



Latvijas Universitāte

Bioloģijas fakultāte

**PROTEOBAKTĒRIJU NEKLASISKI
SEKRETĒJAMO OLBALTUMVIELU SEKVENČU
DISKRIMINATĪVĀS PAZĪMES**

**DISCRIMINATIVE FEATURES AMONG THE SEQUENCES OF NON-
CLASSICALLY SECRETED PROTEINS FROM PROTEOBACTERIA**

Promocijas darbs

Autors: Ināra Kampenusa (dz.Andersone)

Darba vadītājs: Dr.habil.biol. Pēteris Zikmanis

Dr.biol. Olga Mutere

Doktora zinātniskā grāda iegūšanai bioloģijā

Apakšnozare: Mikrobioloģija

Rīga 2010

Mg. biol. Ināras Kampenusa dz. Andersones promocijas darbs “**Proteobaktēriju neklasiski sekretējamo olbaltumvielu sekvenču diskriminatīvās pazīmes**” izstrādāts Latvijas Universitātes Mikrobioloģijas un Biotehnoloģijas institūtā (LU MBI) laika posmā no 2006. līdz 2010. gadam.

Zinātniskais vadītājs: Pēteris Zikmanis, Dr. habil. biol., vadošais pētnieks (LU Mikrobioloģijas un Biotehnoloģijas institūts (LU MBI), Rīga, Latvija) un Olga Mutere, Dr. biol., vadošais pētnieks (LU Mikrobioloģijas un Biotehnoloģijas institūts (LU MBI), Rīga, Latvija).

Promocijas darba recenzenti

Dr. habil. biol. Pauls Pumpēns (Latvijas Biomedicīnas pētījumu un studiju centrs)

Dr. biol. Nīls Rostoks (LU Bioloģijas fakultāte)

Dr. biol. Kaspars Tārs (Latvijas Biomedicīnas pētījumu un studiju centrs)

Promocijas darba aizstāvēšana notiks 2011. gada 30. martā LU Bioloģijas fakultātes nozares promocijas padomes atklātā sēdē LU Bioloģijas fakultātē, Rīgā, Kronvalda b. 4, __. auditorijā.

Promocijas padomes sastāvs

Dr. habil. biol. Pauls Pumpēns (padomes priekšsēdētājs)

Dr. biol. Uldis Kalnenieks

Dr. habil. biol. Indriķis Muižnieks

Dr. habil. biol. Aleksandrs Rapoportis

Dr. sc. comp. Juris Vīksna

Dr. habil. biol. Henriks Zenkevičs

Promocijas darba izstrāde ir **saistīta ar dalību** (2006.-2009.) **LZP projektā** Nr. 04.1099. „Levānsaharāzes sekrēcijas fizioloģiskie nosacījumi baktērijās *Zymomonas mobilis*”.

Promocijas darba izstrādāšanai **saņemts Eiropas Sociālā Fonda projekta** Nr. 2009/0138/1DP/1.1.2.1.2/09IPI/VIAA/004 "Atbalsts doktora studijām Latvijas Universitātē" atbalsts (līgums Nr. 09/142/38).

Ar promocijas darbu var iepazīties LU Bibliotēkā, Kalpaka b. 4, Rīgā.

Atsauksmes, lūdzu, sūtīt: Daina Eze, Bioloģijas zinātņu promocijas padomes zinātniskā sekretāre, Bioloģijas fakultāte, Latvijas Universitāte, Kronvalda b. 4, Rīga, Latvija, LV-1586, E-pasts: daina.eze@lu.lv

Anotācija

Promocijas darbā atrastas un raksturotas no proteobaktērijām neklasiski sekretējamu (I, III, IV un VI tips) olbaltumvielu sekvenču un to terminālo fragmentu strukturālās pazīmes, no kurām diskriminantanalīzē izdalītas atsevišķas raksturīgo pazīmju (efektīvie mainīgie) kombinācijas un izveidotas statistiski būtiskas diskriminantfunkcijas, kuru precizitāte ļauj klasificēt sekrēcijas substrātus un prognozēt to atbilstību noteiktam olbaltumvielu translokācijas mehānismam.

Promocijas darba rezultāti publicēti četrās zinātniskajās publikācijās un triju starptautisku konferenču materiālos. Promocijas darbs izstrādāts 2006. – 2010. gadā Latvijas Universitātes Mikrobioloģijas un Biotehnoloģijas institūtā (LU MBI) Dr.habil.biol. Pētera Zikmaņa vadībā.

Atslēgvārdi: Olbaltumvielu sekrēcija, Kodonu izlietojums, Aminokābju sastāvs, Diskriminantanalīze, Proteobaktērijas

Abstract

The promotion work deals with the comparison of the sequence structural properties between the groups of non-classically secreted proteins (types I, III, IV, VI) from proteobacteria. The selected combinations of characteristics (effective variables) maximize the differences observed within the whole and terminal sequences of proteins and the developed significant discriminant functions appear to be reliable for the classification and affiliation of the secretion substrates to a definite type of protein translocation.

The results of the promotion work are published in four research papers and three international conference abstract books. The promotion research was carried out between 2006-2010 at the Institute of Microbiology and Biotechnology, the University of Latvia (LU MBI) under supervision of Dr.habil.biol. Pēteris Zikmanis.

Key words: Protein secretion, Codon usage, Amino acid composition, Discriminant analysis, Proteobacteria

Saīsinājumi

CAI – kodonu adaptācijas indekss (*codon adaptation index*)

KCV – k -reizes atkārtotā vairākkopu krosvalidācija (*k-fold crossvalidation*),

LOOCV – k -reizes atkārtotā vairākkopu krosvalidācija, kuras gadījumā k ir vienāds ar kopējo paraugu skaitu oriģinālajā datu kopā (*leave-one-out cross validation*)

N_c – efektīvo kodonu skaits (*number of effective codons*)

Q – Press's Q indekss

T1SS – 1. tipa sekrēcijas sistēma (*type 1 secretion pathway*)

T2SS – 2. tipa sekrēcijas sistēma (*type 2 secretion pathway*)

T3SS – 3. tipa sekrēcijas sistēma (*type 3 secretion pathway*)

T4SS – 4. tipa sekrēcijas sistēma (*type 4 secretion pathway*)

T5SS – 5. tipa sekrēcijas sistēma (*type 5 secretion pathway*)

T6SS – 6. tipa sekrēcijas sistēma (*type 6 secretion pathway*)

T7SS – 7. tipa sekrēcijas sistēma (*type 7 secretion pathway*)

VK – variācijas koeficients

Saturs

Ievads.....	9
Zinātniskā darba aprobācija.....	11
1. Literatūras apskats.....	14
1.1. Olbaltumvielu transportfunkcijas daudzveidība un loma	14
1.2. Olbaltumvielu sekrēcijas sistēmas proteobaktērijās.....	15
1.2.1. Pārskats par I tipa sekrēcijas sistēmu komponentu un substrātu īpatnībām.....	15
1.2.2. Pārskats par III tipa sekrēcijas sistēmu komponentu un substrātu īpatnībām....	16
1.2.3. Pārskats par IV tipa sekrēcijas sistēmu komponentu un substrātu īpatnībām. .	17
1.2.4. Pārskats par VI tipa sekrēcijas sistēmu komponentu un substrātu īpatnībām. .	18
1.3. Olbaltumvielu sekrēcijas sistēmu izplatība un daudzveidība	18
1.4. Pārskats par sekrēcijas signāliem un to nozīmi.....	19
1.5. Olbaltumvielu sekrēcijas regulācija	20
1.6. Sekretējamo olbaltumvielu biotehnoloģiskais potenciāls.....	21
1.7. Sekretējamo olbaltumvielu eksperimentālas identificēšanas un raksturošanas metodika un prognozēšanas metožu izstrāde.....	22
1.8. Olbaltumvielu sekrēcijas norišu jēdzieniskā daudzveidība un nomenklatūras problēmas	26
2. Metodes.....	30
1.1. Izstrādes kopas olbaltumvielu sekvenses	30
2.2. Neatkarīgas paraugkopas olbaltumvielu sekvenses	31
2.3. Aminoskābju frekvenču, kodonu izlietojuma un nukleotīdu satura un sekvenču kompozicionālo raksturlielumu aprēķināšana.....	31
2.4. Aminoskābju izkārtojuma periodisko īpašību un otrējās struktūras elementu sadalījuma prognozēšana.....	32
2.5. Svešas izcelsmes gēnu olbaltumvielu sekvenču noteikšana.....	33
2.6. Sekrēcijas substrātu grupu īpašību salīdzināšana.....	34
2.7. Diskrimantanalīze.....	34
2.8. Datorprogrammas.....	35
3. Rezultāti.....	36
3.1. Vispārējo sekvenču raksturlielumu atšķirības neklasiskās sekrēcijas substrātos.....	36

3.2. Aminoskābju un kodonu izlietojuma diskriminatīvā potenciāla salīdzinājums.....	37
3.3. Kodonu un aminoskābju izlietojuma variācija un informācijas entropija.....	39
3.4. Efektīvo kodonu un aminoskābju atlase 4-grupu diskriminantanalīzei.....	40
3.5. Neatkarīga olbaltumvielu sekvenču kopa.....	45
3.6. Efektīvo kodonu un aminoskābju kombināciju atlase lineārai grupu pāru diskriminantanalīzei.....	45
3.7. Efektīvo kodonu un aminoskābju kombināciju atlase pāru diskriminantanalīzē individuālu substrātu grupu atšķiršanai no apvienotās pārējo trīs sekrēcijas tipu substrātu kopas.....	47
3.8. Efektīvo kodonu un aminoskābju saistība ar olbaltumvielu vispārējām, t.sk. otrējās struktūras īpašībām	49
3.9. Sekrēcijas substrātu pārstāvība proteobaktēriju subklasēs.....	52
3.10. Svešas izcelsmes gēni.....	53
3.11. Patiesi un kļūdaini klasificēto sekvenču īpašību atšķirības.....	55
3.12. Piemērs baktēriju eksoproteoma precizēšanai ar iegūtajām diskriminantfunkcijām...	55
3.13. Sekvenču terminālo fragmentu salīdzināšana.....	56
 4. Diskusija.....	 60
 Secinājumi.....	 71
 Pateicības.....	 72
 Literatūras saraksts.....	 73
 Pielikums.....	 89

Pielikumu saraksts

1. Pielikums

Publikācija I: Zikmanis P., Andersone I., Baltkalne M. 2006. Discriminative features of type I and type III secreted proteins from Gram-negative bacteria. – Central European Journal of Biology, 1(1):124-136

2. Pielikums

Publikācija II: Andersone I., Zikmanis P. 2007. Distinctive amino acid residue periodicities in terminal sequences of type III and type I secreted proteins from proteobacteria. – Central European Journal of Biology, 2(2):192-205.

3. Pielikums

Publikācija III: Kampenusa I., Zikmanis P. 2008. Distinctive attributes for predicted secondary structures at terminal sequences of non-classically secreted proteins from proteobacteria. – Central European Journal of Biology, 3(3):320-326.

4. Pielikums

Publikācija IV: Kampenusa I., Zikmanis P. 2010. Distinguishable codon usage and amino acid composition patterns among substrates of leaderless secretory pathways from proteobacteria. – Applied Microbiology and Biotechnology, 86(1):285-293.

Ievads

Āršūnas olbaltumvielu ievērojamais īpatsvars baktēriju kopējā proteomā, to nozīme šūnu dzīvības procesu nodrošināšanā, arī daudzpusīgās tehnoloģiskā pielietojuma iespējas nosaka nepieciešamību pēc efektīvākām to identificēšanas un raksturošanas metodēm. Tādēļ ir svarīgi padziļināt zināšanas un attīstīt priekšstatus par baktēriju sekrēcijas sistēmām, īpašu vērību piegriežot to substrātu īpašībām. Tā, proteobaktērijām šobrīd pazīstamas vismaz sešu tipu sekrēcijas sistēmas, no kurām četras nav raksturojamas ar N- terminālā signālpeptīda atšķelšanu no sekrēcijas substrāta, tādēļ visbiežāk tiek apzīmētas kā 'neklasiskās' sekrēcijas sistēmas. Jāuzsver, ka olbaltumvielas ar eksperimentāli apstiprinātu atbilstību kādam noteiktam sekrēcijas mehānismam joprojām pārstāv tikai niecīgu daļu no nepārtraukti pieaugošā sekvencēto eksoproteomu apjoma. To zināmā mērā kavē daudzveidīgās un savstarpēji atšķirīgās, taču vienlaikus arī līdzīgās sekrēcijas kompleksu īpašības.

Olbaltumvielu, t.sk. baktēriju sekrēcijas substrātu, raksturošanai līdztekus eksperimentālajām pieejām plaši izmanto dažādas *in silico* prognozēšanas metodes, kuras, lielākoties, ļauj noteikt vienīgi olbaltumvielu lokalizāciju šūnā un, raksturojot sekrēcijas substrātus, un apstiprināt vai noliegt vispārējam sekrēcijas ceļam raksturīgā N- terminālā signālpeptīda klātbūtni sekretējamā olbaltumvielā, kas nav pietiekami, lai sniegtu norādi uz tām atbilstošu sekrēcijas tipu. Turklāt šobrīd izmantojamo precizējošāku metožu izstrāde pamatojas uz visai šauru organismu spektru un / vai aptver tikai nedaudzas olbaltumvielu funkcionālās klases, un tādējādi ir piemērotas tikai atsevišķa neklasiskās sekrēcijas tipa (galvenokārt T3SS) sekrēcijas substrātu, t.sk. efektoru olbaltumvielu atpazīšanai. Tāpat nav pamata pārvērtēt substrātu identifikācijas un raksturošanas iespējas, izmantojot vispārējo sekvenču salīdzināšanas (BLASTP/HAMAP) metodi. Tādēļ jāuzskata, ka neklasiskās sekrēcijas substrātu prognozēšanas metožu izstrāde, atvieglotu šādu sekrēcijas sistēmu turpmāku eksperimentālu identificēšanu, kā arī ļautu novērtēt atsevišķu tipu sekrēcijas sistēmu īpatsvaru dažādās proteobaktēriju kultūrās, tā skaidrojot arī to tehnoloģiskās izmantošanas perspektīvas.

Lai arī pētījumi par sekvenču īpašību izmantošanu olbaltumvielu klasifikācijai un to piederības prognozēšanai literatūrā ir pārstāvēti salīdzinoši plaši, tomēr joprojām nav rodamas atbildes, vai neklasiskās sekrēcijas substrātu grupas ir raksturojamas un / vai atšķiramas pēc to vispārējām sekvenču īpašībām un vai tās ietver kādam noteiktam neklasiskās sekrēcijas tipam specifisku pazīmju kopumu. Tādēļ šī promocijas darba **mērķis** ir salīdzinoši izvērtēt bez

N- signālpeptīda līdzdalības sekretējamo olbaltumvielu sekvenču un to terminālo fragmentu pazīmes, atlasīt no tām efektīvo pazīmju kopas, novērtēt šo kopu piemērotību sekrēcijas substrātu klasifikācijai un prognozēt to atbilstību kādai no proteobaktēriju I, III, IV vai VI tipa sekrēcijas sistēmām.

Darba mērķa sasniegšanai izvirzīti šādi **uzdevumi**:

1. Raksturot aminoskābju satura un kodonu izlietojuma atšķirības proteobaktēriju neklasiski sekretējamo olbaltumvielu grupās.

2. Novērtēt prognozēto otrējās struktūras elementu, kā arī ar kopīgām īpašībām raksturojamu aminoskābju grupu izvietojuma un sastāva atšķirības neklasiskās sekrēcijas substrātu sekvenču terminālajos fragmentos.

3. Izmantojot sekvenču vispārējās un struktūru raksturojošās īpašības, atlasīt efektīvas neatkarīgo mainīgo kombinācijas un statistiski novērtēt atbilstošo diskriminantfunkciju efektivitāti.

4. Pārbaudīt izvēlēto neatkarīgo mainīgo un tām atbilstošo diskriminantfunkciju savstarpējo atbilstību, savietojamību un precizitāti, klasificējot un prognozējot proteobaktēriju neklasiskās sekrēcijas substrātu piederību noteiktai sekrēcijas sistēmai.

5. Pārbaudīt izvēlēto diskriminantfunkciju iespējas prognozēt neatkarīgas anotētu olbaltumvielu sekvenču paraugkopas sadalījumu atbilstoši šobrīd pazīstamajām proteobaktēriju neklasiskās sekrēcijas sistēmām.

Promocijas darba **novitāte un zinātniskais nozīmīgums**

1. Pirmo reizi apstiprināta iespēja izmantot aminoskābju un kodonu izlietojuma frekvences kā raksturīgas pazīmes neklasiskās sekrēcijas substrātu klasifikācijai.

2. Parādīts, ka no aminoskābju un kodonu izlietojuma frekvenču vektoriem izvēlētas neatkarīgo mainīgo kopas dod iespēju ar augstu precizitāti klasificēt neklasiskās sekrēcijas substrātus un prognozēt to atbilstību noteiktai proteobaktēriju sekrēcijas sistēmai.

3. Atrasts, ka prognozēto otrējās struktūras elementu un vairāku grupu aminoskābju izvietojums sekvenču C- terminālajos fragmentos veido sekrēcijas substrātu grupām specifiskas pazīmju kopas, kuras ar pietiekamu precizitāti ļauj klasificēt sekrēcijas substrātu grupas, tomēr nesasniedzot, ar sekvenču vispārējiem raksturlielumiem iegūto, izšķirtspēju.

4. Pamatojoties uz multiplās diskriminantanalīzes funkciju statistisko izvērtējumu, pirmo reizi izveidotas praktiski pielietojamas diskriminantfunkcijas, kuras ļauj klasificēt proteobaktēriju ekstracellulārās olbaltumvielas pēc to sekvenču īpašībām un prognozēt to piederību kādam no šobrīd zināmajiem neklasiskās sekrēcijas tipiem.

5. Apstiprinātas izveidoto klasifikatoru pielietojuma iespējas, novērtējot filoģenētiski un funkcionāli atšķirīgu olbaltumvielu sekrēcijas norises.

Zinātniskā darba aprobācija

Promocijas darba rezultāti ir apkopoti un publicēti četrās zinātniskajās publikācijās starptautiski recenzējamās ISI žurnālos.

1. **Kampenusa I., Zikmanis P. 2010.** Distinguishable codon usage and amino acid composition patterns among substrates of leaderless secretory pathways from proteobacteria. – Applied Microbiology and Biotechnology, 86(1):285-293.

2. **Kampenusa I., Zikmanis P. 2008.** Distinctive attributes for predicted secondary structures at terminal sequences of non-classically secreted proteins from proteobacteria. – Central European Journal of Biology, 3(3):320-326.

3. **Andersone I., Zikmanis P. 2007.** Distinctive amino acid residue periodicities in terminal sequences of type III and type I secreted proteins from proteobacteria. – Central European Journal of Biology, 2(2):192-205.

4. Zikmanis P., **Andersone I., Baltkalne M. 2006.** Discriminative features of type I and type III secreted proteins from Gram-negative bacteria. – Central European Journal of Biology, 1(1):124-136.

Par promocijas darba rezultātiem ir ziņots trijās starptautiskajās zinātniskajās konferencēs.

1. **Kampenusa I., Zikmanis P. 2010.** Discriminant functions for classification and prediction of T1SS, T3SS, T4SS and T6SS secreted proteins from proteobacteria. – 9th European Conference on Computational Biology (ECCB), Ghent, Belgium, Sep 26-29, 2010: D2 (*stenda referāts*).

2. **Kampenusa I., Zikmanis P. 2009.** Discriminative features of non-classically secreted proteins from Proteobacteria. – 3rd Congress of European Microbiologists – FEMS, Gothenburg, Sweden, Jun 28-Jul 2, 2009: A-151-0016-00282 (*stenda referāts*).

3. **Kampenusa I., Zikmanis P. 2008.** Distinctive compositional age of non-classically secreted proteins from proteobacteria. – Mini EURO conference on Computational Biology, Bioinformatics and Medicine (CCBBM), Rome, Italy, Sep 15-17, 2008: M4B.3 (*mutisks ziņojums*).

Par promocijas darba rezultātiem ir ziņots četrās Latvijas Universitātes konferencēs.

