A. Loregian, et al. (2004). "The cytomegalovirus DNA polymerase subunit UL44 forms a C clamp-shaped dimer." Mol Cell **15**(2): 233-244.
6. Aravind, L. ir E. V. Koonin (1999). "Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches." J Mol Biol **287**(5): 1023-1040.
7. Aravind, L., D. D. Leipe, et al. (1998). "Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins." Nucleic Acids Res **26**(18): 4205-4213.
8. Ashkenazy, H., E. Erez, et al. (2010). "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids." Nucleic Acids Res **38**(Web Server issue): W529-533.
9. Azvolinsky, A., S. Dunaway, et al. (2006). "The S. cerevisiae Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes." Genes Dev **20**(22): 3104-3116.
10. Bailey, S., R. A. Wing, et al. (2006). "The structure of T. aquaticus DNA polymerase III is distinct from eukaryotic replicative DNA polymerases." Cell **126**(5): 893-904.
11. Baker, N. A., D. Sept, et al. (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc Natl Acad Sci U S A **98**(18): 10037-10041.
12. Bedford, E., S. Tabor, et al. (1997). "The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on Escherichia coli DNA polymerase I." Proc Natl Acad Sci U S A **94**(2): 479-484.
13. Benarroch, D. ir S. Shuman (2006). "Characterization of mimivirus NAD+-dependent DNA ligase." Virology **353**(1): 133-143.

14. Berdis, A. J., P. Soumillion, et al. (1996). "The carboxyl terminus of the bacteriophage T4 DNA polymerase is required for holoenzyme complex formation." *Proc Natl Acad Sci U S A* **93**(23): 12822-12827.
15. Bezalel-Buch, R., N. Yaffe, et al. (2013). Replication Machinery of Kinetoplast DNA. *Trypanosomatid Diseases*, Wiley-VCH Verlag GmbH & Co. KGaA: 243-260.
16. Bochkareva, E., S. Korolev, et al. (2002). "Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA." *EMBO J* **21**(7): 1855-1863.
17. Bochman, M. L., N. Sabouri, et al. (2010). "Unwinding the functions of the Pif1 family helicases." *DNA Repair (Amst)* **9**(3): 237-249.
18. Boehmer, P. E. ir I. R. Lehman (1997). "Herpes simplex virus DNA replication." *Annu Rev Biochem* **66**: 347-384.
19. Boyle, K. ir P. Traktman (2009). Poxviruses. *Viral Genome Replication*. K. D. Raney, M. Gotte ir C. E. Cameron, Springer US: 225-247.
20. Bowman, G. D., M. O'Donnell, et al. (2004). "Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex." *Nature* **429**(6993): 724-730.
21. Bruning, J. B. ir Y. Shamo (2004). "Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1." *Structure* **12**(12): 2209-2219.
22. Buisson, M., T. Geoui, et al. (2009). "A bridge crosses the active-site canyon of the Epstein-Barr virus nuclease with DNase and RNase activities." *J Mol Biol* **391**(4): 717-728.
23. Burtelow, M. A., P. M. Roos-Mattjus, et al. (2001). "Reconstitution and molecular analysis of the hRad9-hHus1-hRad1 (9-1-1) DNA damage responsive checkpoint complex." *J Biol Chem* **276**(28): 25903-25909.
24. Buttner, K., S. Nehring, et al. (2007). "Structural basis for DNA duplex separation by a superfamily-2 helicase." *Nat Struct Mol Biol* **14**(7): 647-652.
25. Carver, T. J., K. M. Rutherford, et al. (2005). "ACT: the Artemis Comparison Tool." *Bioinformatics* **21**(16): 3422-3423.
26. Cerritelli, S. M. ir R. J. Crouch (2009). "Ribonuclease H: the enzymes in eukaryotes." *FEBS J* **276**(6): 1494-1505.
27. Chen, Y., P. Bai, et al. (2011). "Herpes simplex virus type 1 helicase-primase: DNA binding and consequent protein oligomerization and primase activation." *J Virol* **85**(2): 968-978.
28. Chen, S. H., N. L. Chan, et al. (2013). "New mechanistic and functional insights into DNA topoisomerases." *Annu Rev Biochem* **82**: 139-170.
29. Contesto-Richefeu, C., N. Tarbouriech, et al. (2014). "Crystal structure of the vaccinia virus DNA polymerase holoenzyme subunit D4 in complex with the A20 N-terminal domain." *PLoS Pathog* **10**(3): e1003978.
30. Dahlen, M., T. Olsson, et al. (1998). "Regulation of telomere length by checkpoint genes in *Schizosaccharomyces pombe*." *Mol Biol Cell* **9**(3): 611-621.

31. Dalrymple, B. P., K. Kongsuwan, et al. (2001). "A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems." Proc Natl Acad Sci U S A **98**(20): 11627-11632.
32. Delhon, G., E. R. Tulman, et al. (2006). "Genome of invertebrate iridescent virus type 3 (mosquito iridescent virus)." J Virol **80**(17): 8439-8449.
33. Derelle, E., C. Ferraz, et al. (2006). "Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features." Proc Natl Acad Sci U S A **103**(31): 11647-11652.
34. Dodson, M. S. ir I. R. Lehman (1991). "Association of DNA helicase and primase activities with a subassembly of the herpes simplex virus 1 helicase-primase composed of the UL5 and UL52 gene products." Proc Natl Acad Sci U S A **88**(4): 1105-1109.
35. Dolinsky, T. J., J. E. Nielsen, et al. (2004). "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations." Nucleic Acids Res **32**(Web Server issue): W665-667.
36. Dong, G., J. Nowakowski, et al. (2002). "Structure of small protein B: the protein component of the tmRNA-SmpB system for ribosome rescue." EMBO J **21**(7): 1845-1854.
37. Druck Shudofsky, A. M., J. E. Silverman, et al. (2010). "Vaccinia virus D4 mutants defective in processive DNA synthesis retain binding to A20 and DNA." J Virol **84**(23): 12325-12335.
38. Eddy, S. R. (2011). "Accelerated Profile HMM Searches." PLoS Comput Biol **7**(10): e1002195.
39. Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Res **32**(5): 1792-1797.
40. Fairman-Williams, M. E., U. P. Guenther, et al. (2010). "SF1 and SF2 helicases: family matters." Curr Opin Struct Biol **20**(3): 313-324.
41. Federici, B. A. ir Y. Bigot (2010). Evolution of Immunosuppressive Organelles from DNA Viruses in Insects. Evolutionary Biology - Concepts, Molecular and Morphological Evolution. P. Pontarotti, Springer Berlin Heidelberg: 229-248.
42. Franklin, M. C., J. Wang, et al. (2001). "Structure of the replicating complex of a pol alpha family DNA polymerase." Cell **105**(5): 657-667.
43. Frickey, T. ir A. Lupas (2004). "CLANS: a Java application for visualizing protein families based on pairwise similarity." Bioinformatics **20**(18): 3702-3704.
44. Gammon, D. B. ir D. H. Evans (2009). "The 3'-to-5' exonuclease activity of vaccinia virus DNA polymerase is essential and plays a role in promoting virus genetic recombination." J Virol **83**(9): 4236-4250.
45. Garcia-Gomez, S., A. Reyes, et al. (2013). "PrimPol, an Archaic Primase/Polymerase Operating in Human Cells." Mol Cell **52**(4): 541-553.
46. Gardner, A. F., S. D. Bell, et al. (2014). "Protein-Protein Interactions Leading to Recruitment of the Host DNA Sliding Clamp by the Hyperthermophilic *Sulfolobus islandicus* Rod-Shaped Virus 2." J Virol **88**(12): 7105-7108.