1. **Kampenusa I., Zikmanis P. 2010.** Diskriminatīvo pazīmju novērtēšana proteobaktēriju neklasiski sekretējamo olbaltumvielu sekvencēs. LU 68. zinātniskā konference, Mikrobioloģijas un Biotehnoloģijas sekcija. Latvijas Universitāte, Rīga, 10. februārī (*mutisks ziņojums*).

2. **Kampenusa I., Zikmanis P. 2008.** Otrējās struktūras pazīmju atšķirības proteobaktēriju ‘neklasiski’ sekretējamo olbaltumvielu sekvencēs. LU 66. zinātniskā konference, Mikrobioloģijas un Biotehnoloģijas sekcija. Latvijas Universitāte, Rīga, 7. februārī (*mutisks ziņojums*).

3. **Andersone I., Zikmanis P. 2007.** Aminokābju periodiskā izvietojuma atšķirības proteobaktēriju ‘neklasiski’ sekretējamo olbaltumvielu sekvencēs. LU 65. zinātniskā konference, Mikrobioloģijas un Biotehnoloģijas sekcija. Latvijas Universitāte, Rīga, 15. februārī (*mutisks ziņojums*).

4. **Andersone I., Baltkalne M., Zikmanis P. 2006.** Gram-negatīvo baktēriju atšķirīgi sekretējamo olbaltumvielu diskriminatīvās pazīmes. LU 64. zinātniskā konference,

Mikrobioloģijas un Biotehnoloģijas sekcija. Latvijas Universitāte, Rīga, 7. februārī (*mutisks ziņojums*).

Aizstāvēšanai izvirzītās tēzes

1. Neatkarīgi no organismu filoģenētiskās piederības un olbaltumvielu funkcionālajām īpašībām proteobaktēriju neklasiskās sekrēcijas substrātu aminoskābju un kodējošo sekvenču raksturīgo pazīmju atšķirīgo sadalījumu nosaka olbaltumvielu atbilstība I, III, IV vai VI sekrēcijas tipam.

2. Neklasiski sekretējamo olbaltumvielu un tās kodējošo sekvenču kvantitatīvās pazīmes atspoguļo fizikāli-ķīmisko īpašību un sekvenču ģenēzes atšķirības, tostarp horizontālās gēnu pārnese izpausmes, dažādās no signālpeptīdiem neatkarīgu sekrēcijas sistēmu substrātu grupās.

3. Atbilstoši proteobaktērijās definētajiem sekrēcijas tipiem no ekstracellulāro olbaltumvielu sekvenču kvantitatīvo pazīmju kopuma var izdalīt atsevišķas efektīvo mainīgo grupas, kuru veidotās diskriminantfunkcijas ļauj klasificēt un prognozēt sekrēcijas substrātu atbilstību noteiktam olbaltumvielu translokācijas mehānismam, sasniedzot 97% precizitāti.

1. Literatūras apskats

1.1. Olbaltumvielu transportfunkcijas daudzveidība un loma

Āršūnas olbaltumvielām ir ievērojama loma mikroorganismu dzīvības procesos. Tās piedalās šūnas sienīņu izveidē un barības vielu uzņemšanā, kā arī nodrošina šūnu kustīgumu, to adhēzijas spēju un starpšūnu komunikāciju. Turklāt baktēriju piesaistīšanās biotiskām un abiotiskām virsmām vienlaikus ar olbaltumvielu (toksīni un efektorās olbaltumvielas (*effector*)) sekrēciju ir būtisks saimniekšūnu organismu kolonizēšanas norises priekšnosacījums (Gerlach un Hensel 2007). Olbaltumvielu sekrēciju vai tiešu to tiešu nogādi (*injection*) saimniekorganismu šūnās realizē vairākas sekrēcijas sistēmas, kuru evolucionārajā attīstībā, darbības mehānismā un nepieciešamībā pēc enerģētiskā nodrošinājuma ir daudz kopīgu iezīmju (Holland 2004; Saier 2006; DeBuck *et al.* 2007). Atšķirībā no eikariotiem raksturīgajām membrānu transportieru grupām, kas lielākoties ir sugām specifiskas, prokariotu transportsistēmu komponenti baktērijās ir plaši izplatīti un tiešā veidā neatspoguļo organismu filoģenētisko piederību (Ren un Paulsen 2005). Tomēr transportsistēmas specifiskums attiecībā pret noteiktu substrātu ir uzskatāms par evolucionārās attīstības priekšnoteikumu (*evolutionary trait*), jo sekrēcijas substrātu un transportsistēmu evolūcija vispārējā gadījumā ir virzīta uz optimālu savstarpējo saderību (Economou 2002). To apstiprina novērojumi, ka membrānu transportieri ar līdzīgām funkcijām filoģenētiskajās analīzēs parasti izvietojas vienkopus (Ren un Paulsen 2005). Transportsistēmu daudzveidība vienas ģints baktēriju sugām parasti, ar atsevišķiem izņēmumiem, var atspoguļot vides apstākļu atšķirības. Vienšūņiem, t.sk. baktērijām, lielāks skaits transportsistēmu ļauj aptvert plašu spektru iespējamo substrātu, kamēr daudzšūnu organismiem raksturīgs relatīvi lielāks transportsistēmu paralogu apjoms (Ren un Paulsen 2005). Turklāt fundamentālas atšķirības nepieciešamībā pēc olbaltumvielu transportfunkciju intensitātes atsevišķu baktēriju sugu ietvaros atspoguļojas to genomā, kur viens genoms var kodēt kādas transportsistēmas paralogas kopijas un / vai vairāku sekrēcijas sistēmu komponentu gēnus (Gerlach un Hensel 2007; Tseng *et al.* 2009). Daļu no zīdītājiem patogēnām baktērijām pārnēsā arī protozoji, kukaiņi, u.c. bezmugurkaulnieki. Tādējādi virkni šūnas procesu (t.sk. olbaltumvielu sekrēciju), kurus šobrīd pirmām kārtām attiecina uz baktēriju virulenci, var precīzāk raksturot, kā līdzekļus baktēriju dzīvotspējas (*survival*) nodrošināšanai atšķirīgos ārvides apstākļos un saimniekorganismos (Flintoft 2009, Bleasdale *et al.* 2009). Ievērojamu daļu sekrēcijas substrātu attiecina uz t.s.

efektorām olbaltumvielām, kuras tiek ievadītas saimniekorganismā, lai izmainītu tā šūnas(-u) morfoloģiskās, fizioloģiskās, bioķīmiskās vai ģenētiskās īpašības (Chibucos *et al.* 2009). Ir pierādīta efektoru olbaltumvielu koevolucionēšana līdz ar saimniekorganismu rezistences mehānismu pilnveidošanos (Alfano un Collmer 2004), un nereti to sekvenču satur eikariotiskas izcelsmes olbaltumvielām raksturīgus motīvus, kas nodrošina efektoru olbaltumvielu iesaistīšanos saimniekorganisma šūnu funkciju regulēšanā (Boyer *et al.* 2009).

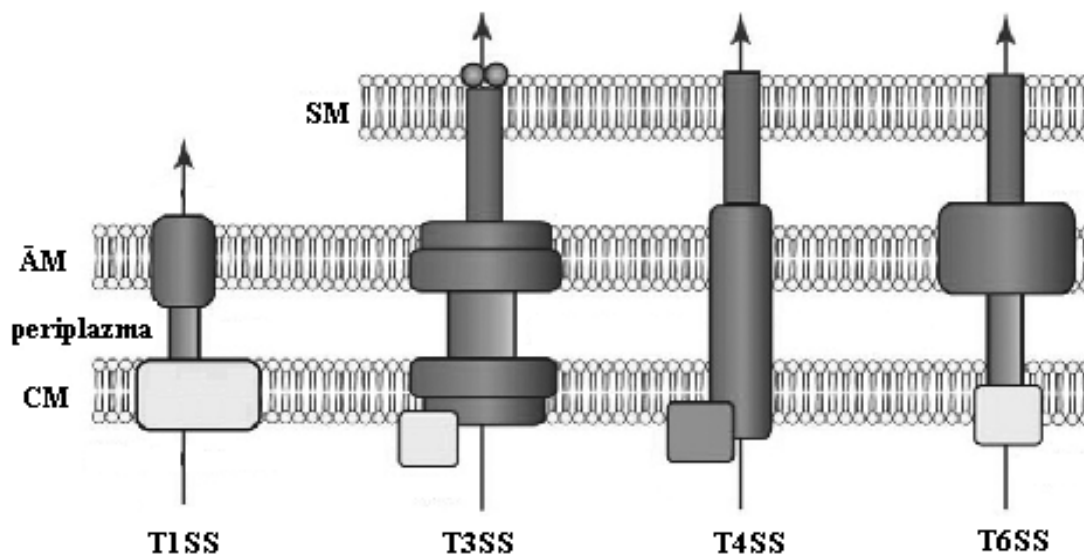
1.2. Olbaltumvielu sekrēcijas sistēmas proteobaktērijās

Proteobaktērijās olbaltumvielu sekrēciju nodrošina vismaz sešas (I – VI tips) sekrēcijas sistēmas, kuru sekrēcijas substrāti nav tieši klasificējami pēc to funkcijas, bet var atšķirties pēc specifiskām signālsekvencēm N-terminālajos fragmentos vai arī neatšķelamu sekrēcijas signālu klātbūtnes (Tseng *et al.* 2009; Desvaux *et al.* 2009). Aptuveni 20% proteobaktēriju eksoproteoma (ekstracelulāro olbaltumvielu kopums) nav konstatēta signālsekvenču klātbūtne (Bendtsen *et al.* 2005, Tseng *et al.* 2009), tādēļ tās visbiežāk definē kā „neklasiski” sekretējamas (Bendtsen *et al.* 2005). Sekrēciju bez signālpeptīda atšķelšanas nodrošina salīdzinoši vienkāršā I tipa sekrēcijas sistēma (T1SS), kompozicionāli sarežģītākās III un IV tipa sistēmas (T3SS, T4SS), kā arī samērā neseno raksturotā VI tipa sekrēcijas sistēma (T6SS; 1. Att.). Savukārt, Sec- un Tat- signālpeptīdus saturošas olbaltumvielas sekretē II un / vai V tipa sistēmas (T2SS, T5SS). Tāpat, dažu T4SS un T6SS sekrēcijas substrātu sekrēcijai nepieciešama T2SS. Attiecīgi, baktērijās *Bordetella pertussis* toksīna sekrēcija ir divpakāpju process, kur T4SS nodrošina ārējās membrānas šķērsošanu (Cascales un Christie 2003), un T6SS substrātam RbsB baktērijās *Rhizobium leguminosarum* ir konstatēta Sec-atkarīgā signālpeptīda klātbūtne (Shao *et al.*, 2007 *cit. pēc* Filoux *et al.* 2008). Analogas sekrēcijas sistēmas raksturīgas arī citām baktēriju grupām, tomēr tās plašāk pārstāvētas un izsmeltošāk analizētas tieši proteobaktērijās (McCann un Guttman 2008; Tseng *et al.* 2009).

1.2.1. Pārskats par I tipa sekrēcijas sistēmu komponentu un substrātu īpatnībām

T1SS uzbūves vienkāršība (1. Att.) atšķir to no citām sekrēcijas sistēmām (Koronakis *et al.* 1991; Holland *et al.* 2005; Saier 2006; Gerlach un Hensel 2007). Tā ir veidota no ABC eksportiera (saīsinājums ABC atvasināts no jēdziena '*ATP-binding cassette*'), baktērijas citoplazmatisko un ārējo membrānu sajūdošām olbaltumvielām (*membrane fusion proteins MFP*) un ārējās membrānas olbaltumvielas TolC (*outer membrane protein OMP*). Vēsturiski

pirmais eksperimentāli apstiprinātais un anotētais T1SS sekrēcijas substrāts ir *Escherichia coli* hemolizīns HlyA (Lory 1992; Bendtsen *et al.* 2005), tādēļ nereti T1SS literatūrā dēvē par hemolizīna sekrēcijas sistēmu. Tās substrāti tiek sekretēti natīvajā formā (Holland 2004; Binet *et al.* 1997 *cit. pēc* Sandkvist 2001), un sekrēcijas process var būt cieši saistīts ar olbaltumvielas translāciju vai noritēt vienlaikus ar sekrēcijas substrātu foldingu (Masi un Wandersman 2010; Sapriel *et al.* 2002). Tāpat ir ziņots par iespējamu T1SS piemērotību arī lielmolekulāru eksopolisaharīdu sekrēcijai (Tseng *et al.* 2009).



1. att. Proteobaktēriju neklasiskās sekrēcijas sistēmu vispārēja shēma (Tseng *et al.* 2009). Apzīmējumi: SM – saimniekorganisma šūnas membrāna; ĀM – ārējā membrāna; CM – citoplazmatiskā membrāna.

Fig. 1 Generalized scheme of non-classical secretion systems in proteobacteria (Tseng *et al.* 2009). Abbreviations: SM – host cell membrane; ĀM – outer membrane; IM – inner membrane.

1.2.2. Pārskats par III tipa sekrēcijas sistēmu komponentu un substrātu īpatnībām

Vismaz septiņu T3SS kompleksu paveidus (Hueck 1998; Ghosh 2004) veido aptuveni 20-30 komponenti (1. Att.), no kuriem konservatīvi ir tikai deviņi (Tseng *et al.* 2009). Dažu T3SS komponentu sekvenču N-terminālajos fragmentos ir lokalizēts Sec- signālpeptīds, un T3SS ārējās membrānas komponents ir homologs ārējās membrānas poru veidojošajai T2SS olbaltumvielai PulD, kas norāda uz iespējamu Sec- sistēmas lomu T3SS darbības nodrošināšanā (Hueck 1998). T3SS raksturīga adatveida (*needle-like*) struktūra, kas atsevišķu baktēriju gadījumā var atšķirties. Tā, *Pseudomonas syringae* T3SS ir saistīta ar piliem (*pilus*); *Salmonella sp.* gadījumā ar filamentiem (*filament*), bet *Yersinia* ģints baterijās tā līdzinās labai adatai

(*hollow needle*; Chibucos *et al.* 2009). T3SS ir piemērota baktēriju efektoro olbaltumvielu tiešai nogādei saimniekorganisma citosolā (Cornelis 2002; Gophna *et al.* 2003; Ma *et al.* 2006). Savukārt, *Pseudomonas syringae* Hrp (saīsinājums atvasināts no '*harpin*') olbaltumvielas tiek sekretētas starpšūnu telpā (Tseng *et al.* 2009; Li 2007). T3SS ir morfoloģiski līdzīga flagellārajam aparātam, un to filoģenētiskajā attīstībā ir saskatāma virkne kopīgu iezīmju (Gophna *et al.* 2003; McCann un Guttman 2008). Lai gan, atšķirībā no flagellu sistēmas, T3SS kodējošie gēni vairumā gadījumu ir izvietoti baktēriju plazmīdu t.s. patogenitātes saliņās (*pathogenicity islands PAI*; McCann un Guttman 2008; Tseng *et al.* 2009). T3SS sekrēcijas sistēmas gēnos guanīna un citozīna (GC) satura (40-45%) novērtējums ir salīdzinoši zemāks nekā pārējam organismu genomam. Tādēļ par iespējamo T3SS izcelšanās avotu hipotētiski ir izvirzītas *Chlamydia* ģints baktērijas, kurām kopumā raksturīgs zems GC saturs (Hueck 1998). Turklāt T3SS gēni šajās baktērijās nav izvietoti atsevišķās patogenitātes saliņās (Arnold *et al.* 2009), un to flagellu komponentiem ir cieša saistība ar T3SS darbību (Snyder *et al.* 2008), kas netieši norāda uz T3SS senāku izcelsmi šajās baktērijās. T3SS darbība lielākoties ir raksturota kā kotranslacionāla, tomēr ir ziņots par *Yersinia enterocolitica* efektorās olbaltumvielas YopQ posttranslacionālas sekrēcijas norises iespēju (Cambronne un Roy, 2006, Trček *et al.* 2002). T3SS efektorajām olbaltumvielām nereti novērota N- un C- terminālo fragmentu homologija, kas, iespējams, veidojusies, saplūstot divu efektoru olbaltumvielu, vai arī efektoru un citas olbaltumvielas, gēniem (Stavrinides *et al.* 2006).

1.2.3. Pārskats par IV tipa sekrēcijas sistēmu komponentu un substrātu īpatnībām

T4SS sekrēcijas sistēmas adatveida struktūru veido vismaz 10 komponenti (1. Att.). Tā raksturojama ar salīdzinoši vēlīnu izcelšanos, pārejot no sākotnējā baktēriju konjugācijas kompleksa formas uz efektoru olbaltumvielu sekrēcijas sistēmu (Cascales un Christie 2003; Cao un Saier 2001). Ir ziņota T4SS un konjugācijas sistēmu savstarpēja mijiedarbība, veidojot potenciālu to izmantošanai DNS tiešai nogādei eikarotu šūnās (DePaz *et al.* 2005). Baktērijās *Agrobacterium tumefaciens* T4SS komponents VirB10 ir raksturots kā vienīgā identificētā olbaltumviela, kas proteobaktērijās šķērso abas membrānas (Chandran *et al.* 2009).

1.2.4. Pārskats par VI tipa sekrēcijas sistēmu komponentu un substrātu īpatnībām

T6SS sekrēcijas sistēmai (1. Att.) ir raksturīga adatveida struktūra (*phage-tail-spike-like injectisome*; Tseng *et al.* 2009). To kodējošie aptuveni 15 gēni ir izvietoti t.s. patogenitātes salīnās (Bingle *et al.* 2008; Filloux *et al.* 2008). T6SS gēni ir pārstāvēti aptuveni ceturtdaļā šobrīd sekvenčēto baktēriju genomu un atbilst vismaz pieciem T6SS paveidiem (Bingle *et al.* 2008; Tseng *et al.* 2009), turklāt nav izslēgta T6SS sekrēcijas sistēmu evolucionēšana no bakteriofāga olbaltumvielām (Cascales 2008; Leiman *et al.* 2009). Joprojām sastopami pretrunīgi uzskati par T6SS klātbūtni nepatogēnu baktēriju celmu genomos (Boyer *et al.* 2009; Shrivastava un Mande 2008). Tāpat ir pretrunīgi ziņojumi par to, vai T6SS atkarīgā sekrēcija ir vienpakāpju vai divpakāpju process (Tseng *et al.* 2009, Cascales 2008). ATF-āzei ClpV ir ziņota T6SS šaperona aktivitāte baktērijās *Vibrio cholerae* un būtiska loma T6SS kompleksa aktivizēšanā baktērijās *Pseudomonas aeruginosa* (Bönemann *et al.* 2009). *Vibrio cholerae* T6SS komponentam VrgG-1 ir novērotas efektoras funkcijas makrofāgu šūnās, un ir izvirzīta hipotēze, ka T6SS komponenti kopumā var kalpot kā efektori, tātad var tikt konceptuāli salīdzināti ar T5SS autotransportieriem (Pukatzki *et al.* 2007, Pukatzki *et al.* 2009). Homologi sekvenču domēni ir novēroti T6SS ClpB ATF-āzei un T3SS komponentam YscN, kā arī transmembranālajiem VasF un VasK komponentiem ar IcmF un IcmH (T4SS), savukārt VrgG un Hcp veidoto sekrēcijas kanālu morfoloģija ir līdzīga ar HlyA un RtxA kanāliem (T1SS; Shrivastava un Mande 2008).

1.3. Olbaltumvielu sekrēcijas sistēmu izplatība un daudzveidība

Tā kā olbaltumvielu neklasiskās sekrēcijas sistēmas sākotnēji identificētas patogēnās baktērijās, tā šādu transportsistēmu pārstāvība baktērijas genomā joprojām tiek uzskatīta par patogenitātes apliecinājumu. Tomēr vispārējā skatījumā tas norāda uz baktēriju un saimniekorganismu asociāciju iespējamību vai līdzdalību bioplēves izveidē, u.tml., nevis obligātu patogenitāti (McCann un Guttman 2008, Tseng *et al.* 2009, Coombes 2008, Boyer *et al.* 2009). Līdzīgi, patogēnās un nepatogēnās baktērijās ir novērots savstarpēji samērojams Sec-atkarīgo sekrēcijas substrātu apjoms (Song *et al.* 2009). Vairumam patogēno proteobaktēriju sekrēcijas sistēmu paralogās kopijas ir iesaistītas dažādu patoģenēzes fāžu norisē, tādējādi norādot, ka straujāku baktēriju evolūciju un to adaptāciju jaunās ekoloģiskās nišās visdrīzāk ir nodrošinājusi horizontāli pārnestu sekrēcijas sistēmu gēnu klāsteru atkārtota pārņemšana (*acquisition*), nevis jau esošu sekrēcijas sistēmu pielāgošanās jaunām funkcijām (Gerlach un

Hensel 2007). Tā, vairākas baktērijas (*Salmonella* sp., *Yersinia* sp.) var kodēt vairāk kā vienu no septiņiem T3SS kompleksiem (Troisfontaines un Cornelis 2005; Ghosh 2004), kamēr T4SS kompleksi baktēriju *Legionella pneumophila* celmos ir savstarpēji atšķirīgi (DeBuck *et al.* 2007). Baktēriju *Aeromonas* spp. T3SS paveidiem novērotas atšķirības to regulācijā, un eksperimentāli novērota T3SS efektoro olbaltumvielu sekrēcijas ciešā saistība ar citiem procesiem šūnā (Vilches *et al.* 2009). Vairāku sekrēcijas sistēmu un to paveidu ekspresija ir būtiska baktēriju dzīvotspējai dažādos saimniekorganismos (Bleasdale *et al.* 2009), nav izslēgta arī vienas un tās pašas efektorās olbaltumvielas atšķirīga loma mutuālu, komensālu vai parazītisku asociāciju gadījumos (Coombes 2008). Baktērijās *Salmonella enterica* un *Pseudomonas aeruginosa* eksperimentāli nevērota savstarpēji izslēdzoša T3SS un T6SS regulācija, tā norādot uz abu sekrēcijas sistēmu darbību atšķirīgās infekcijas fāzēs (Parsons un Heffron 2005 *cit. pēc* Cascales 2008). Tādējādi jāsecina, ka sekrēcijas kompleksu atšķirības var nodrošināt augstu sekrēcijas substrātu specifiskumu, vienlaicīgi arī sarežģot noteiktai olbaltumvielai atbilstošās sekrēcijas sistēmas eksperimentālu noteikšanu.

1.4. Pārskats par sekrēcijas signāliem un to nozīmi

Sekrēcijas norisē izšķirīga ir sekrēcijas substrātu atpazīšana pēc sekrēcijas signālu lokalizācijas sekvencēs (Holland 2004; Cambronne un Roy 2006; Bendtsen *et al.* 2005; Gerlach un Hensel 2007, DeBuck *et al.* 2007). Turklāt Sec- atkarīgās sekrēcijas signālpeptīdi, ne vien iezīmē sekretējamo olbaltumvielu, bet arī allostēriski aktivē Sec- translokāzi (Gouridis *et al.* 2009).

Neklasiskās sekrēcijas signālu sekvenču nav konservatīvas, tomēr vairākas sekvenču terminālo fragmentu sastāva un otrējās struktūras pazīmes tiek uzskatītas arī par neatšķelamo sekrēcijas signālu indikatoriem. Šeit jāņem vērā, ka sekrēcijas substrāta funkcionālā un sekrēcijai piemērotā (*secretion competent*) struktūra var būt atšķirīga (Izadi-Pruneyre *et al.* 1999; Evdokimov *et al.* 2001, Chervaux un Holland 1996; Zhang *et al.* 1993; Gerlach un Hensel 2007). T1SS, T4SS un T6SS sekrēcijas substrātiem par sekrēcijas signāla apgabalu uzskata aptuveni 50 aminoskābju fragmentu C- terminālajā apgabalā (Holland 2004; Izadi-Pruneyre *et al.* 1999; Christie un Cascales 2005; Cambronne un Roy 2006; Boyer *et al.* 2009).

Baktērijās *E. chrysantiemi* par T1SS sekrēcijas signālu uzskata sekvenču C- terminālo fragmentu kur negatīvi lādētai aminoskābei seko 3-5 hidrofobu aminoskābju atlikumi (Izadi-

Pruneyre *et al.* 1999). *Escherichia coli* hemolizīna T1SS sekrēcijas signālam ir raksturota vāja otrējās struktūras konservativitāte (Stanley *et al.* 1991 *cit. pēc* Gerlach un Hensel 2007). *Serratia marcescens* hemoforam HasA bez C- terminālā sekrēcijas signāla iespējami papildus signāl- apgabali, kas izvietoti visas olbaltumvielas garumā, t.sk. to N- terminālajā galā, un nodrošina T1SS komponentu sagrupēšanos (*assembly*), stabilizējot C- terminālā signāla mijiedarbību ar T1SS komponentiem (Masi un Wandersman 2010; Sapriel *et al.* 2002). Tāpat ir ziņots par diviem T1SS vai vismaz to sekrēcijas signālu paveidiem (Holland 2004), kas zināmā mērā skaidro pretrunīgos ziņojumus par cisteīna klātbūtni un tā lomu T1SS sekrēcijas substrātu sekvencēs (Angkawidjaja un Kanaya 2006). Tā, eikariotiskas izcelsmes olbaltumvielai ar pievienotu *Erwinia chrysanthemi* proteāzes C- terminālo domēnu nebija iespējama T1SS sekrēcija, norādot, ka T1SS sekrēcijas substrātu peptīdiem nav pieļaujama disulfīdu saites veidošanās (Palacios *et al.* 2001 *cit. pēc* Angkawidjaja un Kanaya 2006). Savukārt sekmīgai sekrēcijas norisei *Escherichia coli* hemolizīna T1SS sekrēcijas sistēmas gadījumā disulfīdu saites klātbūtne substrātā ir ziņota kā absolūti nepieciešama pazīme (Fernández un Lorenzo 2001), un tieši hemolizīna C- terminālā sekrēcijas signāla pievienošana cisteīnu saturošai *E. coli* fosfatāzei (*alkaline phosphatase*) nodrošina tās T1SS sekrēciju (Hess *et al.* 1990 *cit. pēc* Angkawidjaja un Kanaya 2006).

Ir ziņots par T3SS sekrēcijas signālu sekrēcijas substrātu sekvenču vai to mRNS N- terminālā galā, kā arī sekrēcijas kompleksa receptora vai šaperona inducētas konformācijas izmaiņas sekrēcijas substrātu N- terminālajos fragmentos (Birtalan *et al.* 2002; Ghosh 2004, Alfano un Collmer 2004). T3SS šaperoniem un efektorajām olbaltumvielām raksturīgo konservatīvo otrējo struktūru apliecina dažādu sugu baktēriju efektoru-šaperonu pāru savstarpējā aizvietojamība (Locher *et al.* 2005).

1.5. Olbaltumvielu sekrēcijas regulācija

Ir ziņoti atšķirīgi olbaltumvielu sekrēcijas regulācijas mehānismi, gan sekrēcijas substrātu translācijas līmenī, gan arī kontrolējot transportistēmu gēnu ekspresiju. Kvoruma sajūta (*quorum sensing*) ir būtisks faktors *Escherichia coli* un *Pseudomonas aeruginosa* T3SS efektoru olbaltumvielu sekrēcijai (Sperandio *et al.* 1999; Wolfgang *et al.* 2003), norādot, ka infekcijas norisei ir nepieciešama noteikta baktēriju koncentrācija. Un, gluži pretēji, baktērijās *Vibrio harveyi* kvoruma sajūtas inducētā *lux* operona ekspresija aptur T3SS darbību (Winans 2004).

Līdzīgi, ir norādes, ka T2SS sekrēcijas substrātu ekspresija un sistēmas komponentu sagrupēšanās baktērijās *Pseudomonas aeruginosa* ir atkarīga no apkārtējās vides apstākļiem, baktēriju augšanas fāzes un kvoruma sajūtas (Michel un Voulhoux 2009).

Siltasiņu dzīvnieku patogēnu virulences gēnus inducē saimniekorganismiem raksturīgā temperatūra. *Yersinia enterocolitica* T3SS un flagellārās sistēmas reversā regulācija pie augstākas temperatūras (37°C) nodrošina baktēriju virulenci, bet pie zemākas (28°C) to kustīgumu (Kapatral *et al.* 1996 *cit. pēc* Wilharm *et al.* 2004). Turpretī augu patogēna *Pseudomonas syringae* virulences gēnu ekspresija ir novērojama pie temperatūras, kas zemāka par saimniekauga attīstībai optimālo. Tā, sojas pupiņas parasti tiek inficētas agri pavasarī, kad temperatūra ir piemērota baktērijas attīstībai, bet ir pārāk zema, lai nodrošinātu optimālu auga aizsargmehānismu darbību (Li 2001). Tādā pat mērā olbaltumvielu sekrēcija ir tieši saistīta ar baktēriju augšanai piemērotāko temperatūru. Proti, T1SS atkarīgās lipāzes sekrēcijai psihofilās baktērijās *Pseudomonas fluorescens* ir piemērota temperatūra, kas zemāka par 30°C, un, gluži pretēji, *Pseudomonas aeruginosa* lipāzes sekrēcijai optimālā temperatūra ir 37°C (Ahn *et al.* 1999).

Baktēriju *Burkholderia thailandensis* T3SS inducē samazināts vides pH (Jitprasutwit *et al.* 2010). Līdzīgi, saimniekšūnu vakuolās dzīvojošo baktēriju *Salmonella enterica* T3SS efektoro olbaltumvielu sekrēcija notiek pie zema šūnu pH, kad degradējas regulatorā kompleksa olbaltumvielas SsaM, SpiC un SsaL (Yu *et al.* 2010). Fosfātu nepietiekamības apstākļos noris intensīvāka T1SS atkarīgo olbaltumvielu sekrēcija baktērijās *Neisseria meningitides* (Turner *et al.* 2009). Līdzīgi, *Serratia marcescens* T1SS efektorās olbaltumvielas HasA sekrēciju inducē samazināts dzelzs daudzums vidē, un noskaidrots, ka T1SS promoters satur Fur-atkarīgo promoteru (saīsinājums atvasināts no '*Ferric uptake regulator*'), kas badināšanas apstākļos inducē dzelzs uzņemšanas mehānismus (Letoffe *et al.* 1994b un Oxhsner *et al.* 2000 *cit. pēc* Jenewein *et al.* 2009). Savukārt, dizentērijas izraisītāja *Shigella flexneri* T3SS komponenti sagrupējas anaerobā vidē, bet efektoro olbaltumvielu sekrēcijai ir nepieciešama skābekļa klātbūtne (Marteyn *et al.* 2010).

1.6. Sekretējamo olbaltumvielu biotehnoloģiskais potenciāls

Joprojām pieaugošais sekvencēto baktēriju genomu apjoms kopskatā ar zināšanu un priekšstatu attīstību par atsevišķu sekrēcijas sistēmu un to substrātu īpašībām ievērojami

paplašina un veicina tehnoloģiju izstrādi āršūnas olbaltumvielu pielietojumam biomedicīnā un rūpniecībā. Tā, kopš neklasiskās olbaltumvielu sekrēcijas sistēmu atklāšanas patogēnās baktērijās, tiek meklētas un praktiski aprobētas jaunas antibiotikas šo transportsistēmu inaktivēšanai. Ir veidotas pieejas vakcīnu izstrādei, izmantojot baktēriju mutantus. Piemēram, *Bordetella bronchiseptica* un *Brucella abortus* gadījumos, ar inaktivētu, attiecīgi, T3SS vai T4SS sekrēcijas sistēmu (Mann *et al.* 2007; Rolán *et al.* 2009), novērots, ka šādas rekombinantas baktērijas, lai arī saimniekorganismam nepatogēnas, izsauc specifisku imūno atbildi. Tāpat, labas perspektīvas T3SS, T4SS un T6SS izmantošanai dažādu terapeitisko aģentu, tostarp vakcīnu, tiešai nogādei organisma šūnās nosaka šīm sekrēcijas sistēmām raksturīgā efektoro olbaltumvielu translocēšana saimniekorganisma citosolā. Zināšanu kopums par sekrēcijas signāla dabu jau šobrīd ļauj veidot rekombinantas olbaltumvielas, t.sk. himēras, kad mērķolbaltumvielas sekvencei pievieno atbilstošai sekrēcijas sistēmai raksturīgo sekrēcijas signālu (Gentshev *et al.* 2005; Lotter *et al.* 2004; Lacroix *et al.* 2008; Chung *et al.* 2009; Pukatzki *et al.* 2006). Šādu pieeju arvien biežāk izmanto arī efektoro olbaltumvielu sākotnējai atlasei ('*effector detector*' screens), panākot to sekrēciju mērķšūnās, līdztekus novērojot iespējamās to izsauktās izmaiņas šūnā un novērtējot to potenciālu turpmākajam pielietojumam medicīniskajā biotehnoloģijā (Alfano 2009). Lai gan nepieciešamība pievienot/atšķelt sekrēcijas signālu būtiski kavē neklasiskās sekrēcijas sistēmu izmantošanu biotehnoloģijā (Blight un Holland 1994 *cit. pēc* Mergulhão *et al.* 2005), tomēr, augot izpratnei par baktēriju sekrēcijas sistēmu mehānismiem, to komponentu un substrātu mijiedarbību, pieaug arī plašāka pielietojuma iespējas. Līdztekus palielinoties atbilstošu prognozēšanas-raksturošanas metožu klāstam un precizitātei, turpmāk ir iespējama efektīvāka atbilstošākā producenta izvēle, kas ir izšķirīga, lai novērstu, vismaz mazinātu, tādas problēmas, kā mērķolbaltumvielu ekspresēšanas grūtības, to iespējamā toksicitāte producenta šūnā, nepietiekama sekrēcijas kompleksa kapacitāte un proteolīzes izsauktie efekti (Mergulhão *et al.* 2005).

1.7. Sekretējamo olbaltumvielu eksperimentālas identificēšanas un raksturošanas metodika un prognozēšanas metožu izstrāde

Olbaltumvielu identifikācijai un funkcionālo īpašību prognozēšanai vienlīdz sekmīgi izmantojamas gan tīri eksperimentālās, gan skaitļošanas metodes. Jāuzsver, ka pētījumi *in silico* ir uzskatāmi par eksperimentiem, kas veikti ar bioinformātikas metodēm, tomēr promocijas darbā

jēdziens 'eksperimentāls' ir attiecināts uz pētījumiem laboratorijas apstākļos.

Eksperimentāli noteiktās olbaltumvielu īpašības apkopo datubāzēs, kuras izmanto jaunu skaitļošanas metožu izstrādē, tās precizējot līdztekus ar olbaltumvielām prognozēto īpašību eksperimentālu pārbaudi (Emmanuelson *et al.* 2007; Alfano 2009; Abfalg *et al.* 2007; Sleator un Walsh 2010; Schnoes *et al.* 2009). Vērojama vispārēja tendence, ka lielākais vairums datubāzēs norādīto olbaltumvielu anotāciju ir iegūtas, izmantojot prognozēšanas metodes (Schnoes *et al.* 2009), taču kopumā tām piemīt būtisks ierobežojums. Nav apšaubāms, ka iegūtos rezultātus novērtēt *sensus stricto* ir iespējams tikai eksperimentāli, un eksperimentāli iegūtu olbaltumvielu pazīmju anotāciju nepietiekamība liedz izveidot pilnvērtīgas izstrādes sekvenču kopas (*training set*) un sasniegt maksimālu prognozēšanas kvalitāti. Jāņem vērā, ka, novērtējot olbaltumvielu īpašības *in silico* tāpat kā izmantojot eksperimentālas pieejas, ir jāsalīdzina ne vien dažādu prognozēšanas metožu rezultātus, bet arī atšķirīgu olbaltumvielu sekvenču un īpašību datubāžu anotācijas. Turklāt jāņem vērā, ka eksperimentāli pārbaudītās olbaltumvielu īpašības bieži tiek pirmēji apkopotas maza mēroga (*local*), nevis aptverošākajās (*large-scale*) datubāzēs (Schnoes *et al.* 2009).

Baktēriju un eikariotu olbaltumvielām (un tās kodējošajām sekvencēm) šobrīd ir izveidotas sekvenču funkcionālo motīvu, t.sk. to 3-D struktūru, un līdzīgu specifisku signatūru, kā arī ortologo (*orthologous*) sekvenču datubāzes, tāpat ir iespējams novērtēt sekvenču homoloģiju un atbilstošo gēnu lokalizāciju genomā (Sleator un Walsh 2010; Schnoes *et al.* 2009; Emmanuelson *et al.* 2007; Bendtsen *et al.* 2005). Tomēr joprojām liela daļa, ar *in silico* metodēm iegūto, anotāciju ir prognozēta, attiecinot sekvenču funkcionālās īpašības, pamatojoties tikai uz sekvenču homoloģijas novērtējumu, kas, piemēram, fermentu gadījumā ir nepietiekami (Schnoes *et al.* 2009; Sleator un Walsh 2010). Turklāt vairākās datubāzēs ir konstatētas kļūdaini anotētu olbaltumvielu kopas (*clusters of misannotated proteins*). Tās veidojas, uz *de novo* raksturojamajām sekvencēm attiecinot īpašības, kas kļūdaini piemērotas olbaltumvielai, kas izmantota par atskaites sekvenci (Schnoes *et al.* 2009).

Neklasiski sekretējamu olbaltumvielu identificēšanas eksperimentālo metožu pamatā ir virzītā mutāģenēze. Piemēram, Sec- atkarīgās sekrēcijas sistēmas inaktivēšana (Bendtsen *et al.* 2005) un sekrēcijas norisei būtisku olbaltumvielas sekvences fragmentu noteikšana (Lory 1992). Reālā laika PĶR (*Real Time RT-PCR*) ir piemērota infekcijas norisē iesaistītās sekrēcijas sistēmas gēnu transkripcijas konstatēšanai (Bleasdale *et al.* 2009). Taču jāņem vērā, ka jebkura organisma

eksoproteoma sastāvu ietekmē šūnas fizioloģiskais stāvoklis un apkārtējā (t. sk. eksperimentālā) vide, tostarp arī šūnu un olbaltumvielu lizēšanās iespējamība, kas var veidot pretrunīgus, pat savstarpēji izslēdzošus, priekšstatus par āršūnas olbaltumvielu īpašībām (Bendtsen *et al.* 2005). Turklāt individuālu olbaltumvielu raksturlielumi nereti var atšķirties no atbilstošās olbaltumvielu grupas pazīmēm kopumā. Tas izskaidro, kādēļ šobrīd zināmajām neklasiskās sekrēcijas sistēmām atbilstošu un eksperimentāli apstiprinātu sekrēcijas substrātu skaits joprojām pārstāv visai niecīgu daļu no kopējā joprojām pieaugošā sekvencēto eksoproteomu apjoma.

Tiek izstrādātas un nepārtraukti precizētas vairākas *in silico* metodes, kas pielietojamas baktēriju un eikariotu olbaltumvielu lokalizācijas prognozēšanai, kas, savukārt, ļauj spriest arī par olbaltumvielu iespējamo funkcionālo piederību (Emmanuelson *et al.* 2007; Aßfalg *et al.* 2007; Bendtsen *et al.* 2005). Tāpat, pieaugot identificētu T3SS u.c. neklasiski sekretēto olbaltumvielu skaitam, bioinformātikas metožu izstrādē arvien aktuālāks virziens ir efektoru funkciju prognozēšana (Alfano 2009). Gram-pozitīvo un proteobaktēriju āršūnas olbaltumvielas, kuru sekvences nesatur atņēlamos N- signālpeptīdus, var atpazīt, izmantojot, attiecīgi, LocateP un SecretomeP metodes (Zhou *et al.* 2008; Bendtsen *et al.* 2005), kas tomēr nesniedz nepieciešamās norādes uz tām atbilstošāko neklasiskās sekrēcijas tipu. Vispārējā BLASTP/HAMAP metode (saīsinājums atvasināts no '*The basic local alignment tool*' un '*High quality annotated and manual annotation of microbial proteomes*') ar pieņemamu precizitāti ir izmantojama olbaltumvielu sekrēcijas sistēmu komponentu homologu identificēšanai tajā vai citā genomā, par atskaites (*query*) sekvencēm izmantojot attiecīgā sekrēcijas tipa identificēto (datu bāzēs anotēto) komponentu sekvences (Bu *et al.* 2009). Tomēr, lai arī šādi iespējams identificēt atsevišķām funkcionālajām klasēm atbilstošās efektorās olbaltumvielas, kopumā šī metode sekrēcijas substrātu atlasei nav piemērota, jo to sekvences raksturojamas kā mazāk konservatīvas, salīdzinot ar attiecīgo transportsistēmu komponentiem (McCann un Guttman 2008; Arnold *et al.* 2009). 2010. gadā publicētā datubāze CoBaltDB (saīsinājums atvasināts no '*Complete Bacterial and Archaeal Orfomes Subcellular Localization Database*') dod iespēju vienkopus salīdzināt 43 prognozēšanas metožu rezultātus 784 baktēriju un arheobaktēriju proteomiem (Goudenège *et al.* 2010), tādējādi sekmējot sekvenču raksturošanas precizitāti, tomēr neklasiskās sekrēcijas substrātu klasificēšana šobrīd ir iespējama tikai T3SS robežās.