47. Georgescu, R. E., S. S. Kim, et al. (2008). "Structure of a sliding clamp on DNA." Cell **132**(1): 43-54.
48. Gorlas, A., E. V. Koonin, et al. (2012). "TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*." Environ Microbiol **14**(2): 503-516.
49. Greseth, M. D., K. A. Boyle, et al. (2012). "Molecular genetic and biochemical characterization of the vaccinia virus i3 protein, the replicative single-stranded DNA binding protein." J Virol **86**(11): 6197-6209.
50. Gutmann, S., P. W. Haebel, et al. (2003). "Crystal structure of the transfer-RNA domain of transfer-messenger RNA in complex with SmpB." Nature **424**(6949): 699-703.
51. Halgasova, N., I. Mesarosova, et al. (2012). "Identification of a bifunctional primase-polymerase domain of corynephage BFK20 replication protein gp43." Virus Res **163**(2): 454-460.
52. Hertveldt, K., R. Lavigne, et al. (2005). "Genome comparison of *Pseudomonas aeruginosa* large phages." J Mol Biol **354**(3): 536-545.
53. Hyland, E. M., L. F. Rezende, et al. (2003). "The DNA binding domain of the gene 2.5 single-stranded DNA-binding protein of bacteriophage T7." J Biol Chem **278**(9): 7247-7256.
54. Hodgman, T. C. (1988). "A new superfamily of replicative proteins." Nature **333**(6168): 22-23.
55. Holm, L. ir J. Park (2000). "DaliLite workbench for protein structure comparison." Bioinformatics **16**(6): 566-567.
56. Holm, L. ir P. Rosenstrom (2010). "Dali server: conservation mapping in 3D." Nucleic Acids Res **38**(Web Server issue): W545-549.
57. Iyer, L. M., L. Aravind, et al. (2001). "Common origin of four diverse families of large eukaryotic DNA viruses." J Virol **75**(23): 11720-11734.
58. Iyer, L. M., S. Balaji, et al. (2006). "Evolutionary genomics of nucleocytoplasmic large DNA viruses." Virus Res **117**(1): 156-184.
59. Iyer, L. M., E. V. Koonin, et al. (2005). "Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members." Nucleic Acids Res **33**(12): 3875-3896.
60. Iyer, L. M., D. D. Leipe, et al. (2004). "Evolutionary history and higher order classification of AAA+ ATPases." J Struct Biol **146**(1-2): 11-31.
61. Imamura, K., K. Fukunaga, et al. (2007). "Specific interactions of three proliferating cell nuclear antigens with replication-related proteins in *Aeropyrum pernix*." Mol Microbiol **64**(2): 308-318.
62. Indiani, C. ir M. O'Donnell (2006). "The replication clamp-loading machine at work in the three domains of life." Nat Rev Mol Cell Biol **7**(10): 751-761.
63. Ishida, T. ir K. Kinoshita (2007). "PrDOS: prediction of disordered protein regions from amino acid sequence." Nucleic Acids Res **35**(Web Server issue): W460-464.

64. Yanai-Balser, G. M., G. A. Duncan, et al. (2010). "Microarray analysis of *Paramecium bursaria* chlorella virus 1 transcription." J Virol **84**(1): 532-542.
65. Yang, W. (2011). "Nucleases: diversity of structure, function and mechanism." Q Rev Biophys **44**(1): 1-93.
66. Yao, N., L. Coryell, et al. (2003). "Replication factor C clamp loader subunit arrangement within the circular pentamer and its attachment points to proliferating cell nuclear antigen." J Biol Chem **278**(50): 50744-50753.
67. Yao, N. Y., A. Johnson, et al. (2006). "Mechanism of proliferating cell nuclear antigen clamp opening by replication factor C." J Biol Chem **281**(25): 17528-17539.
68. Yoshida, T., K. Nagasaki, et al. (2008). "Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies." J Bacteriol **190**(5): 1762-1772.
69. Yu, M. ir E. B. Carstens (2010). "Identification of a domain of the baculovirus *Autographa californica* multiple nucleopolyhedrovirus single-strand DNA-binding protein LEF-3 essential for viral DNA replication." J Virol **84**(12): 6153-6162.
70. Yutin, N. ir E. V. Koonin (2009). "Evolution of DNA ligases of nucleocytoplasmic large DNA viruses of eukaryotes: a case of hidden complexity." Biol Direct **4**: 51.
71. Johnson, A. ir M. O'Donnell (2005). "Cellular DNA replicases: components and dynamics at the replication fork." Annu Rev Biochem **74**: 283-315.
72. Joosten, R. P., T. A. te Beek, et al. (2011). "A series of PDB related databases for everyday needs." Nucleic Acids Res **39**(Database issue): D411-419.
73. Kallberg, M., H. Wang, et al. (2012). "Template-based protein structure modeling using the RaptorX web server." Nat Protoc **7**(8): 1511-1522.
74. Katoh, K. ir D. M. Standley (2013). "MAFFT multiple sequence alignment software version 7: improvements in performance and usability." Mol Biol Evol **30**(4): 772-780.
75. Kelch, B. A., D. L. Makino, et al. (2012). "Clamp loader ATPases and the evolution of DNA replication machinery." BMC Biol **10**: 34.
76. Kelman, Z. (1997). "PCNA: structure, functions and interactions." Oncogene **14**(6): 629-640.
77. Kelman, Z. ir M. O'Donnell (1995). "Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps." Nucleic Acids Res **23**(18): 3613-3620.
78. Khair, L., Y. T. Chang, et al. (2010). "Roles of the checkpoint sensor clamp Rad9-Rad1-Hus1 (911)-complex and the clamp loaders Rad17-RFC and Ctf18-RFC in *Schizosaccharomyces pombe* telomere maintenance." Cell Cycle **9**(11): 2237-2248.