Olbaltumvielu funkciju prognozēšanas metožu izstrādes pamatā ir eksperimentāli noteiktu un / vai prognozētu sekvenču īpašību profili, t.sk. aminoskābju sastāvs, kodonu izlietojums (Goodarzi *et al.* 2008) sekvences garums, izoelektriskais punkts, otrējā struktūra, utt. (Bendtsen

et al. 2005), pamatojoties uz to, ka funkcionāli līdzīgām vai vienā šūnas segmentā lokalizētām olbaltumvielām mēdz būt kopīgas iezīmes, lai arī kopumā to sekvences var būtiski atšķirties. Tā, anotētu neklasiski sekretējamo olbaltumvielu kopumam, kurš nav grupēts pēc atsevišķiem neklasiskās sekrēcijas tipiem, ir definētas vairākas kopīgas iezīmes to aminoskābju sastāvā, otrējās struktūras īpašībās un tām ir līdzīgs nestabilitātes indekss (*instability index*; Bendtsen *et al.* 2005). *Pseudomonas syringae* pv. *tomato* Hop (saīsinājums atvasināts no '*Hrp-dependent outer protein*') olbaltumvielām ir raksturīgs augsts (>10%) serīna saturs to N50- terminālajā fragmentā, un nav sastopami asparģīnskābes un glutamīnskābes atlikumi to N- terminālajos 12 aminoskābju atlikumos, kā arī alifātisku aminoskābju (Ile, Leu, Val) vai prolīna atlikums to trešajā vai ceturtajā pozīcijā. Šīs īpašības sekmīgi attiecinātas uz *Xanthomonas oryzae* pv. *oryzae* T3SS efektorajām olbaltumvielām un ir piemērotas to atlasei (Furutani *et al.* 2009). Pēdējos gados ziņotas vairākas metodes T3SS sekrēcijas substrātu atpazīšanai atsevišķu baktēriju eksoproteomā (Samudrala *et al.* 2009, Löwer un Schneider 2009, Vinatzer *et al.* 2005), tomēr šie pētījumi aptver visai šauras baktēriju grupas. Arī zīdītāju eksoproteomā ir izdalāmas kā neklasiski sekretējamas, tā N- signālpeptīdus saturošas olbaltumvielas, kas ir savstarpēji atšķiramas pēc vispārējām olbaltumvielu sekvenču īpašībām (Bendtsen *et al.* 2004a; Hung *et al.* 2010). Līdzīgi, arī vienšūņiem *Plasmodium falciparum* ir ziņota sekretējamo olbaltumvielu prognozēšanas iespēja pēc to aminoskābju sastāva, sasniedzot precizitāti 89-91% robežās, turklāt uzsvērts, ka šim uzdevumam ir piemērota tikai daļa aminoskābju (*reduced alphabets*; Zuo un Li 2010).

Viena tipa paralogu sekrēcijas sistēmu substrātu sekrēcijas signāli var atšķirties, tie var būt dažādi arī funkcionāli atšķirīgiem substrātiem, kuri pārstāv vienu sekrēcijas sistēmu (Holland 2004). Tādējādi sekrēcijas sistēmu paveidu un tām atbilstošo sekrēcijas signālu daudzveidība pamato natīvu, jeb, nemodificētu mērķolbaltumvielu sekrēcijai piemērotu sekrēcijas sistēmu atlases iespēju. Turklāt nereti novērotā baktēriju un eikariotu olbaltumvielu homologija, arī baktēriju un eikariotu sekrēcijas sistēmu atšķirīgie sekrēcijas signāli norāda uz baktēriju neklasiskās sekrēcijas sistēmas varbūtējām iespējām sekretēt arī nemodificētas eikariotu izcelsmes olbaltumvielas (Bendtsen *et al.* 2004a; Bendtsen *et al.* 2005).

1.8. Olbaltumvielu sekrēcijas norišu jēdzieniskā daudzveidība un nomenklatūras problēmas

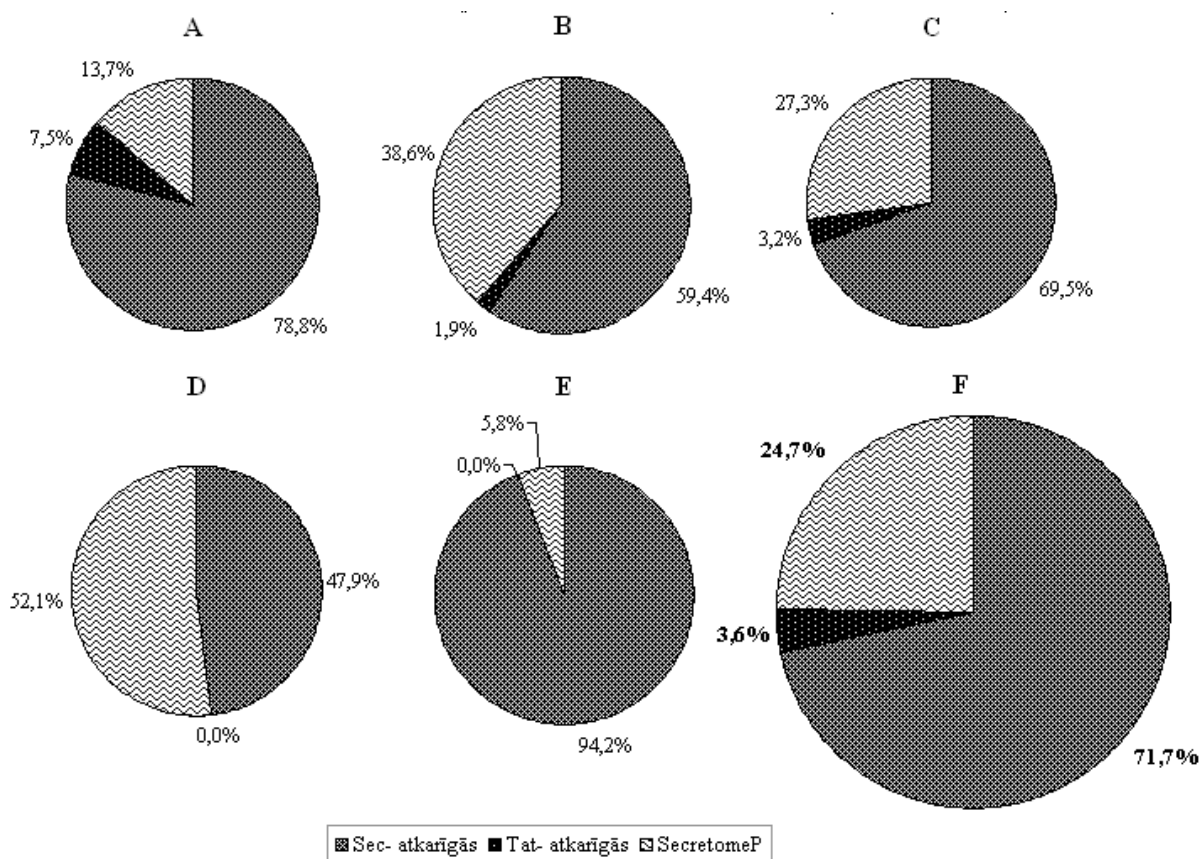
Zināšanas par olbaltumvielu sekrēciju joprojām strauji pieaug. Salīdzinoši neilgā laikā ir veidojušies un būtiski mainījušies priekšstati, arī izmantotie jēdzieni tās raksturošanai, taču vienota jēdzienu sistēma un nomenklatūra tiek veidota tikai pēdējos gados (Chibucos *et al.* 2009; Desvaux *et al.* 2010), kas liek rūpīgi izvērtēt agrīno, piemēram, pirms desmit gadiem, literatūrā publicēto atziņu un šobrīd lietoto jēdzienu saderību.

Saskaņā ar šībrīža priekšstatiem (Chibucos *et al.* 2009; Desvaux *et al.* 2010), jēdziens 'olbaltumvielu sekrēcija' (*protein secretion*) apzīmē olbaltumvielas aktīvu transportu virzienā no šūnas citoplazmas uz ārpusi, kamēr jēdziens 'translokācija' (*translocation*) tikai vispārēji norāda uz šūnas membrānas šķērsošanu, kas var notikt arī bez noteiktu sekrēcijas sistēmu līdzdalības. Protī, olbaltumvielas var nonākt vidē pēc šūnu lizēšanās vai ar membrānu vezikulu starpniecību. Savukārt, jēdziens 'eksports' (*export*) ir attiecināms uz olbaltumvielas aktīvu transportu caur citoplazmatisko membrānu un baktērijās, kurām ir tikai citoplazmatiskā membrāna (*monoderm bacteria*; Desvaux *et al.* 2009), ir uzskatāms par sinonīmu jēdzienam 'sekrēcija'. Savukārt, proteobaktēriju gadījumā jēdziens 'eksports' apzīmē olbaltumvielu transportu periplazmā. Tātad, olbaltumvielu 'eksports' un 'sekrēcija' ir tikai daļēji skaidrojama ar 'translokāciju', un 'translokācija' ne vienmēr ir attiecināma uz 'sekrēciju' vai 'eksportu'.

Tāpat, arī jēdzieni 'āršūnas' vai 'ekstracellulārs' nav uzskatāmi par sinonīmiem jēdzienam 'sekretējams' (Chibucos *et al.* 2009; Desvaux *et al.* 2010), jo 'ekstracellulārs' vispārēji apzīmē ārpus šūnas esošu (ekstracellulāru) olbaltumvielu kopumu, kamēr sekretējamās olbaltumvielas (*secreted proteins*) var palikt piesaistītas šūnas membrānai. Piemēram, N-terminālais signālpeptīds ne visos gadījumos tiek atšķelts no Sec- atkarīgajiem sekrēcijas substrātiem, bet piesaista tos šūnas membrānai (Martoglio un Dobberstein 1998 *cit. pēc* Desvaux *et al.* 2009). Tas nozīmē, ka N-terminālā signālpeptīda konstatēšana olbaltumvielas sekvencē vēl nenorāda tās ekstracellulāru lokalizāciju (*extracellularity*). Likumsakarīgi, ka arī jēdzieni 'eksoproteoms' (*exoproteome*) un 'sekretoms' (*secretome*) nav savsarpēji aizstājami. Turklāt jēdzienu 'sekretoms' ir ieteikts vispārināt (Desvaux *et al.* 2009), šādi apzīmējot ne vien olbaltumvielu sekrēcijas sistēmu komponentus un substrātus, bet arī sekretējamus ogļhidrātus u.c. lielmolekulāras vielas.

Jēdziens 'vispārējais sekrēcijas ceļš' (*general secretory pathway, GSP*) ir veidojies laikā, kad lielākajai daļai šobrīd raksturoto sekretējamo olbaltumvielu tika novērota Sec- signālpeptīda klātbūtne (saīsinājums 'Sec' veidots no vārda '*secretion*'; Pugsley 1993; Lory 1992) un Sec- sistēmas komponentu identificēšana visās baktēriju grupās un eikariotu šūnās ļāva vispārināti raksturot sekrēcijas sistēmas sistēmas izplatību un funkcionālo lomu. Atšķirīga veida N- signālpeptīdi, piemēram, Tat (saīsinājums atvasināts no '*twin arginine motif*', kas apzīmē N- signālpeptīdos raksturotu konservatīvu aminoskābju motīvu Ser-Arg-Arg-x-Phe-Leu-Lys) un SRP (saīsinājums atvasināts no '*signal recognition particle*') tika raksturoti sekrēcijas sistēmu substrātos (Agarraberes un Dice 2001). Minēto priekšstatu kontekstā sekretējamas olbaltumvielas bez šāda veida atšķelama signālpeptīda tika uzskatītas par alternatīvi, jeb, 'neklasiski' sekretējamām (*nonclassical secretion*). Taču, šobrīd Sec- un Tat- signālpeptīdus saturošo olbaltumvielu sekrēcijas sistēmas kopā ar to paveidiem u.tml. sekrēcijas kompleksiem ir klasificētas kā II tipa sekrēcijas sistēma (T2SS; Tseng *et al.* 2009). Turklāt līdz ar sekvencēto genomu skaita pieaugumu, arvien vairāk organismiem ir prognozēts, ka vispārējā sekrēcijas ceļa atkarīgas sekretējamās olbaltumvielas ne vienmēr ir dominējoša sekretoma daļa (2. Att.; Bendtsen *et al.* 2005, Tseng *et al.* 2009). Tādēļ ir likumsakarīgi, ka jau šobrīd apšaubā jēdzienu 'vispārējais sekrēcijas ceļš' un 'neklasiskā sekrēcija' izmantošanas lietderību (Desvaux *et al.* 2009), to vietā piedāvājot vienotu apzīmējumu 'sekrēcija bez signālpeptīdu atšķelšanas' (*leaderless secretion; signal-peptide independent secretion; secretion without signal peptide*) vai arī, gluži pretēji, ar vienu jēdzienu nevispārināt T1SS, T3SS, T4SS un T6SS sekrēcijas sistēmas. Tomēr joprojām literatūrā Sec- atkarīgā sekrēcija tiek minēta kā 'galvenais sekrēcijas ceļš' ('*main bacterial translocase*'; Gouridis *et al.* 2009) un pietiekami plaši izmanto arī jēdzienu 'neklasiskā' sekrēcija (Bendtsen un Wooldridge 2009). Jēdziens 'neklasiskā' sekrēcija izmantots arī promocijas darbā tā pietiekamās jēdzieniskās ietilpības un lakonisma dēļ.

Vienotas nomenklatūras trūkums un izmantoto abreviatūru daudzveidība, apzīmējot vienas un tā pašas sekrēcijas sistēmas, var veicināt neprecīzus, pat kļūdainus, pieņēmumus un secinājumus. Piemēram, saīsinājums ABC tiek plaši lietots tādos jēdzienos kā 'olbaltumvielu sekrēcija pa ABC- eksportieri' (*protein secretion by ABC-exporter*), 'ABC sekrēcijas sistēma' (*ABC secretion pathway*), u.c., kas vienlaicīgi apzīmē gan olbaltumvielas eksportu pār iekšējo membrānu, gan olbaltumvielas sekrēciju pa T1SS, kuras ietvaros ABC- transportieris ir tikai daļa no sekrēcijas kompleksa (Saier 2006). Kopumā sekrēcijas sistēmas atsevišķu organismu ietvaros



2. att. Prognozētais Sec- un Tat- atkarīgās sekrēcijas un N-signalpeptīdus nesaturošu (secretomeP) olbaltumvielu samērs baktērijās *Bordetella bronchiseptica* (A), *Zymomonas mobilis* (B), *Escherichia coli* (C), *Rickettsia conorii* (D), *Legionella pneumophila* str:Lens (E), kā arī proteobaktēriju (132 sugu 168 celmi) eksoprotomā kopumā (F). Dati no Genome Atlas datubāzes (<http://www.cbs.dtu.dk/>).

Fig. 2 Distribution of the predicted Sec- and Tat- dependent as well as leaderless secretory proteins within *Bordetella bronchiseptica* (A), *Zymomonas mobilis* (B), *Escherichia coli* (C), *Rickettsia conorii* (D), *Legionella pneumophila* str:Lens (E), and total exoproteome of sequenced proteobacteria (168 strains of 132 species). Data source: Genome Atlas database (<http://www.cbs.dtu.dk/>).

tiek apzīmētas ar dažādiem nosaukumiem, piemēram, T3SS *Burkholderia* un *Yersinia* ģints baktērijās apzīmē ar, attiecīgi, Bsa un Ysc (saīsinājumi atvasināti no '*Burkholderia secretion apparatus*' un '*Yop secretion*'), utt. (Jitprasutwit *et al.* 2010; Cornelis 2002).

Šobrīd ierosināts jēdzieniski nodalīt T3SS sekrēcijas sistēmu iespējas translocēt sekrēcijas substrātus vidē un saimniekorganisma citosolā, ieviešot tādus apzīmējumus, kā 'injektisomas aktivitāte' (*injectisome activity*) un 'tranlokatora aktivitāte' (*translocator activity*; Chibucos *et al.* 2009). Savukārt, apzīmējums 'T7SS' pagaidām ir attiecināts uz divām atšķirīgām olbaltumvielu sekrēcijas sistēmām, proti, uz filamentu ģenēzē iesaistīto CU sistēmu (saīsinājums atvasināts no '*chaperone-usher pathway*'), kā arī ESAT-6 (saīsinājums atvasināts no '*early secretory antigenic*

target') sekrēcijas sistēmu mikobaktērijās (Goudenège *et al.* 2010; Desvaux *et al.* 2009; Bitter *et al.* 2009 *cit. pēc* Goudenège *et al.* 2010).

Vispārējā gadījumā sekrēcijas sistēmas mēdz iedalīt arī pēc to substrātu signālsekvenču un translokatoru īpašībām, substrāta strukturētības translokācijas laikā (natīvs/foldēts) un enerģētiskā nodrošinājuma avota (Holland 2004). Tāpat sekrēcijas sistēmas var iedalīt pēc membrānu šķērsošanas veida (Filloux *et al.* 2008): vienpakāpju (T1SS, T3SS, T4SS, T6SS) vai divpakāpju (T2SS, T5SS). Arī pēc sekrēcijas substrātu nogādes vietas: vidē (T1SS, T2SS, T5SS) vai saimniekšūnu citosolā (T3SS, T4SS, T6SS). Tomēr jebkura šāda veida klasifikācija ir piemērota tikai procesa vispārējam novērtējumam, jo neņem vērā individuālus, t.sk. izņēmuma, gadījumus, kas nereti tiek vispārināti, attiecinot tos uz sekrēcijas sistēmu kopumā. Jāpiekrīt viedoklim, ka ziņojumi par sekrēcijas sistēmu skaitu ir atkarīgi no autoru izvēlētā skatpunkta, t.i., līdzīgas sekrēcijas sistēmas autori var nodalīt kā vienu vai vairākas, kā arī ņemt vērā vai neieskaitīt virsmas olbaltumvielu sekrēcijas sistēmas (Wooldridge 2009).

2. Metodes

1.1. Izstrādes kopas olbaltumvielu sekvences

111 olbaltumvielām ar anotētu atbilstību noteiktai (T1SS, T3SS, T4SS vai T6SS) proteobaktēriju sekrēcijas sistēmai (attiecīgi, 32, 41, 16 un 22 sekvences) UniProt (The UniProt Consortium 2008) un EMBL (Cochrane *et al.* 2008) datubāzēs iegūtas aminoskābju un atbilstošās kodējošās sekvences (1. pielikums Publikācijā IV). Olbaltumvielas hlyt_vibho gadījumā par kodonu izlietojuma frekvencēm izmantots reversās translācijas prognozēšanas metodes “Reverse Translate” (Gasteiger *et al.* 2003) rezultāts, katrai aminoskābei attiecinot viena kodona frekvences.