79. Kim, Y. T. ir C. C. Richardson (1994). "Acidic carboxyl-terminal domain of gene 2.5 protein of bacteriophage T7 is essential for protein-protein interactions." J Biol Chem **269**(7): 5270-5278.
80. Kim, J. H., H. D. Kim, et al. (2006). "Isolation of human Dna2 endonuclease and characterization of its enzymatic properties." Nucleic Acids Res **34**(6): 1854-1864.
81. Knopf, C. W. ir K. Weisshart (1990). "Comparison of exonucleolytic activities of herpes simplex virus type-1 DNA polymerase and DNase." Eur J Biochem **191**(2): 263-273.
82. Komazin-Meredith, G., W. L. Santos, et al. (2008). "The Positively Charged Surface of Herpes Simplex Virus UL42 Mediates DNA Binding." J Biol Chem **283**(10): 6154-6161.
83. Konagurthu, A. S., J. C. Whisstock, et al. (2006). "MUSTANG: a multiple structural alignment algorithm." Proteins **64**(3): 559-574.
84. Kool, M., C. H. Ahrens, et al. (1994). "Identification of genes involved in DNA replication of the Autographa californica baculovirus." Proc Natl Acad Sci U S A **91**(23): 11212-11216.
85. Koonin, E. V. (2006). "Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases." Biol Direct **1**: 39.
86. Koonin, E. V. ir N. Yutin (2010). "Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses." Intervirology **53**(5): 284-292.
87. Kozłowski, L. P. ir J. M. Bujnicki (2012). "MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins." BMC Bioinformatics **13**: 111.
88. Krupovic, M., S. Gribaldo, et al. (2010). "The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements." Mol Biol Evol **27**(12): 2716-2732.
89. Kuchta, R. D. ir G. Stengel (2010). "Mechanism and evolution of DNA primases." Biochim Biophys Acta **1804**(5): 1180-1189.
90. Kurowski, M. A. ir J. M. Bujnicki (2003). "GeneSilico protein structure prediction meta-server." Nucleic Acids Res **31**(13): 3305-3307.
91. Lamers, M. H., R. E. Georgescu, et al. (2006). "Crystal structure of the catalytic alpha subunit of E. coli replicative DNA polymerase III." Cell **126**(5): 881-892.
92. Legendre, M., S. Audic, et al. (2010). "mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus." Genome Res **20**(5): 664-674.
93. Leipe, D. D., L. Aravind, et al. (1999). "Did DNA replication evolve twice independently?" Nucleic Acids Res **27**(17): 3389-3401.
94. Lyubimov, A. Y., M. Strycharska, et al. (2011). "The nuts and bolts of ring-translocase structure and mechanism." Curr Opin Struct Biol **21**(2): 240-248.

95. Lobley, A., M. I. Sadowski, et al. (2009). "pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination." Bioinformatics **25**(14): 1761-1767.
96. Lopez de Saro, F. J. ir M. O'Donnell (2001). "Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I." Proc Natl Acad Sci U S A **98**(15): 8376-8380.
97. Loregian, A., E. Sinigalia, et al. (2007). "Binding parameters and thermodynamics of the interaction of the human cytomegalovirus DNA polymerase accessory protein, UL44, with DNA: implications for the processivity mechanism." Nucleic Acids Res **35**(14): 4779-4791.
98. Maga, G. ir U. Hubscher (2003). "Proliferating cell nuclear antigen (PCNA): a dancer with many partners." J Cell Sci **116**(Pt 15): 3051-3060.
99. Majji, S., V. Thodima, et al. (2009). "Transcriptome analysis of Frog virus 3, the type species of the genus Ranavirus, family Iridoviridae." Virology **391**(2): 293-303.
100. Makarova, K. S. ir E. V. Koonin (2013). "Archaeology of eukaryotic DNA replication." Cold Spring Harb Perspect Biol **5**(11): a012963.
101. Mansky, K. C., A. Batiza, et al. (1997). "Bovine papillomavirus type 1 E1 and simian virus 40 large T antigen share regions of sequence similarity required for multiple functions." J Virol **71**(10): 7600-7608.
102. Mardanov, A. V., T. S. Strakhova, et al. (2004). "RepA protein of the bacteriophage N15 exhibits activity of DNA helicase." Dokl Biochem Biophys **397**: 217-219.
103. Margelevičius, M., M. Laganeckas, et al. (2010). "COMA server for protein distant homology search." Bioinformatics **26**(15): 1905-1906.
104. Margelevičius, M. ir Č. Venclovas (2005). "PSI-BLAST-ISS: an intermediate sequence search tool for estimation of the position-specific alignment reliability." BMC Bioinformatics **6**: 185.
105. Marintcheva, B., S. M. Hamdan, et al. (2006). "Essential residues in the C terminus of the bacteriophage T7 gene 2.5 single-stranded DNA-binding protein." J Biol Chem **281**(35): 25831-25840.
106. Marintcheva, B., A. Marintchev, et al. (2008). "Acidic C-terminal tail of the ssDNA-binding protein of bacteriophage T7 and ssDNA compete for the same binding surface." Proc Natl Acad Sci U S A **105**(6): 1855-1860.
107. Marsden, H. S., G. W. McLean, et al. (1997). "The catalytic subunit of the DNA polymerase of herpes simplex virus type 1 interacts specifically with the C terminus of the UL8 component of the viral helicase-primase complex." J Virol **71**(9): 6390-6397.
108. McCutcheon, J. P. (2010). "The bacterial essence of tiny symbiont genomes." Curr Opin Microbiol **13**(1): 73-78.
109. McGlynn, P. (2013). "Helicases at the replication fork." Adv Exp Med Biol **767**: 97-121.
110. Medagli, B. ir S. Onesti (2013). "Structure and mechanism of hexameric helicases." Adv Exp Med Biol **767**: 75-95.

111. Mesyanzhinov, V. V., J. Robben, et al. (2002). "The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*." J Mol Biol **317**(1): 1-19.
112. Moarefi, I., D. Jeruzalmi, et al. (2000). "Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage." J Mol Biol **296**(5): 1215-1223.
113. Mueser, T. C., J. M. Hinerman, et al. (2010). "Structural analysis of bacteriophage T4 DNA replication: a review in the Virology Journal series on bacteriophage T4 and its relatives." Virology **7**: 359.
114. Muylaert, I., K. W. Tang, et al. (2011). "Replication and recombination of herpes simplex virus DNA." J Biol Chem **286**(18): 15619-15624.
115. Muylaert, I., Z. Zhao, et al. (2012). "Identification of conserved amino acids in the herpes simplex virus type 1 UL8 protein required for DNA synthesis and UL52 primase interaction in the virus replisome." J Biol Chem **287**(40): 33142-33152.
116. Murayama, K., S. Nakayama, et al. (2009). "Crystal structure of Epstein-Barr virus DNA polymerase processivity factor BMRF1." J Biol Chem **284**(51): 35896-35905.
117. Murzin, A. G. (1993). "OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences." Embo J **12**(3): 861-867.
118. Murzin, A. G., S. E. Brenner, et al. (1995). "SCOP: a structural classification of proteins database for the investigation of sequences and structures." J Mol Biol **247**(4): 536-540.
119. O'Shea, V. L. and J. M. Berger (2014). "Loading strategies of ring-shaped nucleic acid translocases and helicases." Curr Opin Struct Biol **25**: 16-24.
120. Paytubi, S., S. A. McMahon, et al. (2012). "Displacement of the canonical single-stranded DNA-binding protein in the Thermoproteales." Proc Natl Acad Sci U S A **109**(7): E398-405.
121. Paran, N., F. S. De Silva, et al. (2009). "Cellular DNA ligase I is recruited to cytoplasmic vaccinia virus factories and masks the role of the vaccinia ligase in viral DNA replication." Cell Host Microbe **6**(6): 563-569.
122. Parry, M. E., N. D. Stow, et al. (1993). "Purification and properties of the herpes simplex virus type 1 UL8 protein." J Gen Virol **74** (Pt 4): 607-612.
123. Patel, S. S., M. Pandey, et al. (2011). "Dynamic coupling between the motors of DNA replication: hexameric helicase, DNA polymerase, and primase." Curr Opin Chem Biol **15**(5): 595-605.
124. Pei, J., B. H. Kim, et al. (2008). "PROMALS3D: a tool for multiple protein sequence and structure alignments." Nucleic Acids Res **36**(7): 2295-2300.
125. Punta, M., P. C. Coggill, et al. (2012). "The Pfam protein families database." Nucleic Acids Res **40**(Database issue): D290-301.

126. Raghunathan, S., C. S. Ricard, et al. (1997). "Crystal structure of the homo-tetrameric DNA binding domain of Escherichia coli single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-Å resolution." Proc Natl Acad Sci U S A **94**(13): 6652-6657.
127. Rice, P., I. Longden, et al. (2000). "EMBOSS: the European Molecular Biology Open Software Suite." Trends Genet **16**(6): 276-277.
128. Rochester, S. C. ir P. Traktman (1998). "Characterization of the single-stranded DNA binding protein encoded by the vaccinia virus I3 gene." J Virol **72**(4): 2917-2926.
129. Roy, A., A. Kucukural, et al. (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." Nat Protoc **5**(4): 725-738.
130. Rothwell, P. J. ir G. Waksman (2005). "Structure and mechanism of DNA polymerases." Adv Protein Chem **71**: 401-440.
131. Rudd, S. G., L. Glover, et al. (2013). "PPL2 translesion polymerase is essential for the completion of chromosomal DNA replication in the African trypanosome." Mol Cell **52**(4): 554-565.
132. Salas, M. (1991). "Protein-priming of DNA replication." Annu Rev Biochem **60**: 39-71.
133. Sarmiento, F., F. Long, et al. (2014). "Diversity of the DNA replication system in the Archaea domain." Archaea **2014**: 675946.
134. Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.
135. Singleton, M. R., M. S. Dillingham, et al. (2007). "Structure and mechanism of helicases and nucleic acid translocases." Annu Rev Biochem **76**: 23-50.
136. Snider, J. ir W. A. Houry (2008). "AAA+ proteins: diversity in function, similarity in structure." Biochem Soc Trans **36**(Pt 1): 72-77.
137. Söding, J. (2005). "Protein homology detection by HMM-HMM comparison." Bioinformatics **21**(7): 951-960.
138. Söding, J., A. Biegert, et al. (2005). "The HHpred interactive server for protein homology detection and structure prediction." Nucleic Acids Res **33**(Web Server issue): W244-248.
139. Stajich, J. E., D. Block, et al. (2002). "The Bioperl toolkit: Perl modules for the life sciences." Genome Res **12**(10): 1611-1618.
140. Steczkiewicz, K., A. Muszewska, et al. (2012). "Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily." Nucleic Acids Res **40**(15): 7016-7045.
141. Suhre, K. (2005). "Gene and genome duplication in Acanthamoeba polyphaga Mimivirus." J Virol **79**(22): 14095-14101.
142. Šali, A. ir T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol **234**(3): 779-815.
143. Tafel, A. A., L. Wu, et al. (2011). "Human HEL308 localizes to damaged replication forks and unwinds lagging strand structures." J Biol Chem **286**(18): 15832-15840.

144. Tahirov, T. H., K. S. Makarova, et al. (2009). "Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors." Biol Direct **4**: 11.
145. Taneja, B., A. Patel, et al. (2006). "Structure of the N-terminal fragment of topoisomerase V reveals a new family of topoisomerases." Embo J **25**(2): 398-408.
146. Thomas, J. A., M. R. Rolando, et al. (2008). "Characterization of Pseudomonas chlororaphis myovirus 201varphi2-1 via genomic sequencing, mass spectrometry, and electron microscopy." Virology **376**(2): 330-338.
147. Timinskas, K., M. Balvočiūtė, et al. (2014). "Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes." Nucleic Acids Res **42**(3): 1393-1413.
148. Tomkinson, A. E., S. Vijayakumar, et al. (2006). "DNA ligases: structure, reaction mechanism, and function." Chem Rev **106**(2): 687-699.
149. Venclovas, Č. ir M. Margelevičius (2009). "The use of automatic tools and human expertise in template-based modeling of CASP8 target proteins." Proteins-Structure Function and Bioinformatics **77**: 81-88.
150. Venclovas, Č. ir M. P. Thelen (2000). "Structure-based predictions of Rad1, Rad9, Hus1 and Rad17 participation in sliding clamp and clamp-loading complexes." Nucleic Acids Res **28**(13): 2481-2493.
151. Vos, S. M., E. M. Tretter, et al. (2011). "All tangled up: how cells direct, manage and exploit topoisomerase function." Nat Rev Mol Cell Biol **12**(12): 827-841.
152. Wang, F. ir W. Yang (2009). "Structural insight into translesion synthesis by DNA Pol II." Cell **139**(7): 1279-1289.
153. Weigel, C. ir H. Seitz (2006). "Bacteriophage replication modules." FEMS Microbiol Rev **30**(3): 321-381.
154. Weller, S. K. ir D. M. Coen (2012). "Herpes simplex viruses: mechanisms of DNA replication." Cold Spring Harb Perspect Biol **4**(9): a013011.
155. Wernersson, R. (2006). "Virtual Ribosome--a comprehensive DNA translation tool with support for integration of sequence feature annotation." Nucleic Acids Res **34**(Web Server issue): W385-388.
156. Wiederstein, M., M. Gruber, et al. (2014). "Structure-Based Characterization of Multiprotein Complexes." Structure.
157. Wiederstein, M. ir M. J. Sippl (2007). "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins." Nucleic Acids Res **35**(Web Server issue): W407-410.
158. Wilson, M. A., Y. Kwon, et al. (2013). "Pif1 helicase and Poldelta promote recombination-coupled DNA synthesis via bubble migration." Nature **502**(7471): 393-396.
159. Wu, S. L., C. C. Li, et al. (2009). "Mutagenesis identifies the critical regions and amino acid residues of suid herpesvirus 1 DNA-binding protein required for DNA binding and strand invasion." Virus Res **140**(1-2): 147-154.

160. Xu, D., L. Jaroszewski, et al. (2013). "FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking." Bioinformatics.
161. Zhuang, Z. ir Y. Ai (2010). "Processivity factor of DNA polymerase and its expanding role in normal and translesion DNA synthesis." Biochim Biophys Acta **1804**(5): 1081-1093.
162. Ziegelin, G., N. Tegtmeyer, et al. (2005). "The repA gene of the linear *Yersinia enterocolitica* prophage PY54 functions as a circular minimal replicon in *Escherichia coli*." J Bacteriol **187**(10): 3445-3454.
163. Zuccola, H. J., D. J. Filman, et al. (2000). "The crystal structure of an unusual processivity factor, herpes simplex virus UL42, bound to the C terminus of its cognate polymerase." Mol Cell **5**(2): 267-278.