T1SS sekrēcijas substrāti: lipa_psefl, hasa_serma, o67991_psefl, o85374_promi, slap_caucr, hlyac_ecoli, prtb_erwch, prtx_erwch, q6sqm7_pseae, cyaa_borpe, nodo_rhylv, lka11_pasha, lka13_pasha, frpa_neimb, q9xb65_erwam, hlyt_vibho, q6d3f9_erwct, paxa_pasae, frpc_neimc, q59932_serma, o87109_serma, prta_phoaz, q93rn3_yerru, hlya_actsu, slap_camfe, prtc_erwch, apra_pseae, przn_sersp, prtg_erwch, alge4_azovi, alge7_azovi, rtx11_actpl.

T3SS sekrēcijas substrāti: sobp_salty, ipab_shifl, q51451_pseae, o30529_pseae, yope_yeren, hrpz_psesy, hrpn_erwam, q87y16_psesm, o33976_ecoli, popa_ralso, q57nf3_salch, q7brz4_yeren, q8vpk4_psefl, hopd2_psesm, q7wlx4_borbr, o87327_psesx, q8zq79_salty, q57qv1_salch, hrpn_peccc, hrpz_pseap, avrd1_psess, ipab_shidy, avrd1_psesg, aext_aersa, q52497_ralso, ipgd_shiso, sobp_saldu, sipa_salen, sptp_salpa, yoph_yerps, ipac_shidy, arpt2_psesm, ipaa_shifl, ipad_shifl, lcrv_yerpe, avrb3_xaneu, yope_yerpe, avrp3_psesh, ypka_yerps, sope_saldu, q9kwh9_ecolx.

T4SS sekrēcijas substrāti: vire2_agrt5, vire2_agrtu, a5wy70_9rhiz, a5wy71_9rhiz, q8vt81_9rhiz, q7d2d9_agrt5, caga_helpj, caga_helpy, cga1_helpy, a5idc4_legpc, q5wv98_legpl, q5wxw8_legpl, q5x6g3_legpa, q6g2a5_barhe, q6g2a4_barhe, q6g2a3_barhe.

T6SS sekrēcijas substrāti: hcp1_pseae, q6ee19_edwta, a8yqr4_edwta, q4v2q6_burma, q6rk44_erwct, q7axz0_salet, a1em96_vibch, q1bs59_burca, q1btx6_burca, q391l0_burs3, q39k65_burs3, q391l3_burs3, q2sv41_burta, a4jvs7_burvg, q1qv84_chrsd, q1qva4_chrsd, q1r140_chrsd, q1d0w1_myxxd, q1i7q4_psee4, q3kdt4_psepf, q0jzd4_raleh, q2p058_xanom.

2.2. Neatkarīgas paraugkopas olbaltumvielu sekvences

Kā neatkarīga olbaltumvielu sekvenču datu kopa diskriminantfunkciju prognozēšanas precizitātes pārbaudei izmantotas 34 *Salmonella typhimurium* T3SS sekrēcijas substrātu (1. Tab. publikācijā Samudrala *et al.* 2009) un astoņas dažādu proteobaktēriju lipāžu sekvences (q9rby1_9psed, q6gx93_psefl, q59933_serma, q76d31_psefl, lipb_psefl, q93hr6_9psed, q9zsg91_psefl, q52515_psesp; T1SS; Angkawidjaja un Kanaya 2006).

Diskriminantfunkciju pielietojuma iespēju pārbaudei izmatoti proteobaktēriju *Pseudomonas syringae* B728a, *Helicobacter pylori* 26695 un *Zymomonas mobilis* ZM4 eksoproteomi, klasificējamo sekvenču kopā iekļaujot, attiecīgi, 51, 39 un 73 olbaltumvielas, kuru garums ir vismaz 161 aminoskābes atlikumi (īsākās sekvences garums izstrādes kopā) un kuras Genome Atlas datubāzē (<http://www.cbs.dtu.dk/>) ir norādītas kā *secretomeP* olbaltumvielas (Bendtsen *et al.* 2005).

2.3. Aminoskābju frekvenču, kodonu izlietojuma un nukleotīdu satura un sekvenču kompozicionālo raksturlielumu aprēķināšana

Aminoskābju un kodona izlietojuma frekvences aprēķinātas, attiecīgi, ar ExPASy/ProtParam (Gasteiger *et al.* 2003) un Emboss/Cusp (Rice *et al.* 2000) metodēm, un identiski rezultāti iegūstami, raksturojot olbaltumvielu aminoskābju sastāvu, gan kā “aminoskābju frekvences sekvencē”, gan kā “sinonīmo kodonu frekvenču summa, kas normēta uz 1000 kodoniem”.

Nukleotīdu saturs, GC- un AT- asimetrijas (Rocha *et al.* 2006) un pirimidīnu-purīnu attiecības noteikšanai, aprēķināts no kodonu izlietojuma tabulām, reizinot katra kodona frekvences, kas normētas uz 1000 kodoniem (% uz 1000), ar attiecīgo bāžu skaitu kodona pozīcijā(s). Efektīvo kodonu skaits N_c (Wright 1990) un kodonu adaptācijas indekss CAI (Sharp un Li 1987) aprēķināts ar Emboss/Chips un Emboss/Cai metodēm (Rice *et al.* 2000), izmantojot „Ecoli.cut” kā atskaites (*reference set*) kodonu izlietojumu.

Olbaltumvielu aminoskābju un kodonu izlietojuma frekvencēm aprēķināta informācijas entropija, un novērtēta tās starpība ar teorētiski iespējamo (attiecīgi, 4.32 un 5.83 biti; Cover un Thomas 1991).

Aminoskābju grupu pārstāvības proporcijas aprēķinātas atbilstoši to iedalījumam pēc fizikāli ķīmiskajām īpašībām (Paila *et al.* 2008) vai to lomas olbaltumvielas otrējās struktūras izveidē (Xia un Xie 2002): hidrofobās (Gly, Ala, Val, Leu, Ile, Phe, Met, Pro, Trp), lādētās (Asp, Glu, Lys, Arg) un polārās (Ser, Thr, Tyr, His, Cys, Asn, Gln) kā arī α spirāles (Ala, Gln, Lys, Glu, Arg, Met, Leu, Asn, Asp), β plāksnes (Ser, His, Cys, Thr, Tyr, Phe, Trp, Ile, Val) un pagriezienus (*turn*; Gly, Pro) veidojošās aminoskābes. FYMINK/GARP indekss aprēķināts pēc aminoskābju iedalījuma AT- bagātajās (Lys, Met, Asn, Ile, Tyr, Phe) un GC- bagātajās (Gly, Ala, Arg, Pro) aminoskābēs (Foster *et al.* 1997).

GRAVI (Kyte un Doolittle 1982), molekulārā tilpuma (Soto un Tohac 1983), metaboliskās dārdzības (Akashi un Gojobori 2002), kompozicionālā vecuma (Sobolevsky un Trifonov 2005, Trifonov *et al.* 2006) un mutacionālās stabilitātes (Grantham 1974) indeksi aprēķināti vispārējai aminoskābju sastāva raksturošanai kā svērtais vidējais no produkta (aminoskābju frekvences) un faktora vērtības katrai aminoskābei attiecīgajā skalā.

2.4. Aminoskābju izkārtojuma periodisko īpašību un otrējās struktūras elementu sadalījuma prognozēšana

Sākotnējās datu kopas olbaltumvielu (55 sekvences, Publikācija I) terminālajiem fragmentiem (50 aminoskābju atlikumi) iegūti to fizikāli ķīmisko un strukturālo īpašību profili ar ExPASy/ProtScale metodi (Gasteiger *et al.* 2003), izmantojot tādas pazīmes kā hidropātija (Kyte un Doolittle 1982), polaritāte (Grantham 1974), α spirāles un β plāksnes (Levitt 1978).

Novērtējot vairāku pēc fizikāli ķīmiskajām īpašībām tuvinātu aminoskābju grupu izkārtojuma periodiskās īpašības (Publikācija II), aminoskābju izkārtojums olbaltumvielu C- un N- terminālajos fragmentos (64 aminoskābju atlikumi) noteikts ar Furjē analīzes (FT) metodi (Pasquier *et al.* 1998). Aminoskābes grupētas atbilstoši t.s. Deihofas alfabētam, kā arī FT metodē piedāvātajam iedalījumam (Dayhoff *et al.* 1978; Pasquier *et al.* 1998), kodējot olbaltumvielu sekvenču fragmentus skaitliskos vektoros, kur "1" un "0", attiecīgi, apzīmē grupai atbilstošās un

pārējās aminoskābes. Atbilstoši mazākajam iespējamajam FT analīzes mērogam (*embedded size*; $D = 2^4 = 64$) katram sekvenču fragmentam iegūti 19 ($n = 3 \div 21$) periodu profili ar tiem novērtēto intensitāti (I). Aminoskābju izvietojums ticami atbilst attiecīgajam periodam, ja $I \geq 2.99$ ($p < 0.05$). Iegūtās periodu profilu intensitātes matricas un to pāru kombinācijas tika izmantotas kā skaitliski vektori neklasiskās sekrēcijas substrātu grupu diskriminatīvo pazīmju kopas atlasei.

Olbaltumvielu C- un N- terminālo fragmentu (64 aminoskābju atlikumi) otrējā struktūra prognozēta ar 3D-PSSM metodi (Kelley *et al.* 2000; Publikācija III). Iegūtie otrējo struktūru elementu profili pārkodēti, izmantojot nevienādas skaitliskās vērtības (*property values*) katram struktūrelementam: „1” α spirālēm (H), „2” β spirālēm (*strand*; E) un „3” nestrukturētiem reģioniem (*coils*; C). Prognozēto otrējās struktūras elementu (H, E, C) konsensus profili katra sekrēcijas tipa olbaltumvielu grupām noteikti ar metodi “Shannon Entropy one” (Publikācija III).

2.5. Svešas izcelsmes gēnu olbaltumvielu sekvenču noteikšana

Olbaltumvielu iespējamā izcelsme no horizontālās gēnu pārnesei (HGT) vai to atbilstība nezināmas izcelsmes (*alien*) sekvenču novērtēta, kvantificējot GC satura atšķirības olbaltumvielas kodējošo sekvenču kodonu 3. pozīcijā (GC3) un attiecīgās (vai vistuvāk radniecīgās) proteobaktērijas genoma vidējo GC3, papildus novērtējot atšķirības katras aminoskābes relatīvajam daudzumam sekvenču salīdzinājumā ar attiecīgā proteoma vidējo aminoskābju sastāvu (Garcia-Vallvé *et al.* 2000). Par horizontāli pārnestām sekvenču uzskatāmas tās, kuru GC3 satura un genoma GC3 satura starpība vismaz 1.5 reizes pārsniedza attiecīgā genoma GC3 satura standartnovirzi un kuru aminoskābju satura un atbilstošā organisma aminoskābju satura starpība katrai aminoskābei iekļaujas organisma aminoskābju satura trīs standartnoviržu robežās (Garcia-Vallvé *et al.* 2000). Par svešas izcelsmes sekvenču uzskatāmas tās, kurām vismaz vienas aminoskābes satura starpība ar tās saturu attiecīgajā organismā vismaz trīs reizes pārsniedza standartnovirzi aminoskābes saturam organismā (Garcia-Vallvé *et al.* 2000). Pētījumā izmantotajām olbaltumvielām atbilstošo proteobaktēriju sugu genomu GC satura, kodonu izlietojuma un aminoskābju sastāva raksturlielumi iegūti Horizontālās gēnu pārnesei datubāzē (HGT-DB; Garcia-Vallvé *et al.* 2003). *Serratia*, *Proteus*, *Edwardsiella* un *Azotobacter* sugu gadījumos izmantotas filoģenētiski tuvu baktēriju genomu un proteomu raksturlielumu vidējās vērtības, aizvietojošas HGT-DB iztrūkstošus datus.

2.6. Sekrēcijas substrātu grupu īpašību salīdzināšana

Kodonu izlietojuma un aminoskābju raksturlielumu atšķirības olbaltumvielās, kas grupētas pēc to sekrēcijas tipa, novērtētas ar neparametrisko Kruskala-Valisa dispersijas analīzi (*Kruskal-Wallis test*) un / vai Frīdmana testu (*Friedman's test*) kombinācijā ar grupu pāru mediānu salīdzināšanu pēc Vilkoksona kritērija. Iespējamās parametru lineārās saistības novērtētas korelācijas un regresijas analīzē.

Prognozēto otrējās struktūras elementu un aminoskābju grupu izvietojuma periodu sadalījuma atšķirības sekrēcijas substrātu C- un N- terminālajos fragmentos un to grupās noteiktas ar standarta dispersijas analīzi (MANOVA), atšķirību būtiskumu novērtējot pēc Fišera kritērija.

2.7. Diskriminantanalīze

Diskriminantanalīzē pakāpeniskas tiešās un reversās atlasē ceļā, iteratīvi izvēloties mainīgo kombinācijas, sākot no viena vai visiem mainīgajiem, attiecīgi, palielinot vai samazinot mainīgo skaitu (*stepwise forward and backward selection*), līdz tiek izvēlētas efektīvu neatkarīgo mainīgo kombinācijas (Klecka 1980, Chou un Zhang 1995), kas piemērotas anotētu āršūnas olbaltumvielu atšķiršanai pēc tām atbilstošā sekrēcijas tipa. Iegūto diskriminantfunkciju precizitāte pārbaudīta ar Press's Q metodi (Chan 2005), aprēķinot Q indeksu pēc formulas $Q = [N - (n * K)]^2 / N(K - 1)$, kur N - kopējais datu kopas apjoms, n - pareizi klasificēto gadījumu skaits, K - grupu skaits, Q - Hī kvadrāts ($v=1$; $Q \leq 3.84$, $p \leq 0.05$; $Q \leq 6.64$, $p \leq 0.01$), un k-reizes atkārtoto vairākkopu krosvalidāciju (*k-fold crossvalidation* KCV), t.sk. LOOCV krosvalidāciju (*leave-one-out cross validation*), kuras gadījumā k ir vienāds ar kopējo paraugu skaitu oriģinālajā datu kopā (Leary *et al.* 2004; Publikācija II). KCV metodes ietvaros datu kopu k reizes pēc nejaušības principa sadala mazākās datu kopās, kurās proporcionāli pārstāvētas visas attiecīgās olbaltumvielu grupas. Pārbaudot klasifikācijas efektivitāti katrai no k- kopām, atlikušās sekvences (k-1) sekvences uzskata par izstrādes kopu (*training set*), un vidējo kļūdu intensitāti aprēķina katra k sadalījuma rezultātiem atsevišķi, ko apkopo kopējā KCV kļūdu intensitātē. Iegūto diskriminantfunkciju, jeb, klasifikatoru, izšķiršanas spēks un precizitāte (LOOCV rezultātam) salīdzināta pēc Mak Nemara metodes (*McNemar's test*; Fielding un Bell 1997).

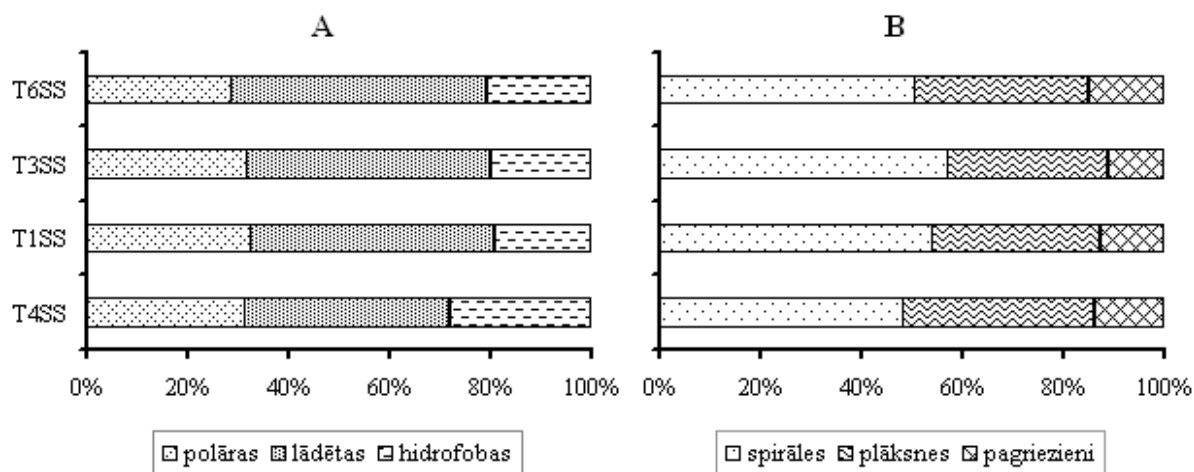
2.8. Datorprogrammas

Statistiskie aprēķini veikti datorprogrammās SPSS 11.0 (SPSS Inc., IL., ASV) un Statgraphics® Plus (Manugistics Inc., Mar., ASV).

3. Rezultāti

3.1. Vispārējo sekvenču raksturlielumu atšķirības neklasiskās sekrēcijas substrātos

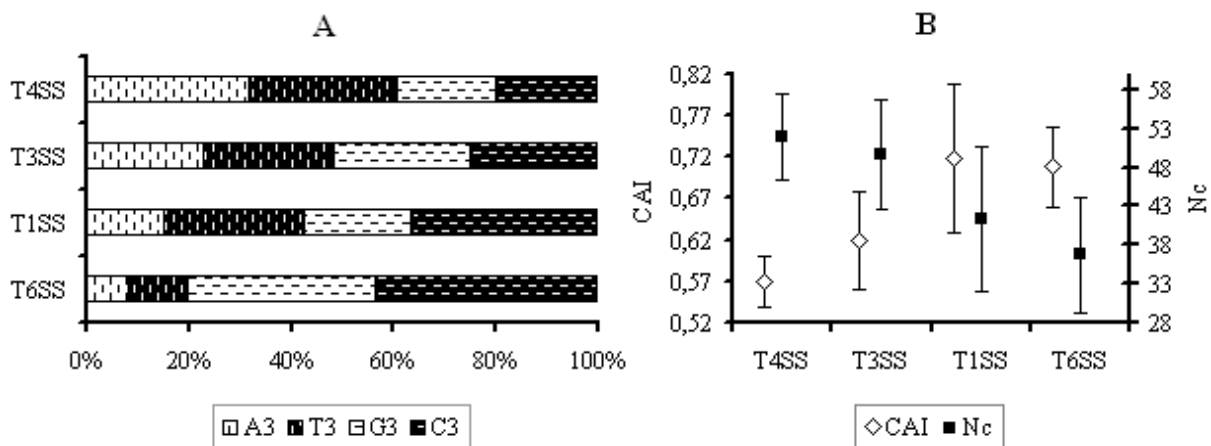
T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu sekvenču grupu aminoskābju sastāva neparametiskā dispersijas analīze uzrādīja vairākas būtiskas, no sekrēcijas tipa atkarīgas, atšķirības ($p < 0.001$, Kruskala-Valisa tests), gan sākotnēji ierobežotajā (42 sekvences), gan paplašinātajā (111 sekvences) T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu sekvenču kopā (1. Att. Publikācijā I; 1. Tab. Publikācijā IV). Tā, salīdzinot aminoskābju sastāva raksturlielumus olbaltumvielu grupu pāros, vismaz trīs no sešām iespējamām kombinācijām būtiski atšķiras ($p < 0.05$, Vilkoksona kritērijs) ar augstāko izšķirtspēju (sešas būtiskas starpības) α spirāles veidojošo aminoskābju kopas frekvenču gadījumā (3. Att.; 1. Tab.; Publikācija IV).



3. att. Polāro, lādēto, hidrofobo (A) un α spirāles, β plāksnes un pagriezienus veidojošo (B) aminoskābju sadalījums T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupās.

Fig. 3 Content of polar, charged and hydrophobic (A) and α helix, β sheet and turn forming (B) amino acids in groups of proteins secreted by T1SS, T3SS, T4SS and T6SS.

Līdzīgi, visi pētījumā izmantotie sekrēcijas substrātus kodējošo sekvenču nukleotīdu satura un kodonu izlietojuma raksturlielumi (izņemot kopējās GC- asimetrijas indeksu) uzrāda samērojamas, lai arī nevienādas olbaltumvielu grupu atšķirības ($p < 0.001$, Kruskala-Valisa tests), tostarp sešas olbaltumvielu grupu pāru kombinācijas atšķirot ($p < 0.05$, Vilkoksona kritērijs) pēc adenīna satura un kopējā nukleotīdu sastāva asimetrijas kodonu 3. pozīcijā (A3, AT3- asimetrija; 4. Att.; 1. Tab.; 1. Tab. Publikācijā IV).