Priedas. Replikacijos baltymų taksonominis pasiskirstymas dgDNR virusuose, išrikiuotuose pagal mažėjantį genomo dydį.

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvumo veiksnys	Žiedo užkėiklis	Replikacinė helikazė	Praimazė	SSB	baltymas pašalinimo	Ligazė	Topoizomerazė
Pandoravirus dulcis	1908524	1	1	1	1	1	1	1	0	0
Acanthamoeba polyphaga mimivirus	1181549	1	1	1	1	1	1	1	1	1
Cafeteria roenbergensis virus BV-PW1	617453	1	1	1	1	1	1	1	1	1
Pithovirus sibericum	610033	1	1	0	1	1	1	1	1	1
Bacillus phage G	497513	1	0	0	1	1	1	1	1	1
Phaeocystis globosa virus	459984	1	1	1	1	1	1	1	1	1
Emiliana huxleyi virus 86	407339	1	1	1	1	1	1	1	1	1
Marseillevirus	368454	1	1	0	1	1	1	1	1	1
Cronobacter phage vB_CsaM_GAP32	358663	1	1	1	1	1	1	1	1	1
Organic Lake phycodnavirus 1	344723	1	1	1	1	1	1	1	1	1
Ectocarpus siliculosus virus 1	335593	1	1	1	1	1	1	1	0	0
Paramecium bursaria Chlorella virus 1	330611	1	1	1	1	1	1	1	1	1
Choristoneura biennis entomopoxvirus 'L'	307691	1	0	0	1	1	0	1	1	1
Shrimp white spot syndrome virus	305108	1	0	0	1	1	0	1	0	0
Cyprinid herpesvirus 3	295146	1	0	0	1	1	0	1	0	0
Pseudomonas phage phiKZ	280334	1	0	0	1	0	0	1	0	0
Anguillid herpesvirus 1	248526	1	0	0	1	1	0	1	0	0
Melanoplus sanguinipes entomopoxvirus	236120	1	0	0	1	1	0	1	1	1
Human herpesvirus 5	235646	1	1	0	1	1	1	1	0	0
Ranid herpesvirus 2	231801	1	0	0	1	1	0	1	0	0
Helicoverpa zea nudivirus 2	231621	1	0	0	1	1	0	1	1	0
Ralstonia phage RSL1	231255	1	1	1	1	1	1	1	1	1
Cowpox virus	224499	1	1	0	1	1	1	1	1	1
Ranid herpesvirus 1	220859	1	0	0	1	1	0	1	0	0
Sphingomonas phage PAU	219372	1	0	1	1	1	1	1	1	1
Bacillus phage 0305phi8-36	218948	1	1	0	1	1	0	1	0	0
Caulobacter phage phiCbK	215710	1	0	0	1	1	0	1	1	0
Invertebrate iridescent virus 6	212482	1	1	0	1	1	0	1	1	1
Pseudomonas phage EL	211215	1	0	0	1	0	0	1	1	0
Ostreid herpesvirus 1	207439	1	0	0	1	1	0	1	0	0
Invertebrate iridescent virus 3	191100	1	1	0	1	1	1	1	1	1
Nile crocodilepox virus	190054	1	1	0	1	1	1	1	1	1
Glossina pallidipes salivary gland hypertrophy virus	190032	1	0	0	1	1	0	1	0	0
Heliothis virescens ascovirus 3e	186262	1	0	0	1	1	0	1	0	0

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvuomo veiksnys	Žiedo užkėiklis	Replikacinė helikazė	Praimazė	SSB	Pradmens pašalinimo baltymas	Ligazė	Topoizomerazė
Lymphocystis disease virus - isolate China	186250	1	1	0	1	1	1	1	0	0
Clostridium phage c-st	185683	1	0	0	1	1	1	1	1	1
Ostreococcus virus OsV5	185373	1	1	0	1	1	1	1	1	1
Xestia c-nigrum granulovirus	178733	1	0	0	1	1	1	1	1	0
Gallid herpesvirus 2	177874	1	1	0	1	1	1	1	0	0
Synechococcus phage syn9	177300	1	1	1	1	1	1	1	1	0
Human herpesvirus 4 type 2	172764	1	1	0	1	1	1	1	0	0
African swine fever virus	170101	1	1	0	1	1	0	1	1	1
Enterobacteria phage T4	168903	1	1	1	1	1	1	1	1	1
Halocynthia phage JM-2012	167292	1	0	0	1	0	0	1	0	0
Psittacid herpesvirus 1	163025	1	1	0	1	1	1	1	0	0
Microcystis aeruginosa phage Ma-LMM01	162109	1	1	1	1	1	0	1	0	0
Cyanophage S-TIM5	161440	1	0	0	1	1	0	1	0	1
Deftia phage phiW-14	157486	1	1	1	1	1	1	1	1	1
Salmonella phage ViI	157061	1	1	1	1	1	1	1	1	1
Mycobacterium phage Cali	155372	1	0	0	1	1	1	1	0	0
Thermus phage phiYS40	152372	1	0	0	1	1	0	1	0	0
Human herpesvirus 1	152261	1	1	0	1	1	1	1	0	0
Gallid herpesvirus 1	148687	1	1	0	1	1	1	1	0	0
Cellulophaga phage phi4:1	145865	1	0	0	1	1	1	1	1	0
Halovirus HGTV-1	143855	1	0	0	1	0	0	1	1	0
Lactobacillus phage Lb338-1	142111	1	0	0	1	1	1	1	0	0
Staphylococcus phage G1	138715	1	0	0	1	1	1	1	0	0
Human herpesvirus 8	137969	1	1	0	1	1	1	1	0	0
Escherichia phage rv5	137947	1	0	0	1	1	1	1	1	0
Campylobacter phage NCTC12673	135041	1	1	1	1	1	1	1	1	1
Bacillus phage SPbeta	134416	1	0	0	1	1	1	0	1	0
Ictalurid herpesvirus 1	134226	1	0	0	1	1	0	1	0	0
Bacillus phage SPO1	132562	1	0	0	1	1	1	1	1	0
Lactobacillus phage LP65	131522	1	0	0	1	1	1	0	0	0
Rhodothermus phage RM378	129908	1	0	0	1	1	0	1	0	0
Bombyx mori nucleopolyhedrovirus	128413	1	0	0	1	1	1	1	0	0
Oryctes rhinoceros virus	127615	1	0	0	1	1	0	1	1	0
Musca domestica salivary gland hypertrophy virus	124279	1	0	0	1	1	0	1	0	0
Enterobacteria phage T5	121750	1	0	0	1	1	1	1	1	0

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvuomo veiksnyvs	Žiedo užkėlėjas	Replikacinė helikazė	Praimazė	SSB	Pradmens pašalinimo baltymas	Ligazė	Topozimerazė
Diadromus pulchellus ascovirus 4a	119343	1	1	0	1	1	0	0	0	0
Infectious spleen and kidney necrosis virus	111362	1	1	0	1	1	1	1	0	0
Mycobacterium phage Omega	110865	0	0	0	1	1	0	1	0	0
Culex nigripalpus nucleopolyhedrovirus	108252	1	0	0	1	1	0	1	0	0
Cyanophage PSS2	107530	0	0	0	1	1	1	1	0	0
Frog virus 3	105903	1	1	0	1	1	1	1	0	0
Colwellia phage 9A	104936	1	0	0	1	1	0	1	1	0
Salmonella phage SSU5	103299	1	0	0	1	1	1	1	1	0
Halovirus HCTV-1	103257	1	1	1	1	1	0	1	0	0
Halovirus HVTV-1	102319	1	1	1	1	1	0	1	0	0
Cellulophaga phage phi14:2	100418	1	0	0	1	1	0	1	0	0
Gryllus bimaculatus nudivirus	96944	1	0	0	1	1	0	1	1	0
Planktothrix phage PaV-LD	95299	0	0	0	1	0	0	0	0	0
Enterobacteria phage P1	94800	1	0	0	1	0	1	0	0	0
Bacillus phage BCD7	93839	1	0	0	1	1	1	1	1	0
Pseudomonas phage PAK_P1	93398	1	0	0	1	1	0	1	1	0
Escherichia phage wV8	88487	1	0	0	1	1	0	1	1	0
Neodiprion abietis NPV	84264	1	0	0	1	1	1	1	0	0
Thermus phage P74-26	83319	1	0	0	1	1	0	1	0	0
Vibrio phage SIO-2	81184	1	0	0	1	1	1	1	0	0
Lactococcus phage KSY1	79232	1	0	0	1	1	0	1	1	0
Cellulophaga phage phiST	79114	1	0	0	1	1	0	1	1	0
Vibrio phage vB_VpaS_MAR10	78751	1	1	1	1	1	1	1	1	0
Mycobacterium phage Wildcat	78441	0	0	0	1	0	1	1	0	0
Enterobacteria phage PhiEco32	77554	1	0	0	1	1	0	1	1	0
Mycobacterium phage Cjw1	75931	0	0	0	1	0	1	1	0	0
Halovirus HF1	75898	1	0	0	1	1	0	1	0	0
Gordonia phage GTE7	74431	0	1	0	1	1	0	1	0	0
Sulfitobacter phage EE36phi1	73325	1	0	0	1	1	1	1	0	0
Cellulophaga phage phi38:1	72534	1	0	0	1	1	0	1	0	0
Burkholderia phage BcepF1	72415	1	0	0	1	1	0	0	0	0
Cellulophaga phage phi13:2	72369	0	0	0	0	0	1	0	0	0
Corynebacterium phage P1201	70579	1	0	0	1	1	0	1	0	0
Escherichia phage N4	70153	1	0	0	1	1	1	1	0	0
Mycobacterium phage Corndog	69777	0	1	0	1	1	0	1	0	0
Synechococcus phage S-CBS4	69420	0	0	0	1	1	0	1	0	0
Listeria phage P70	67170	1	0	0	1	1	0	1	0	0

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvuomo veiksnys	Žiedo užkėlėjas	Replikacinė helikazė	Praimazė	SSB	Pradmens pašalinimo baltymas	Ligazė	Topoizomerazė
Pseudomonas phage 14-1	66235	1	0	0	1	1	0	0	1	0
Pseudomonas phage F116	65195	0	0	0	0	0	1	0	0	0
Burkholderia phage Bcep22	63879	0	0	0	0	0	1	1	0	0
Phage phiJL001	63649	1	0	0	1	1	0	0	0	0
Acidianus two-tailed virus	62730	0	0	0	0	0	0	0	0	0
Stx2 converting phage II	62706	0	0	0	1	0	1	1	0	0
Azospirillum phage Cd	62337	0	0	0	0	0	0	0	0	0
Rhizobium phage 16-3	60195	0	1	0	1	1	1	1	0	0
Lactococcus phage P087	60074	1	0	0	1	0	0	0	0	0
Pseudomonas phage M6	59446	1	0	0	1	1	0	0	0	0
Clostridium phage phiCTP1	59199	1	1	1	1	1	1	1	1	0
Clavibacter phage CMP1	58652	0	0	0	1	1	1	1	0	0
Actinoplanes phage phiAsp2	58638	1	0	0	1	1	0	1	0	0
Natrialba phage PhiCh1	58498	0	1	0	0	0	0	0	0	0
Burkholderia phage AH2	58065	1	1	0	1	1	1	1	0	0
Mycobacteriophage Daenerys	58043	0	0	0	0	0	0	0	0	0
Mycobacterium phage Boomer	58037	0	0	0	0	0	0	0	0	0
Burkholderia phage BcepNazgul	57455	1	1	0	1	1	1	1	0	0
Sinorhizobium phage PBC5	57416	0	0	0	0	0	0	0	0	0
Mycobacterium phage Che9c	57050	0	0	0	0	0	0	1	0	0
Clostridium phage phiC2	56538	0	0	0	0	0	1	0	0	0
Pseudomonas phage D3	56425	0	0	0	1	0	0	1	0	0
Xanthomonas phage Xp15	55770	1	0	0	1	1	0	1	1	0
Lactococcus phage 1706	55597	1	0	0	1	1	1	1	0	0
Puniceispirillum phage HMO-2011	55282	1	0	0	1	1	0	1	0	0
Mycobacterium phage Giles	53746	0	0	0	0	0	0	1	0	0
Cellulophaga phage phi10:1	53664	0	0	0	1	1	1	1	0	0
Burkholderia phage phiE125	53373	0	0	0	0	0	0	0	0	0
Bacillus phage IEBH	53104	0	0	0	0	1	1	0	0	0
Enterobacteria phage phiEcoM-GJ1	52975	1	0	0	1	1	1	1	1	0
Mycobacterium phage TM4	52797	0	0	0	1	1	0	1	0	0
Halovirus HHTV-2	52643	0	0	0	0	0	0	0	0	0
Burkholderia phage BcepGomr	52414	1	1	0	1	1	1	1	0	0
Mycobacterium phage L5	52297	1	0	0	1	1	0	1	0	0
Streptomyces phage Sujidade	51552	1	0	0	1	1	0	1	0	0
Rhodococcus phage REQ1	51342	0	0	0	1	1	0	1	0	0
Pseudomonas phage PA11	49639	1	0	0	1	1	0	1	0	0