4. att. Nukleotīdu sastāvs kodonu 3. pozīcijā (A) un kodonu adaptācijas indekss CAI un efektīvo kodonu skaits Nc (vidējais ± SK) (B) T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupās.

Fig. 4 Nucleotide content in 3rd position of codon sequences (A) and the codon adaptation index CAI and the number of effective codons Nc (average ± SE) (B) in groups of proteins secreted by T1SS, T3SS, T4SS and T6SS.

Pētījumā izmantotie kodonu izlietojuma raksturlielumi uzrāda savstarpēji samērojamus, tomēr ne identiskus rezultātus (4. Att. B; 1. Tab.; 1. Tab. Publikācijā IV). Proti, T3SS un T4SS sekrēcijas substrātu grupas nav atšķiramas pēc sekvenču Efektīvo kodonu skaita (Nc), kamēr T1SS un T6SS grupas ir vienādas pēc Kodonu adaptācijas indeksa (CAI; abos gadījumos $p > 0.05$, Vilksona kritērijs). Līdzīgi atšķiras rezultāti, arī salīdzinot grupu pārus pēc citiem aminoskābju un kodonu izlietojuma raksturlielumiem (1. Tab.).

3.2. Aminoskābju un kodonu izlietojuma diskriminatīvā potenciāla salīdzinājums

Pārbaudot iespējas atšķirt, vismaz principā, proteobaktēriju sekrēcijas sistēmu substrātus pēc katras to aminoskābju un kodonu frekvences, T1SS, T3SS, T4SS un T6SS sekrēcijas substrāti tika attiecīgi apzīmēti kā 1., 2., 3. un 4. atkarīgo mainīgo grupa, par neatkarīgajiem (prognozējošajiem) mainīgajiem multiplā-4-grupu diskriminantanalīzē izmantojot katru pa vienai no 61 kodona un 20 aminoskābju sinonīmo kodonu frekvencēm. Visi kodoni, izņemot CAT (His) un 15 aminoskābes (Gln, Met, Leu, Pro, Glu, Ser, Arg, Lys, Cys, Trp, Asn, Phe, Tyr, Gly, Asp) uzrādīja augstāku klasifikācijas precizitāti salīdzinot ar nejaušam gadījumam atbilstošo ($Q > 3.84$; $p < 0.05$), tā norādot uz iespējamām kodonu izlietojuma priekšrocībām indicēt būtiskās atšķirības četru neklasiskās sekrēcijas tipu substrātu sekvencēs (5. Att.).

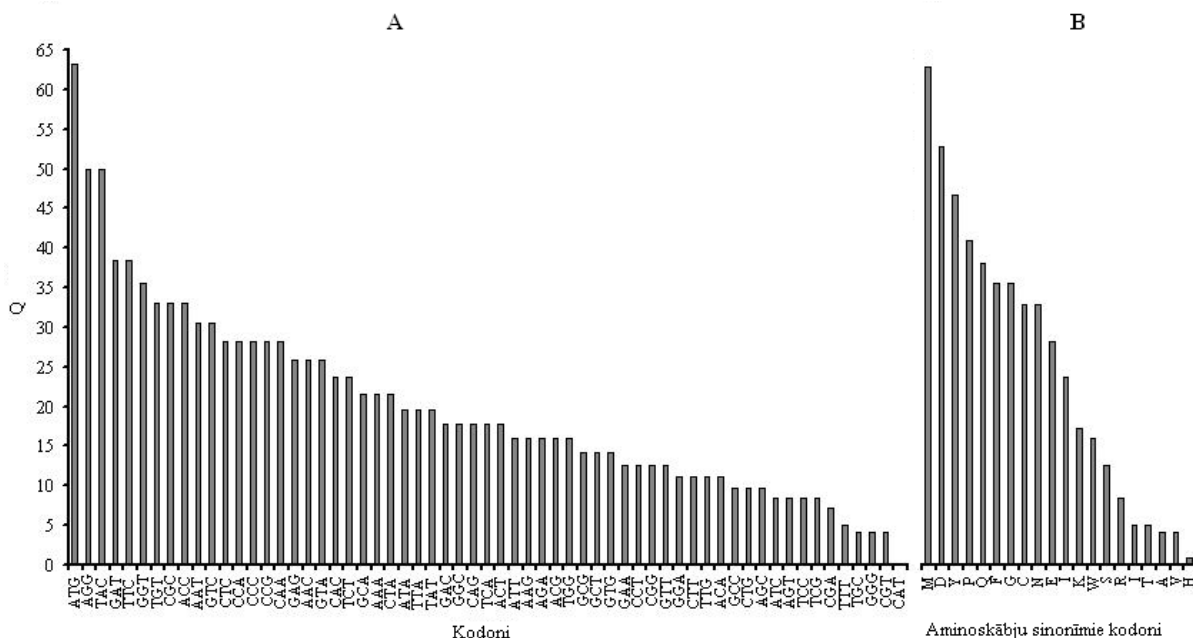
Aminoskābju un nukleotīdu satura un kodonu izlietojuma salīdzināšana T1SS, T3SS, T4SS un T6SS sekrēcijas substrātos (attiecīgi, 1., 2., 3. un 4. grupa) pēc Vilksona kritērija.

Apzīmējumi: grupu pāru mediānu atšķirība „+” ir būtiska ($p < 0.05$), „N” - nav ($p > 0.05$) būtiska.

Result of pair-wise comparison (Mann-Whitney test) between compositional properties of the T1SS, T3SS, T4SS and T6SS secreted proteins (group 1, group 2, group 3 and group 4, respectively) in terms of amino acid, nucleotide content and codon usage indices.

Abbreviations: “+” for significant ($p < 0.05$) and “N” for absence of significant differences ($p > 0.05$) between medians of two groups.

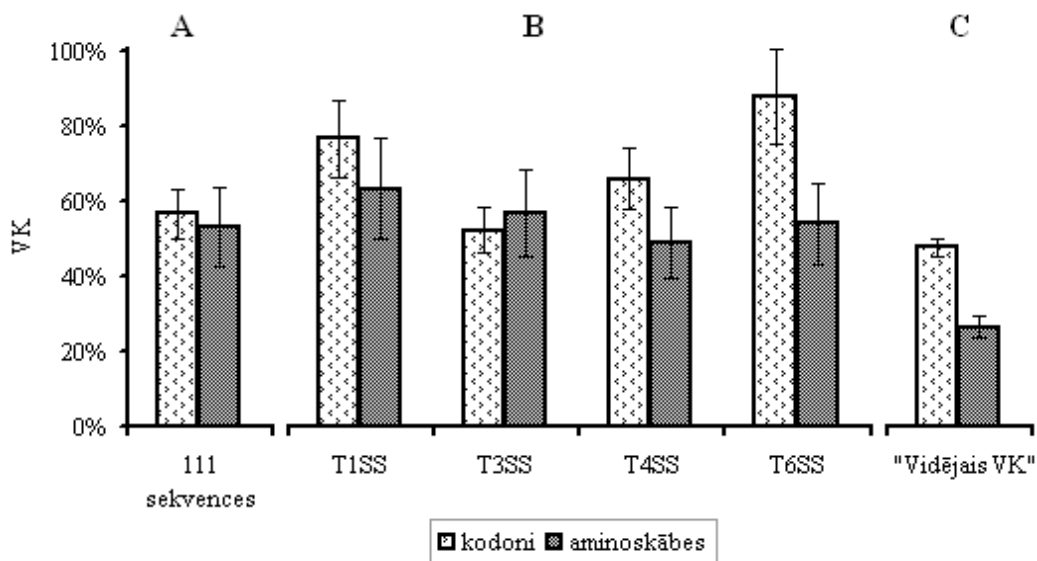
Sekvenču raksturlielumi <i>Characteristics of sequence composition</i>	Olbaltumvielu grupu pāri <i>Pairs of protein groups</i>					
	1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
Nukleotīdu satura un kodonu izlietojuma raksturlielumi						
CAI	+	+	N	+	+	+
Nc	+	+	+	N	+	+
Pirimidīnu-purīnu attiecība	+	+	N	+	+	+
GC (kopējais)	N	+	+	+	+	+
GC1	N	N	N	+	+	+
GC2	N	N	+	+	N	+
GC3	N	+	+	+	+	+
A (kopējais)	N	+	+	+	+	+
C (kopējais)	N	N	+	N	+	+
G (kopējais)	N	+	+	+	+	+
T (kopējais)	N	N	+	N	+	+
A3	+	+	+	+	+	+
C3	+	+	N	N	+	+
G3	+	N	+	+	+	+
T3	N	N	N	+	+	+
GC asimetrija (kopējā)	N	N	N	N	N	N
AT1 asimetrija	+	+	N	N	+	+
AT2 asimetrija	+	+	+	+	N	+
AT3 asimetrija	+	+	+	+	+	+
Aminoskābju satura raksturlielumi						
α spirāles veidojošo aminoskābju saturs	+	+	+	+	+	+
GRAVI indekss	N	+	N	+	N	+
Glicīna saturs, %	+	+	+	+	+	+
Kompozicionālais vecums	N	+	N	+	N	+
Molekulārais tilpums	+	+	+	N	N	N
Mutacionālā stabilitāte	+	+	+	N	+	+
FYMINK/GARP indekss	N	N	+	N	+	+
Informācijas entropija						
Kodoni	+	+	N	N	+	+
Aminoskābes	+	N	N	N	+	N



5. att. T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu multiplās diskriminantanalīzes Press's Q indekss individuālu kodonu izlietojuma frekvenču (A) un aminoskābju sinonīmo kodonu kopu frekvenču (B) gadījumā.
 Fig. 5 Press's Q statistics for multiple four-group discrimination of T1SS, T3SS, T4SS and T6SS secreted proteins by means of individual normalized codon frequencies (A) and summarized normalized frequencies of synonymous codon families, i.e., amino acid frequencies of occurrence (B).

3.3. Kodonu un aminoskābju izlietojuma variācija un informācijas entropija

Negrupētā 111 olbaltumvielu sekvenču kopā, kā arī T1SS, T3SS, T4SS un T6SS substrātu grupās atsevišķi, kvantificētas katra atsevišķa kodona un aminoskābes sinonīmo kodonu izlietojuma atšķirības pēc to frekvenču variācijas koeficientiem (turpmāk – VK). Kodona un aminoskābju kopējā variācija starp sekrēcijas substrātu grupām aprēķināta kā “vidējais VK” no katra kodona un aminoskābes VK-iem sekrēcijas tipa ietvaros. Negrupētu 111 sekvenču kopai novēroti savstarpēji samērojami individuālo kodonu un aminoskābju sinonīmo kodonu izlietojuma VK-i (attiecīgi, $56.7 \pm 6.6 \%$ un $53.3 \pm 10.6 \%$; 6. Att. A). Līdzīga tendence novērota arī sekrēcijas substrātu grupās atsevišķi (6. Att. B). Tomēr kodonu izlietojuma kopējā variācija starp sekrēcijas substrātu grupām būtiski ($p < 0.001$, Vilkoksona kritērijs) pārsniedza aminoskābju izlietojuma “VK vidējo vērtību” (6. Att. C). Tas norāda uz lielāku asimetriju tieši individuālu kodonu izlietojuma frekvenču gadījumā, uzsverot iespējamās priekšrocības to izmantošanai klasificējot neklasiskās sekrēcijas substrātu grupas.



6. att. Kodonu un aminokābju izlietojuma frekvenču variācijas koeficients ($VK \pm S_{VK}$) nēgrupētu 111 olbaltumvielu sekvencēs (A); T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupās (B) un sekrēcijas substrātu grupu „vidējais VK \pm SK” (C).

Fig. 6 Coefficient of variation (CV) of amino acid and codon usage frequencies: within ungrouped set of 111 nonclassically secreted proteins (A); within the groups of proteins secreted by T1SS, T3SS, T4SS and T6SS (B) and “average of CVs \pm SE” among groups of secretion substrates (C).

Līdzīgi, nēgrupētu 111 sekvenču kopas individuālajiem kodoniem novērota augstāka informācijas entropija nekā aminokābju sinonīmo kodonu izlietojuma gadījumā, attiecīgi, 390.93 ± 29.98 biti (5.21 ± 0.03 biti uz atlikumu (*bits per residue*)) un 212.56 ± 14.03 biti (3.99 ± 0.01 biti uz atlikumu). Tāpat, sekrēcijas substrātu grupās novērotas (Kruskala-Valisa tests) būtiskas kodonu un aminokābju sinonīmo kodonu izlietojuma informācijas entropijas vienību atšķirības (attiecīgi, $p < 0.001$ un $p < 0.05$). Likumsakarīgi, ka trīs un divi no sešiem iespējamajiem sekrēcijas substrātu grupu pāriem būtiski atšķirās ($p < 0.05$; Vilksona kritērijs), attiecīgi, pēc to kodonu un aminokābju izlietojuma informācijas entropijas vienībām (1. Tab.). Tādējādi individuālu kodonu izlietojuma frekvenču gadījumā novērotā augstāka variabilitāte, arī augstākas entropijas vienības var norādīt uz to iespējami augstāko jutību, tātad būtiskām priekšrocībām atšķirt neklasiskās sekrēcijas substrātu grupas.

3.4. Efektīvo kodonu un aminokābju atlase 4-grupu diskriminantanalīzei

Ievērojamās aminokābju satura un kodonu izlietojuma atšķirības T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupās (3. Att.; 4. Att.; 1. Tab.; 1. Tab. Publikācijā IV) norādīja, uz

iespēju šādi vispārinātus sekvenču raksturlielumus izmantot kā diskriminatīvās pazīmes neklasiski sekretējamu olbaltumvielu klasifikācijai. Multiplā diskriminantanalīzē ar iteratīvo tiešo (*forward*) un reverso (*backward*) atlasē metodi, izmantojot divus datu vektorus (61x111 un 20x111), kuri, attiecīgi, pārstāv 111 olbaltumvielu normētās 61 kodona un 20 aminoskābju sinonīmo kodonu frekvences, tika izdalītas divas efektīvo neatkarīgo mainīgo kopas, kuras vektorā „61x111” pārstāv 24 kodonu frekvences (7. Att.), un daļēji attiecināmas uz to kodētajām 16 aminoskābēm (Cys, Asp, Gly, Ile, Lys, Leu, Met, Asn, Pro, Gln, Arg, Ser, Thr, Val, Trp, Tyr).

		Otrā pozīcija								
		T	C	A	G					
Pirmā pozīcija	T	TTT	Phe (F)	TCT	Ser (S)	TAT	Tyr (Y)	TGT	Cys (C)	T
		TTC		TCC		TAC		TGC		C
		TTA	Leu (L)	TCA	TAA	Stop	TGA	Stop	A	
		TTG		TCC	TAG		TGG	Trp (W)	G	
	C	CTT	Leu (L)	CCT	Pro (P)	CAT	His (H)	CGT	Arg (R)	T
		CTC		CCC		CAC		CGC		C
		CTA		CCA		CAA	CGA	A		
		CTG		CCG		CAG	CGG	G		
	A	ATT	Ile (I)	ACT	Thr (T)	AAT	Asn (N)	AGT	Ser (S)	T
		ATC		ACC		AAC		AGC		C
		ATA		ACA		AAA	AGA	A		
		ATG	Met (M)	ACG		AAG	Lys (K)	AGG	Arg (R)	G
	G	GTT	Val (V)	GCT	Ala (A)	GAT	Asp (D)	GGT	Gly (G)	T
		GTC		GCC		GAC		GCC		C
		GTA		GCA		GAA	GGA	A		
		GTG		GCG		GAG	GGG	G		

A - efektīvie mainīgie

7. Att. No aminoskābju un kodonu izlietojuma frekvenču profilu vektoriem izvēlēto efektīvo mainīgo kopu attēlojums ģenētiskā kodā tabulā.

Fig. 7 The sets of predictor variables selected from the data vectors of amino acid and codon usage frequencies visualized within the table of the genetic code.

Līdzīgā veidā no vektora „20x111” izdalītas 10 aminoskābes kodējošo sinonīmo kodonu kopējās, t.i., aminoskābju sastāvu raksturojošās, frekvences (7. Att.). Abu izvēlēto efektīvo mainīgo kopu diskriminantfunkciju augstie kvalitātes rādītāji (2. Tab.; 3. Tab.) un LOOCV krosvalidācijā 97.3% un 92.8% atbilstoši klasificētu olbaltumvielu ($Q=309.43$ un $Q=272.08$; $p<0.001$) ievērojami pārsniedza nejaušā gadījumā sasniedzamo precizitāti ($Q>3.84$; $p<0.05$).

Diskriminantfunkciju kvalitātes rādītāji efektīvo mainīgo kopām, kas izvēlētas no aminoskābju un kodonu izlietojuma frekvenču profilu vektoriem.

- a Prognozējošie mainīgie, kas izvēlēti 61 kodona frekvencēm
 b Prognozējošie mainīgie, kas izvēlēti 20 aminoskābju sinonīmo kodonu frekvencēm
 c Trīs diskriminantfunkcijas multiplai-4-grupu diskriminantanalīzei
 d 1. funkcija pret 3. funkciju, 2. funkcija pret 3. funkciju, 3. funkcija
 e Diskriminantes punkta vērtības
 f Atkarīgo mainīgo grupas: 1. grupa – T1SS, 2. grupa – T3SS, 3. grupa – T4SS, 4. grupa – T6SS sekretācijas substrāti

The quality measures of discriminant functions for the sets of predictor variables selected from the data vectors of amino acid and codon usage frequencies.

- a The predictor variables selected from codon frequencies
 b The predictor variables selected from summarised frequencies of synonymous codons for ten amino acids
 c Three discriminant functions for multiple 4 group discrimination.
 d Function 1 through function 3, function 2 through function 3, function 3
 e Mean discrimination scores.
 f Groups of dependent variables: group 1 – T1SS, group 2 – T3SS, group 3 – T4SS, group 4 – T6SS secreted proteins.

Kvalitātes rādītāji <i>Quality measures</i>	Kodoni ^a <i>Codons</i>			Aminoskābes ^b <i>Amino acids</i>		
	Funkcija ^c <i>Function</i>			Funkcija ^c <i>Function</i>		
	1	2	3	1	2	3
Korelācijas koeficients (bp)	0.96 (72)	0.91 (46)	0.80 (22)	0.93 (30)	0.71 (18)	0.65 (8)
Dispersija, %	65.00	25.40	9.50	77.70	12.90	9.40
Eigen vērtība	12.37	4.84	1.81	5.98	0.99	0.72
Vilksa λ^d	0.01	0.06	0.36	0.04	0.29	0.58
Hī-kvadrāts ^d	517.64	268.72	99.33	327.08	127.01	55.89
Centroīdi^e						
1. grupa ^f	-4.37	-1.46	-0.85	-3.43	-0.12	0.53
2. grupa	2.91	1.36	-1.02	2.03	-0.97	0.12
3. grupa	4.11	-3.91	1.48	2.35	1.95	0.83
4. grupa	-2.06	2.42	2.06	-0.49	0.55	-1.60

24 efektīvo kodonu kopu raksturoja nedaudz augstāki diskriminantfunkciju kvalitātes rādītāji (2. Tab.; 3. Tab.), kas atbilst novērotajām augstākām VK vērtībām (6. Att. C). Taču, tā kā abu efektīvo mainīgo kopu (t.i., klasifikatoru) kļūdu intensitāte (LOOCV krosvalidācijas gadījumā) pēc Mak Nemara metodes neatšķiras ($\chi=1.78$, $p>0.05$, $n=111$; 4. Tab.), to efektivitāte ir uzskatāma par statistiski vienlīdz samērojamu.