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvuomo veiksnys	Žiedo užkėlėjas	Replikacinė helikazė	Praimazė	SSB	Pradmens pašalinimo baltymas	Ligazė	Topoizomerazė
Myxococcus phage Mx8	49534	0	0	0	0	0	0	0	0	0
Bacillus phage Andromeda	49259	1	0	0	1	1	1	1	0	0
Streptomyces phage VWB	49220	0	0	0	0	1	0	1	0	0
Halovirus HHTV-1	49107	0	1	0	0	0	0	0	0	0
Enterobacteria phage T1	48836	0	0	0	1	1	1	1	0	0
Burkholderia phage Bcep781	48247	1	1	0	1	1	1	1	0	0
Listeria phage B054	48172	0	0	0	0	0	0	0	0	0
Clostridium phage phi8074-B1	47595	1	0	0	1	1	1	1	1	0
Iodobacteriophage phiPLPE	47453	0	0	0	0	0	0	1	0	0
Burkholderia phage BcepB1A	47399	1	0	0	1	1	1	1	0	0
Flavobacterium phage 6H	46978	0	0	0	1	1	0	0	0	0
Xanthomonas phage OP2	46643	1	0	0	1	1	1	1	0	0
Rhodococcus phage RER2	46586	1	0	0	1	1	0	1	0	0
Enterobacteria phage N15	46375	0	0	0	1	1	0	0	0	0
Microbacterium phage Min1	46365	0	0	0	0	0	1	1	0	0
Yersinia phage PY54	46339	0	0	0	1	1	0	0	0	0
Synechococcus phage Syn5	46214	1	0	0	1	1	1	1	0	0
Vibrio phage VpV262	46012	1	0	0	1	1	1	1	0	0
Bdellovibrio phage phi1422	45354	0	0	0	1	1	0	1	0	0
Prochlorococcus phage P-SSP7	45176	1	0	0	1	1	1	1	0	0
Sodalis phage SO-1	45169	1	1	0	1	1	1	1	0	0
Bacteroides phage B40-8	44929	0	0	0	0	0	1	1	0	0
Cellulophaga phage phiSM	44557	0	1	0	0	1	0	1	0	0
Xanthomonas phage Xop411	44520	1	0	0	1	1	0	1	1	0
Staphylococcus phage phiN315	44082	0	0	0	0	0	1	0	1	0
Staphylococcus phage PH15	44041	0	0	0	1	0	1	0	0	0
Bacillus phage SPP1	44010	0	0	0	1	0	1	1	0	0
Vibrio phage VP93	43931	1	0	0	1	1	0	1	0	0
Staphylococcus phage 53	43883	0	0	0	0	0	1	0	0	0
Fowl adenovirus A	43804	1	0	0	0	0	1	0	0	0
Salisaeta icosahedral phage 1	43788	0	0	0	1	1	0	0	0	0
Lactobacillus phage phiadh	43785	0	0	0	1	1	1	0	0	0
Enterobacteria phage SP6	43769	1	0	0	1	1	0	1	1	0
Pseudomonas phage 119X	43365	1	0	0	1	0	0	1	0	0
Phormidium phage Pf-WMP3	43249	1	0	0	1	1	0	1	0	0
Ralstonia phage RSB1	43079	1	0	0	1	1	0	1	1	0
Haemophilus phage Aaphi23	43033	0	0	0	1	0	0	1	0	0

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvuomo veiksnys	Žiedo užkėlėjas	Replikacinė helikazė	Praimazė	SSB	Pradmens pašalinimo baltymas	Ligazė	Topozimerazė
Salmonella phage epsilon34	43016	0	0	0	1	1	1	0	0	0
Pseudomonas phage 73	42999	1	1	0	1	1	1	1	0	0
Corynebacterium phage BFK20	42969	1	0	0	1	1	0	1	0	0
Streptococcus phage EJ-1	42935	0	0	0	0	0	0	1	0	0
Bordetella phage BIP-1	42638	1	0	0	1	1	1	1	0	0
Pseudomonas phage phiKMV	42519	1	0	0	1	1	0	1	1	0
Archaeal BJ1 virus	42271	0	0	0	1	0	0	0	0	0
Lactobacillus phage phig1e	42259	0	0	0	0	0	1	0	0	0
Thermoanaerobacterium phage THSA-485A	41938	0	0	0	1	1	0	0	0	0
Staphylococcus phage 55	41902	0	0	0	1	0	1	0	0	0
Mycobacterium phage Bps	41901	0	0	0	0	0	0	1	0	0
Enterobacteria phage P22	41724	0	0	0	1	0	0	0	0	0
Streptomyces phage phiC31	41491	1	0	0	1	1	0	1	0	0
Sulfitobacter phage pCB2047-A	41374	0	0	0	0	0	0	0	0	0
Brucella phage Tb	41148	0	0	0	1	1	1	0	0	0
Bacillus phage BCJA1c	41092	0	1	0	1	1	0	0	0	0
Clostridium phage phiCD38-2	41090	0	0	0	0	0	1	0	0	0
Phormidium phage Pf-WMP4	40938	1	0	0	1	1	1	1	0	0
Listeria phage A118	40834	0	0	0	0	0	1	1	0	0
Streptococcus phage P9	40539	1	0	0	1	1	1	1	0	0
Rhodobacter phage RcapNL	40489	0	0	0	1	1	0	0	0	0
Stenotrophomonas phage S1	40287	0	0	0	1	1	0	0	0	0
Lactobacillus phage Lrm1	39989	0	0	0	0	0	1	0	0	0
Roseobacter phage SIO1	39898	1	0	0	1	1	1	1	0	0
Vibrio phage VP5	39786	1	0	0	1	1	1	0	0	0
Acidianus filamentous virus 6	39577	0	0	0	1	0	0	1	0	0
Bacillus phage phi105	39325	0	0	0	1	1	1	0	0	0
Cellulophaga phage phi18:1	39189	0	0	0	1	1	1	0	0	0
Tetraselmis viridis virus S20	38987	0	0	0	1	0	0	1	0	0
Lactobacillus phage Lv-1	38934	0	0	0	0	0	1	0	0	0
Cyanophage MED4-117	38834	0	0	0	0	0	0	1	0	0
Enterobacteria phage mEpX2	38759	0	0	0	1	0	1	0	0	0
Clostridium phage phiCP39-O	38753	0	0	0	1	1	1	0	0	0
Salmonella phage SPN1S	38684	0	0	0	0	0	0	1	0	0
Streptococcus phage phi3396	38528	0	0	0	0	0	1	0	0	0
Lactobacillus prophage Lj928	38384	0	0	0	0	0	0	0	0	0
Enterobacteria phage T3	38208	1	0	0	1	1	1	1	1	0

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvuomo veiksnys	Žiedo užkėlėjas	Replikacinė helikazė	Praimazė	SSB	Pradmens pašalinimo baltymas	Ligazė	Topoizomerazė
Streptomyces phage mu1/6	38194	0	0	0	1	0	0	1	0	0
Clostridium phage phiSM101	38092	0	0	0	0	0	0	0	0	0
Listonella phage phiHSIC	37966	0	1	0	1	1	0	1	0	0
Paenibacillus phage PG1	37644	0	0	0	0	0	1	1	0	0
Burkholderia phage phi52237	37639	0	0	0	1	1	0	0	0	0
Burkholderia phage KS10	37635	0	0	0	0	0	0	0	0	0
Bacillus phage TP21-L	37456	0	0	0	1	0	0	0	0	0
Burkholderia phage phiE255	37446	0	0	0	0	0	0	0	0	0
Streptococcus phage Sfi19	37370	0	0	0	1	1	1	0	0	0
Thalassomonas phage BA3	37313	0	0	0	0	0	0	1	0	0
Bacillus phage Gamma	37253	0	0	0	0	0	0	0	0	0
Escherichia phage D108	37235	0	0	0	0	0	0	0	0	0
Streptomyces phage phiSASD1	37068	1	0	0	1	1	0	1	0	0
Pseudomonas phage MP38	36885	0	0	0	0	0	0	0	0	0
Enterobacteria phage Mu	36717	0	0	0	0	0	0	0	0	0
Lactobacillus phage phiJL-1	36674	0	0	0	1	1	0	1	0	0
Streptococcus phage SMP	36019	0	0	0	1	1	1	0	0	0
Flavobacterium phage 11b	36012	0	0	0	1	0	1	0	0	0
Halovirus HRTV-4	35722	0	0	0	0	0	0	0	0	0
Persicivirga phage P12024S	35700	0	0	0	1	1	1	1	0	0
Listeria phage P40	35638	1	0	0	1	1	0	0	0	0
Sulfolobus islandicus rod-shaped virus 2	35450	0	0	0	0	0	0	1	0	0
Human adenovirus D	35083	1	0	0	0	0	1	0	0	0
Abalone shriveling syndrome-associated virus	34952	0	0	0	1	1	1	1	0	0
Pelagibacter phage HTVC010P	34892	0	0	0	0	0	0	1	0	0
Cellulophaga phage phi46:1	34844	0	0	0	1	1	0	1	0	0
Streptococcus phage SM1	34692	0	0	0	0	0	0	0	0	0
Geobacillus phage GBSV1	34683	0	0	0	0	0	0	1	0	0
Lactobacillus phage LL-H	34659	0	0	0	0	0	1	0	0	0
Mannheimia phage phiMHaA1	34525	0	0	0	0	0	1	0	1	0
Weissella phage phiYS61	33594	1	0	0	0	0	0	0	0	0
Enterobacteria phage P2	33593	0	0	0	0	0	0	0	0	0
Clostridium phage phi3626	33507	0	0	0	1	0	1	0	0	0
Vibrio phage K139	33106	0	0	0	0	0	0	0	1	0
Haemophilus phage HP1	32355	0	0	0	0	0	0	0	0	0
Acidianus filamentous virus 2	31787	0	0	0	0	0	0	0	0	0
Lactococcus phage bIL170	31754	0	0	0	1	0	1	0	0	0

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvuomo veiksnys	Žiedo užkėlėjas	Replikacinė helikazė	Praimazė	SSB	Pradmens pašalinimo baltymas	Ligazė	Topoizomerazė
Haloarcula phage SH1	30889	0	0	0	0	0	0	0	0	0
Enterococcus phage EF62phi	30505	1	0	0	0	0	0	0	0	0
Propionibacterium phage PA6	29739	0	0	0	1	1	0	1	0	0
Cellulophaga phage phi39:1	28760	0	0	0	1	0	0	0	0	0
Pyrobaculum spherical virus	28337	0	0	0	0	0	0	0	0	0
Snake adenovirus 1	27751	1	0	0	0	0	1	0	0	0
Lactococcus phage Q54	26537	0	0	0	0	0	1	0	0	0
Tetraselmis viridis virus S1	26407	0	0	0	1	0	0	0	0	0
Helicobacter phage KHP30	26215	0	0	0	1	1	0	0	0	0
Leuconostoc phage Lmd1	26201	1	0	0	1	1	0	0	0	0
Frog adenovirus 1	26163	1	0	0	0	0	1	0	0	0
Methanobacterium phage psiM2	26111	0	0	0	0	0	0	0	0	0
Bdellovibrio phage phi1402	23931	0	0	0	0	0	0	0	0	0
Acidianus bottle-shaped virus	23814	1	0	0	0	0	0	0	0	0
Hyperthermophilic Archaeal Virus 1	22743	0	0	0	0	0	0	1	0	0
Lactococcus phage bIL67	22195	0	0	0	1	0	1	0	0	0
Thermococcus prieurii virus 1	21592	0	0	0	1	0	0	0	0	0
Bacillus phage GA-1	21129	1	0	0	0	0	1	0	0	0
Enterococcus phage EFAP-1	21115	0	0	0	0	0	0	0	0	0
Thermoproteus tenax spherical virus 1	20933	0	0	0	0	0	0	0	0	0
Acidianus filamentous virus 1	20869	0	0	0	0	0	0	1	0	0
Phaeocystis globosa virus virophage	19527	1	0	0	1	0	0	0	0	0
Streptococcus phage Cp-1	19343	1	0	0	0	0	0	0	0	0
Bacillus phage phi29	19282	1	0	0	0	0	1	0	0	0
Mavirus	19063	1	0	0	1	0	0	0	0	0
Lactococcus phage ascphi28	18762	1	0	0	0	0	1	0	0	0
Bacillus phage B103	18630	1	0	0	0	0	1	0	0	0
Sputnik virophage	18343	1	0	0	1	0	0	0	0	0
Staphylococcus phage P68	18227	1	0	0	0	0	0	0	0	0
Pyrococcus abyssi virus 1	18098	0	0	0	0	0	0	0	0	0
Hyperthermophilic Archaeal Virus 2	17666	0	0	0	1	1	0	0	0	0
Actinomyces phage Av-1	17171	1	0	0	0	0	1	0	0	0
Thermus phage P23-77	17036	0	0	0	0	0	0	0	0	0
Streptococcus phage C1	16687	1	0	0	0	0	0	0	0	0
Sulfolobus turreted icosahedral virus 2	16622	0	0	0	0	0	0	0	0	0
His2 virus	16067	1	0	0	0	0	0	0	0	0
Mycoplasma phage MAV1	15644	0	0	0	1	0	0	0	0	0

Virusas	Genomo dydis (bp)	DNR polimerazė	Procesyvumo veiksnys	Žiedo užkėlėjas	Replikacinė helikazė	Praimazė	SSB	Pradmens pašalinimo baltymas	Ligazė	Topoizomerazė
Staphylococcus phage PT1028	15603	0	0	0	1	1	0	0	0	0
Lactococcus phage bIL312	15179	0	0	0	1	0	0	0	0	0
Sulfolobus spindle-shaped virus 4	15135	0	0	0	0	0	0	1	0	0
Enterobacteria phage PRD1	14927	1	0	0	0	0	1	0	0	0
Bacillus phage GIL16c	14844	1	0	0	0	0	1	0	0	0
His 1 virus	14462	1	0	0	0	0	0	0	0	0
Acholeplasma phage L2	11965	0	0	0	0	0	0	0	0	0
Mycoplasma phage P1	11660	1	0	0	0	0	0	0	0	0
Enterobacteria phage P4	11624	0	0	0	1	1	0	0	0	0
Pseudoalteromonas phage PM2	10079	0	0	0	0	0	0	0	0	0
Deltapapillomavirus 2	8374	0	0	0	1	0	0	0	0	0
Haloarcula hispanica pleomorphic virus 1	8082	0	0	0	0	0	0	0	0	0
Human papillomavirus type 41	7614	0	0	0	1	0	0	0	0	0
Thetapapillomavirus 1	7304	0	0	0	1	0	0	0	0	0
Human papillomavirus type 48	7100	0	0	0	1	0	0	0	0	0
Murine polyomavirus	5297	0	0	0	1	0	0	0	0	0
WU Polyomavirus	5229	0	0	0	1	0	0	0	0	0
Bovine polyomavirus	4697	0	0	0	1	0	0	0	0	0