Diskriminantfunkciju precizitātes novērtējums ar k - reizes atkārtoto vairākkopu krosvalidāciju (KCV), klasificējot T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupas.

a, b – sk. 2. tab. apzīmējumus a, b

c $k=111$

Estimates of prediction accuracy obtained by k - fold cross-validation (KCV) with varying folds of original dataset for type T1SS, T3SS, T4SS and T6SS secreted proteins.

a, b as a, b in Table 2

c $k=111$

Prognozējošie mainīgie <i>Predictor variables</i>	Kļūdu intensitāte ± SK <i>Error rate ± SE</i>	
	Kodoni ^a <i>Codons</i>	Aminoskābes ^b <i>Amino acids</i>
Iterāciju skaits k <i>Number of folds k</i>		
3	0.078 ± 0.036	0.181 ± 0.053
4	0.075 ± 0.041	0.126 ± 0.042
6	0.090 ± 0.034	0.109 ± 0.039
8	0.083 ± 0.033	0.082 ± 0.030
10	0.021 ± 0.012	0.068 ± 0.023
LOOCV ^c	0.027 ± 0.016	0.072 ± 0.025
vidēji	0.062 ± 0.012	0.106 ± 0.015

T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu diskriminācijai izvēlēto efektīvo mainīgo salīdzinājums Mak Nemara metodes 2x2 matricā.

a, b – sk. 2. tab. apzīmējumus a, b

The cross-table and test statistics of McNemar's test for comparison of the predictor variables for discrimination among groups of T1SS, T3SS, T4SS and T6SS secreted proteins.

a, b as a, b in Table 2

Aminoskābes ^b <i>Amino acids</i>	Kodoni ^a <i>Codons</i>	
	Patiesi <i>True</i>	Kļūdaini <i>False</i>
Patiesi	101	2
Kļūdaini	7	1

Abu klasifikatoru LOOCV krosvalidācijas rezultāts katrai olbaltumvielu grupai apkopots neskaidrību matricās, atsevišķi izdalot patiesi un kļūdaini pozitīvos/negatīvos klasifikācijas gadījumus (5. Tab.). Abu izvēlēto efektīvo mainīgo kopu (1. Att.) diskriminantfunkciju augsto precizitāti, kas būtiski pārsniedz nejaušā gadījumā sagaidāmo, apstiprināja no neskaidrību matricām atvasinātie diskriminantanalīzes efektivitātes raksturlielumi un Press's Q kritērijs (3. Tab.).

Klaisifikatoru kļūdu intensitātes T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupā salīdzinājums neskaidrības matricā.

a, b – sk. 2. tab. apzīmējumus a, b

The confusion matrices for comparison of the error rates of both classifiers accordingly to individual group of T1SS, T3SS, T4SS and T6SS secreted proteins.

a, b as a, b in Table 2

Definētā piederība grupai <i>Known class</i>	Prognozētā piederība grupai <i>Predicted class</i>							
	Kodoni ^a <i>Codons</i>				Aminoskābes ^b <i>Amino acids</i>			
	T1SS	T3SS	T4SS	T6SS	T1SS	T3SS	T4SS	T6SS
T1SS	32	0	0	0	32	0	0	0
T3SS	0	39	1	1	0	36	3	2
T4SS	0	0	16	0	0	0	16	0
T6SS	1	0	0	21	1	1	1	19

Efektīvo mainīgo kopu (7. Att.) samērojami augstā jutība, nosakot T1SS un T4SS sekrēcijas substrātu sekvences (5. Tab.; 6. Tab.) apstiprināja vienādi augsto efektivitāti šo attiecīgo tipu sekrēcijas substrātu atpazīšanā. Savukārt, augstais T3SS sekrēcijas substrātu

No neskaidrību matricām atvasinātie raksturlielumi.

a, b – sk. 2. tab. apzīmējumus a, b

Summary statistics derived from the confusion matrices.

a, b as a, b in Table 2

Raksturlielums <i>Measure of statistics</i>	Efektīvie mainīgie <i>Effective variables</i>	Olbaltumvielu grupa <i>Protein group</i>			
		T1SS	T3SS	T4SS	T6SS
Precizitāte	Kodoni ^a	0.97	1	0.94	0.96
	Aminoskābes ^b	0.97	0.97	0.80	0.91
Jutība	Kodoni ^a	1	0.95	1	0.96
	Aminoskābes ^b	1	0.88	1	0.86
Specifiskums	Kodoni ^a	0.99	1	0.99	0.99
	Aminoskābes ^b	0.99	0.99	0.96	0.98
Kopējā klasifikatora kļūda	Kodoni ^a	0.01	0.02	0.01	0.02
	Aminoskābes ^b	0.01	0.05	0.04	0.05
Kappa indekss	Kodoni ^a	0.98	0.96	0.96	0.94
	Aminoskābes ^b	0.98	0.88	0.87	0.86
Press's Q indekss	Kodoni ^a	107.04	103.14	107.04	103.14
	Aminoskābes ^b	107.04	88.30	95.58	91.90

atpazīšanas specifiskums varētu norādīt uz 24 efektīvo kodonu kopas priekšrocībām (5. Tab.). Tomēr novērotā abu klasifikatoru skaitliski samērojamā efektivitāte (4. Tab.) un nevienāda

kļūdu sadalījums atsevišķās olbaltumvielu grupās (5. Tab.) ir nepieciešamais priekšnosacījums abu klasifikatoru (7. Att.) komplementārai izmantošanai klasificējot/prognozējot sekrēcijas substrātus.

3.5. Neatkarīga olbaltumvielu sekvenču kopa

Samērojama abu klasifikatoru (7. Att.) precizitāte ($p > 0.05$, Mak Nemara kritērijs) tika novērota arī neatkarīgas olbaltumvielu sekvenču kopas gadījumā (7. Tab.), atbilstoši klasificējot dažādu proteobaktēriju lipāzes un 76.5% *Salmonella typhimurium* T3SS sekrēcijas substrātu.

7. Tabula/ Table 7

Izvēlēto efektīvo mainīgo kopu veiktspējas klasificēt neatkarīgu sekrēcijas substrātu sekvenču kopu salīdzinājums Mak Nemara metodes 2x2 matricā.

a, b – sk. 2. tab. apzīmējumus a, b

The cross-table of McNemar's test for comparison of the performance of selected predictor variables to classify the independent testing set of secreted proteins.

a, b as a, b in Table 2

Aminoskābes ^b <i>Amino acids</i>	Kodoni ^a <i>Codons</i>	
	Patiesi <i>True</i>	Kļūdaini <i>False</i>
Patiesi	34	3
Kļūdaini	2	3

3.6. Efektīvo kodonu un aminoskābju kombināciju atlase lineārai grupu pāru diskriminantanalīzei

No aminoskābju un kodonu izlietojuma vektoriem tika papildus izvēlētas citas efektīvo mainīgo kombinācijas neklasiskās sekrēcijas substrātu grupu pāru diskriminācijai, tā pārbaudot, vai šādos gadījumos spīgtāk izpaudīsies iepriekš novērotās nebūtiskās kodonu izlietojuma frekvenču priekšrocības, salīdzinot ar aminoskābju sinonīmo kodonu iespējām (4. Tab.) nodrošināt augstāku diskriminantanalīzes precizitāti.

Vismaz vienreiz kādā no sešu grupu pāru diskriminēšanas efektīvo mainīgo kopu gadījumiem izvēlēti 70% aminoskābju (8. Tab.). Dažādo atsevišķu aminoskābju nozīmi noteiktu sekrēcijas substrātu grupu diskriminācijai apstiprina apstākļi, ka divu aminoskābju (Arg un Asn)

Binārai T1SS, T3SS, T4SS un T6SS sekrecijas substrātu grupu diskriminācijai izvēlētās efektīvo mainīgo kopas un diskriminantfunkciju precizitāte.

a, b, c – sk. 2. tab. apzīmējumus a, b, f

d Oriģināli grupēto gadījumu klasificēšanas precizitāte (%).

e LOOCV krosvalidācijas precizitāte (%).

The predictor variables selected by backward and forward stepwise procedure for binary discrimination among groups of T1SS, T3SS, T4SS and T6SS secreted proteins.

a, b, c as a, b, f in Table 2

d % of original grouped cases correctly classified.

e % of cross-validated grouped cases (LOOCV) correctly classified.

Grupās ^c <i>Groups</i>	Kodoni ^a <i>Codons</i>	Klasifikācijas precizitāte ^{d,e} <i>Classification accuracy</i>	Aminoskābes ^b <i>Amino acids</i>	Klasifikācijas precizitāte ^{d,e} <i>Classification accuracy</i>
	Prognozējošie mainīgie <i>Predictor variables</i>		Prognozējošie mainīgie <i>Predictor variables</i>	
1 & 2	GCG, GAG, ATA, CTC, CTT, TTA, ATG, AAT, CCC, CCT, CAA, CAG, AGA, TCA, TCT, ACC, TAC	100 / 100	R, E, G, I, M, P, S, Y, V	100 / 100
1 & 3	GCA, GCC, GCG, GCT, TGC, TGT, GAT, GAA, GAG, TTC, TTT, GGA, GGG, GGT, CAT, CTA, CTT, TTA, TTG, ATG, AAC, AAT, CCC, CCG, CCT, CAA, AGA, AGG, CGA, CGC, TGG, TAT	100 / 97.9	F, G, I, R, N, S, V	100 / 97.9
1 & 4	GCA, GAT, GAA, GGA, CAC, ATC, CTT, TTG, ATG, CCA, CAG, AGA, AGG, CGG, TCG, ACA, GTG	100 / 100	G, M, N, Q, R, Y	100 / 96.3
2 & 3	GCA, GCC, GCG, GCT, TGC, GAA, GAG, TTT, GGT, CAC, CAT, ATA, ATC, ATT, AAA, CTC, CTG, CTT, TTA, TTG, AAT, CCA, CCC, CCG, CCT, CAA, CAG, AGC, AGT, TCC, TCG, TCT, ACA, ACG	100 / 100	E, F, H, I, K, N, P, Q, R	98.2 / 94.7
2 & 4	GCA, GCT, TGT, GAC, GAA, GAG, TTC, GGG, GGT, CAC, CAT, ATA, ATC, ATT, AAA, AAG, CTG, CTT, TTG, ATG, AAC, AAT, CCC, CCG, CCT, CAA, AGA, AGG, CGA, CGC, CGG, CGT, AGC, TCA, TCC, TCT, ACC, ACG, ACT, GTC, GTG, TGG	100 / 100	E, G, M, N, Y	100 / 100
3 & 4	GCG, GCT, TGC, GAC, GAT, GAG, TTC, GGA, GGC, GGG, GGT, ATA, ATT, AAA, AAG, CTA, CTG, CTT, TTA, TTG, ATG, AAC, AGA	100 / 100	E, K, N, Q, R	100 / 100

frekvences tika izvēlētas kā diskriminatīvās pazīme piecu grupu pāru atšķiršanai, kamēr histidīns (Hys) ir izvēlēts tikai vienā gadījumā, proti, T3SS un T4SS sekrēcijas substrātu diskriminācijai. Turklāt vismaz vienā grupu pāra diskriminatīvo mainīgo gadījumā pārstāvētas deviņas no desmit (izņemot Trp) T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu multiplai diskriminācijai izvēlētajām efektīvajām aminoskābēm (7. Att.). Tanī pat laikā 96% kodonu ir izdalīti vismaz vienā no sešām efektīvo mainīgo kodonu kopām (8. Tab.), turklāt to skaitā neiekļāvās tikai viens (GTA) no efektīvajiem multiplās-4-grupu diskriminantanalīzes 24 kodoniem (7. Att.).

No kodonu izlietojuma frekvencēm atlasīto efektīvo mainīgo kopas piecos no sešiem grupu pāru kombināciju gadījumiem nodrošināja diskrimināciju ar 100% precizitāti (LOOCV gadījumā), kamēr aminoskābju sinonīmo kodonu efektīvo kombināciju gadījumā tika novērota lielāka kļūdu intensitāte (8. Tab.), tomēr Mak Nemara kritērijs norādīja uz aminoskābju un kodonu efektīvo mainīgo kopu kopējās precizitātes statistisko samērojamību ($p > 0.05$). Tāpat, arī grupu pāru un multiplās diskriminantanalīzes precizitātes izrādījās savstarpēji samērojamas ($p > 0.05$, Mak Nemara kritērijs), lai gan arī grupu pāri varētu tikt salīdzināti, lai precizētu atsevišķus neviennozīmīgus diskriminantanalīzes rezultātus (8. Tab.).

3.7. Efektīvo kodonu un aminoskābju kombināciju atlase pāru diskriminantanalīzē individuālu substrātu grupu atšķiršanai no apvienotās pārējo trīs sekrēcijas tipu substrātu kopas

Lai noskaidrotu, vai kādas sekrēcijas substrātu grupas kodonu un aminoskābju sastāvs izteikti atšķiras no citu sekrēcijas substrātu apvienotās kopas, no kodonu un aminoskābju sinonīmo kodonu frekvenču kopas tika papildus atlasītas tādas efektīvo mainīgo kopas, lai pāru diskriminantanalīzē varētu atšķirt katru sekrēcijas substrātu grupu no apvienotās pārējo trīs sekrēcijas tipu substrātu grupas. Kā multiplās, tā sekrēcijas substrātu grupu pāru diskriminācijas gadījumā novērota samērojama precizitāte (9. Tab.; 2. Tab.; 8. Tab.), lai gan ar būtiskām efektīvo kodonu priekšrocībām klasificējot/prognozējot T3SS un T6SS sekrēcijas substrātu grupas (9. Tab.).

Efektīvo mainīgo kopas un diskriminantfunkciju precizitāte un tās salīdzinājums pēc Mak Nemara kritērija bināras atsevišķu T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupu diskriminācijai pret pārējo triju sekrēcijas tipu substrātu apvienoto grupu.

a, b, d, e sk. 8. tab. apzīmējumus a, b, d, e

c Atkarīgo mainīgo grupas: 1. grupa – atsevišķa sekrēcijas substrātu grupa, 2. grupa – pārējo triju sekrēcijas tipu substrātu apvienotā kopa.

f Mak Nemara kritērija būtiskums; N=111.

The predictor variables quality measures and test statistics of McNemar's test for comparison of the predictor variables for binary discrimination of individual groups of T1SS, T3SS, T4SS and T6SS secreted proteins against combined sets of the other three protein groups.

a, b, d, e as a, b, d, e in Table 2

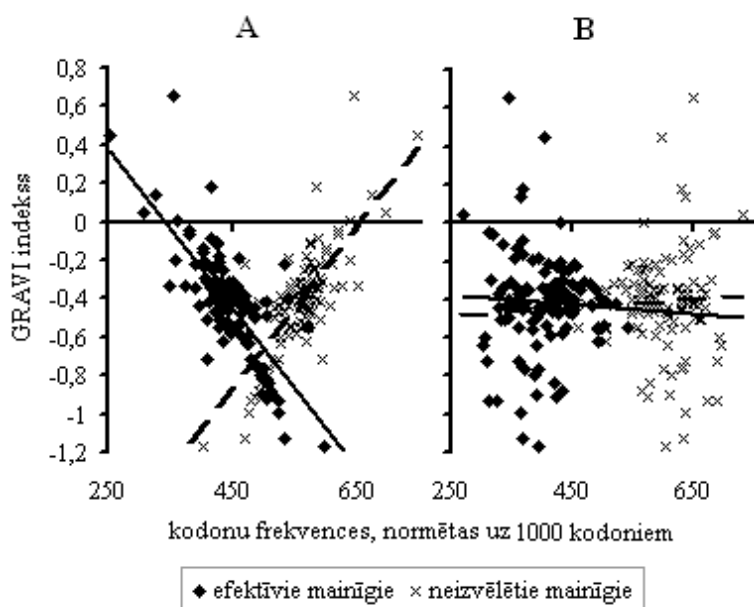
c Groups of dependent variables: group 1 - individual group of secreted proteins, group 2 – combined set of proteins from the other three protein groups

f Significance for statistics of McNemar's test; N=111

1. Grupa ^c <i>Group 1</i>	Kodoni ^a <i>Codons</i>	Klasifikācijas precizitāte ^{d,e} <i>Classification accuracy</i>	Aminoskābes ^b <i>Amino acids</i>		Būtiskums ^f <i>Sign.</i>
	Prognozējošie mainīgie <i>Predictor variables</i>		Prognozējošie mainīgie <i>Predictor variables</i>	Klasifikācijas precizitāte ^{d,e} <i>Classification accuracy</i>	
T1SS	GCC, GCG, GAA, GAG, GGT, CAC, AAA, ATG, AAT, CCA, CCC, CAG, TCG, ACC, GTT, TAC, TAT	100 / 100	Q, E, G, M, P, W, Y	99.1 / 99.1	>0.05
T3SS	GCT, TGT, GAC, GAA, TTC, GGC, CTC, TTA, ATG, CCA, AGG, CGC, CGT, AGT, TCA, ACC, ACG, GTC, TAC, TAT	99.1 / 99.1	A, N, D, C, Q, E, G, H, I, L, M, P, S, T, V	95.5 / 90.1	<0.01
T4SS	TGT, GAA, GAG, TTC, ATA, AAT, CCG, CCT, CAA, AGG, CGA, TCT, GTA	100 / 100	R, N, M, C, Q, E, I, K, F, T	100 / 98.2	>0.05
T6SS	GCA, GCC, GAC, GAT, CAT, AAA, TTA, TTG, ATG, AAC, CCG, AGG, CGA, CGC, AGC, TCT, ACC, GTA, GTC, TGG, TAC	99.1 / 96.4	N, W	87.4 / 86.5	<0.05

3.8. Efektīvo kodonu un aminoskābju saistība ar olbaltumvielu vispārējām, t.sk. otrējās struktūras īpašībām

Katram efektīvajam kodonam un aminoskābei (7. Att.) novērotā augstā T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu klasifikācijas precizitāte ($Q \leq 3.84$; $p > 0.05$; 5. Att.), pamatoja nepieciešamību salīdzināt efektīvo un „neizvēlēto” kodonu un aminoskābju iespējamās izpausmes sekrēcijas substrātu vispārējās īpašībās. Tā, sekrēcijas substrātiem novērota būtiska (korelācijas koeficients -0.759 , $p < 0.001$, $N=111$) 10 efektīvo aminoskābju frekvenču saistība ar olbaltumvielu GRAVI indeksu (8. Att. A), kamēr 24 efektīvo kodonu gadījumā (8. Att. B) šādas sakarības nav ($p > 0.05$).



8. Att. T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu apvienotās kopas GRAVI indeksa izmaiņa atkarībā no efektīvo un neizvēlēto aminoskābju sinonīmo kodonu (A) un kodonu (B) daudzuma.

Fig. 8 GRAVY index of the combined set of T1SS, T3SS, T4SS and T6SS secreted proteins upon the values of codon content attributed to the selected and rest variables extracted from the sets of summarized frequencies of synonymous codons (A) and individual codon frequencies (B).

Tāpat, atsevišķi summējot efektīvo un neizvēlēto aminoskābju sinonīmo kodonu frekvences pēc to fizikāli-ķīmiskajām īpašībām, vērojamas polāro, lādēto un hidrofobo efektīvo aminoskābju satura būtiskas atšķirības ($p < 0.05$, Vilksona kritērijs) piecos olbaltumvielu grupu pāros (10. Tab.), tādējādi pārsniedzot novēroto pēc diskriminācijas efektivitātes negrupētu aminoskābju fizikāli-ķīmisko grupu gadījumā (trīs būtiskās starpības; 1. Tab. Publikācijā IV).

Sekrēcijas substrātu grupas būtiski neatšķirās pēc to kopējā GC sastāva asimetrijas (1.Tab. Publikācijā IV), tāpat arī neizvēlēto kodonu un aminoskābju sinonīmo kodonu gadījumā ($p>0.05$, Vilksona kritērijs, Kruskala-Valisa tests), tomēr 24 efektīvo kodonu un 10 efektīvo aminoskābju sinonīmo kodonu kopas kopējā GC- asimetrija apstiprinājās kā būtisks faktors sekrēcijas substrātu grupu atšķiršanai (attiecīgi, četras un divas būtiskas starpības, $p<0.05$, Vilksona kritērijs).

Savukārt, atšķirīgie pozicionālās AT- un GC- asimetrijas rādītāji (9. Att.; 1. Att. Publikācijā IV) neizvēlēto un efektīvo kodonu un aminoskābju sinonīmo kodonu (7. Att.) nukleotīdos, kā arī to atšķirīgās ($p<0.05$, $N=111$) lineārās sakarības ar olbaltumvielu sastāva un struktūras raksturlielumiem (2. Att. un 3. Piel. Publikācijā IV) var norādīt uz atšķirīgu evolucionārās attīstības atspoguļojumu neklasiskās sekrēcijas substrātu efektīvo un neizvēlēto kodonu un aminoskābju kopās.

10. Tabula / Table 10

Efektīvo un neizvēlēto polāro, lādēto un hidrofobo aminoskābju sinonīmo kodonu frekvenču summa (vidējais \pm SK) T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupās.

a būtiski atšķirīgo ($p<0.05$) grupu pāru skaits pēc Vilksona kritērija

b sk. 2. tab. apzīmējumu b

c neizvēlētas aminoskābes

Summary frequencies of polar, charged and hydrophobic amino acids selected and non-selected as predictor variables in proteins secreted by T1SS, T3SS, T4SS and T6SS (averages \pm SE).

a the number of significant ($p<0.05$) differences between pairs of individual protein groups obtained by Mann-Whitney test

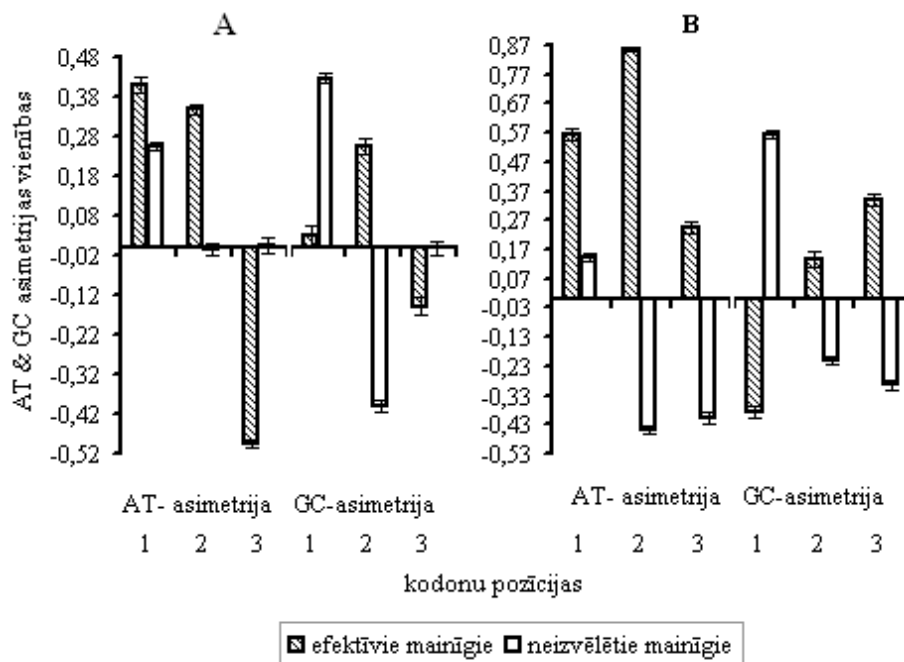
b as b in Table 2

c amino acids non-selected

Aminoskābju grupa <i>Group of amino acids</i>	Aminoskābes <i>Amino acids</i>	Sekrēcijas tips <i>Secretion type</i>				N ^a
		T1SS	T3SS	T4SS	T6SS	
Polāras	S, T, Y, H, C, N, Q	32.44 \pm 0.69	31.48 \pm 0.74	31.02 \pm 0.68	28.52 \pm 0.41	3
	Y, N, Q ^b	14.70 \pm 0.49	13.09 \pm 0.45	14.83 \pm 0.58	11.50 \pm 0.37	5
	S, T, H, C ^c	17.74 \pm 0.53	18.39 \pm 0.55	16.19 \pm 0.56	17.03 \pm 0.35	1
Lādētas	D, E, K, R	48.21 \pm 0.56	48.67 \pm 0.91	41.08 \pm 0.57	50.67 \pm 0.66	4
	E, R, K ^b	10.95 \pm 0.71	14.66 \pm 0.58	22.27 \pm 0.58	14.96 \pm 0.43	5
	D ^c	8.39 \pm 0.36	5.18 \pm 0.27	5.63 \pm 0.39	5.81 \pm 0.93	3
Hidrofobas	G, A, V, L, I, F, M, P, W	19.34 \pm 0.78	19.84 \pm 0.60	27.89 \pm 0.53	20.77 \pm 0.54	3
	G, M, P, W ^b	15.68 \pm 0.41	15.94 \pm 0.86	13.34 \pm 0.74	18.10 \pm 0.44	5
	A, V, L, I, F ^c	32.53 \pm 0.71	32.73 \pm 0.75	27.74 \pm 0.75	32.57 \pm 0.49	3

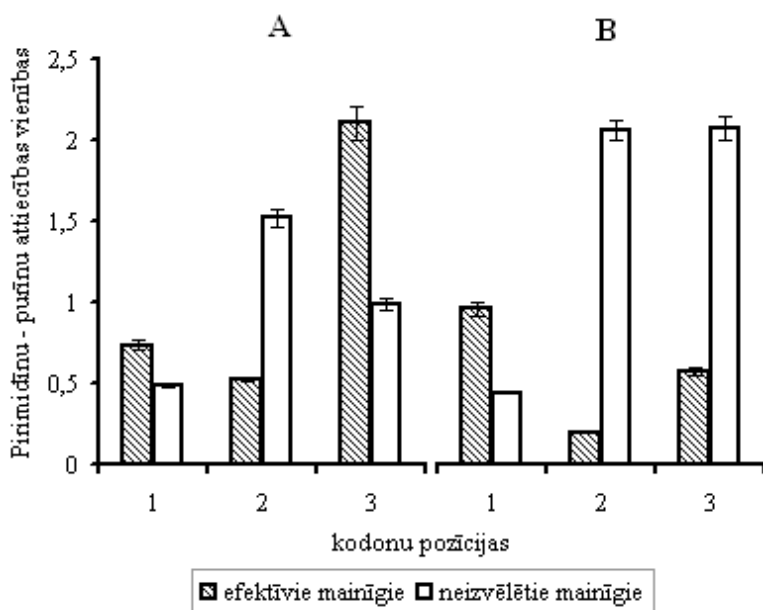
Salīdzinot atsevišķi T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupas pēc efektīvo un neatlasīto kodonu kopu (7. Att.) kopējās pirimidīnu-purīnu attiecības (ņemot vērā visas kodonu pozīcijas), neatlasīto kodonu kopā, gluži kā negrupētu kodonu gadījumā (1. Tab.), novērotas piecu grupu pāru būtiskas atšķirības ($p < 0.05$, Vilksona kritērijs). Tāpat arī sekrēcijas substrātu grupas neatšķirās ($p > 0.05$) efektīvo kodonu gadījumā.

Līdzīgi samērojot efektīvo un neatlasīto aminoskābju sinonīmo kodonu (7. Att.) pirimidīnu-purīnu attiecības, sekrēcijas substrātu grupām novērotas, attiecīgi, četras un divas būtiskas atšķirības ($p < 0.05$). Tāpat sekrēcijas substrātu kopumā (111 sekvenču) būtiski atšķirās ($p < 0.001$) efektīvo un neizvēlēto kodonu un aminoskābju sinonīmo kodonu kopu (7. Att.) pirimidīnu-purīnu pozicionālās attiecības (10. Att.).



9. Att. T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu apvienotās kopas (111 sekvenču) kodējošo sekvenču AT- un GC- asimetrija (vidējais \pm SK) katrai kodonu pozīcijai izvēlēto un neizvēlēto kodonu (A) un aminoskābju sinonīmo kodonu (B) kopās.

Fig. 9 The patterns of AT-skew and GC-skew (average \pm SE) attributed to codon positions of the non-selected and predictor variables (shaded bars) extracted from the sets of individual codon frequencies (A) and summarized frequencies of synonymous codons (B) representing the combined set (111 sequences) of T1SS, T3SS, T4SS and T6SS secreted proteins.

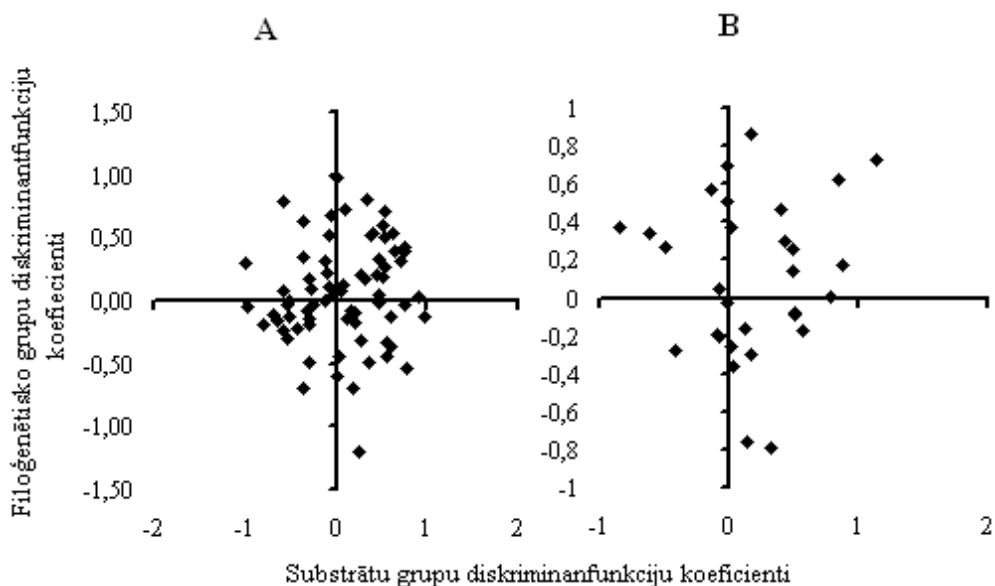


10. Att. T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu apvienotās grupas (111 sekvences) pirimidīnu-purīnu attiecības (vidējais ± SE) atsevišķās kodonu pozīcijās efektīvo un neatlasīto kodonu (A) un aminoskābju sinonīmo kodonu (B) kopās.

Fig. 10 The patterns of pyrimidine versus purine ratio (average ± SE) attributed to codon positions of the non-selected and predictor variables (shaded bars) extracted from the sets of individual codon frequencies (A) and summarized frequencies of synonymous codons (B) representing the combined sets (111 sequences) of T1SS, T3SS, T4SS and T6SS secreted proteins.

3.9. Sekrēcijas substrātu pārstāvība proteobaktēriju subklasēs

T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupu kopā ir atšķirīgi pārstāvētas proteobaktēriju α , β , γ , δ un ξ subklasēm piederīgas baktēriju ģintis (attiecīgi, 11, 15, 80, 1 un 4 sekvences). Novērtējot šādas nevienmērīgās pārstāvības iespējamo ietekmi uz diskriminantfunkciju efektivitāti (2. Tab.), 111 sekvences tika pārgrupētas pēc to atbilstības α , γ (80 sekvences), δ un ξ (apvienotā grupa) vai β proteobaktēriju subklasei (attiecīgi, 1-4 grupa), par diskriminācijas pazīmēm izmantojot jau iegūtās efektīvo mainīgo kopas (7. Att.). Efektīvo mainīgo kopas uzrādīja ievērojamu precizitāti, klasificējot olbaltumvielas pēc to piederības proteobaktēriju subklasēm (LOOCV; 83.8%; $Q=204.57$), tomēr nebija vērojamas ($p>0.05$) sakarības starp šo abu diskriminantanalīžu standartizēto kanonisko funkciju koeficientiem (11. Att.). Tas liecina, ka nevienmērīgais proteobaktēriju subšklašu sadalījums neklasiskās sekrēcijas substrātu grupās būtiski neietekmē efektīvo mainīgo (7. Att.) diskriminantfunkcijas (2. Tab. un 2. Piel. Publikācijā IV).



11. Att. Efektīvo kodonu (A) un efektīvo aminoskābju sinonīmo kodonu (B) standartizēto kanonisko diskriminantfunkciju koeficientu novērtējums sekrēcijas substrātu grupu un proteobaktēriju subklašu grupu multīplo diskrimināciju gadījumā.

Fig. 11 Comparison of the changes of standardized canonical discriminant function coefficients for the effective codons (A) and effective synonyms amino acids (B) in regard to multiple-4-group discrimination of the groups of secretion substrates and groups of representatives of proteobacterial subclasses.

3.10. Svešas izcelsmes gēni

Vairumam (63%) neklasiskās sekrēcijas substrātu, kuri atšķirīgā mērā pārstāvēti T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupās (attiecīgi, 65%, 76%, 81% un 23%), to kodējošo sekvenču un aminoskābju saturs atšķirās no atbilstošam genomam raksturīgā statistiski vidējā kodonu izlietojuma (11. Tab.; 4.Tab. Publikācijā IV) un tādējādi šādas sekvences tika klasificētas kā svešas izcelsmes (*foreign*) gēnus pārstāvošas (Garcia-Vallvé *et al.* 2000).

Atsevišķi salīdzinot laterāli pārnesto olbaltumvielu (41 sekvenču) raksturlielumus četrās olbaltumvielu grupās, T4SS sekrēcijas substrātu kompozicionālais vecums (Sobolevsky un Trifonov 2005) būtiski atšķirās ($p < 0.05$, Vilkoksona kritērijs) no T1SS, T3SS un T6SS sekrēcijas substrātiem (3.Att. Publikācijā IV), kas ir samērojami ar novērotajām T4SS sekrēcijas substrātu atšķirībām pilnajā, 111 sekvenču, olbaltumvielu kopā (1. Tab.; 1.Tab. Publikācijā IV) un liecina par visumā augstāku „jaunāko” aminoskābju īpatsvaru T4SS sekrēcijas substrātu sekvencēs.

11. Tabula./ Table 11

T1SS, T3SS, T4SS un T6SS sekretācijas substrātu sekvences, kas atpazītas kā horizontāli pārnestas un nezināmas izcelsmes sekvences grupās.

Sequences of T1SS, T3SS, T4SS and T6SS secreted proteins recognized as horizontally transferred and putative alien sequences.

Sekretācijas tips <i>Secretion type</i>	Horizontāli pārnestas <i>Horizontally transferred</i>	Nezināmas izcelsmes sekvences <i>Alien sequences</i>
T1SS	hlyac_ecoli, paxa_pasae, hlya_actsu	slap_caucr , nodo_rhylv, frpa_neimb, frpc_neimc, slap_camfe, hasa_serma, o67991_psefl, o85374_promi, prtx_erwch, q6sqm7_pseae, q9xb65_erwam, hlyt_vibho, q59932_serma, o87109_serma, q93m3_yerru, przn_sersp, alge4_azovi, alge7_azovi
T3SS	ipab_shifl, q51451_pseae, q57nf3_salch, hopd2_psesm, q8zq79_salty, q57qv1_salch, ipab_shidy, ipgd_shiso, yoph_yerps, arpt2_psesm, lrv_yerpe, avrp3_psesm, q9kwh9_ecolx	popa_ralso, yope_yeren, hrpn_erwam, q87y16_psesm, o33976_ecoli, o87327_psesx, hrpn_peccc, hrpz_pseap, avrd1_psess, avrd1_psesg, sopb_saldu, sipa_salen, sptp_salpa, ipac_shidy, ipaa_shifl, ipad_shifl, yope_yerpe, sope_saldu
T4SS	vire2_agrt5, vire2_agrtu, a5wy70_9rhiz, a5wy71_9rhiz, q8vt81_9rhiz, q7d2d9_agrt5, a5idc4_legpc, q5wv98_legpl	q6g2a5_barhe, q6g2a4_barhe, q6g2a3_barhe, q5wxw8_legpl, q5x6g3_legpa
T6SS	q2p058_xanom	q4v2q6_burma, q6ee19_edwta, a8yqr4_edwta, q7axz0_salet

Abām efektīvo mainīgo kopām (7. Att.) atsevišķi novērtētie sekvenču kodonu adaptācijas (CAI) indeksi būtiski atšķīrās no neizvēlēto kodonu un aminoskābju sinonīmo kodonu kopu CAI indeksiem ($p < 0.05$, Vilkoksona kritērijs). Tāpat būtiski atšķīrās 70 svešas un 41 laterālas izcelsmes sekvenču kopu CAI indeksi, kā arī nukleotīdu un aminoskābju satura raksturlielumi ($p < 0.05$, Vilkoksona kritērijs; 4. Piel. Publikācijā IV). Papildus salīdzinot laterālās pārneses un svešas izcelsmes sekvences pēc to CAI indeksa abās efektīvo un neizvēlēto mainīgo kopās (7. Att.), sekvenču grupas neatšķīrās ($p > 0.05$) efektīvo mainīgo gadījumā, bet būtiski atšķīrās ($p < 0.01$) pēc neizvēlēto kodonu un aminoskābju sinonīmo kodonu CAI vērtībām. Tas norāda, ka sekretācijas substrātu gēnu atšķirīgā izcelsme neatspoguļojas efektīvo mainīgo kopās.

3.11. Patiesi un kļūdaini klasificēto sekvenču īpašību atšķirības

Salīdzinot abu izmantoto klasifikatoru (2. Tab.) precizitāti un sekvenču īpašības, novērots, ka 101 patiesi (4. Tab.) un desmit kļūdaini klasificētajām sekvencēm (q8vpk4_psefl, ipgd_shiso, arpt2_psesm, ipaa_shifl, lcrv_yerpe, sope_saldu, a8yqr4_edwta, q7axz0_salet, q2p058_xanom) būtiski atšķīrās ($p < 0.05$, Vilkoksona kritērijs) to kompozicionālais vecums, metaboliskās 'dārdzības' un molekulārā tilpuma indeksi, lai gan atšķirības nebija vērojamas ($p > 0.05$, Vilkoksona kritērijs) to vispārējos kodonu izlietojuma raksturlielumos (CAI, Nc, AT- un GC-asimetrija, u.c.). Tomēr, papildus salīdzinot efektīvo kodonu (7. Att.) raksturlielumus, kļūdaini klasificēto sekvenču gadījumā novērotas augstākas pirimidīnu-purīnu attiecības un samazinātas GC-asimetrijas rādītāja vienības efektīvo kodonu 2. pozīcijā ($p < 0.05$, Vilkoksona kritērijs). Šie rezultāti kopumā norāda uz lielāku 'jaunāko', t.i., 'dārgāko' un 'apjomīgāko' aminoskābju īpatsvaru kļūdaini klasificēto olbaltumvielu sekvencēs. Tādēļ ir pamats uzskatīt, ka šīs sekvences ir attīstījušās pēc atšķirīgiem selekcijas nosacījumiem (Akashi un Gojobori 2002, Sobolevsky un Trifonov 2005, Rocha et al. 2006, Soto un Tohac 1983; Knight *et al.* 2001; Chen *et al.* 2004), ko papildus akcentē efektīvo kodonu kopas (2. Tab.; 6. Tab.) raksturlielumu atšķirības.

3.12. Piemērs baktēriju eksoproteoma precizēšanai ar iegūtajām diskriminantfunkcijām

Lai pārbaudītu no aminoskābju un kodonu izlietojuma frekvenču profiliem iegūto klasifikatoru pielietojuma iespējas proteobaktēriju *Pseudomonas syringae* B728a, *Helicobacter pylori* 26695 un *Zymomonas mobilis* ZM4 eksoproteoma (*secretomeP* olbaltumvielu) precizēšanai, sekvenču sadalījums sākotnēji novērtēts, komplementāri izmantojot 24 efektīvo kodonu un 10 efektīvo aminoskābju kopu (7. Att.) veidotās diskriminantfunkcijas, kur, attiecīgi, 53%, 46% un 49% pārbaudīto sekvenču ar abām klasifikācijas metodēm tika saskanīgi prognozēta to piederība kādai no T1SS, T3SS, T4SS vai T6SS substrātu grupām (12. Tab.). Savukārt, tās sekvences, kuru iespējamā piederība, pēc abām klasifikācijas metodēm atšķīrās vienīgi divu sekrēcijas tipu ietvaros (T1SS-T6SS, T4SS-T6SS vai T3SS-T4SS), tika precizētas, komplementāri izmantojot lineārās diskriminantfunkcijas, kuras veido šo grupu pāru efektīvo aminoskābju un kodonu kopas (8. Tab.), tā palielinot saskanīgi prognozēto sekvenču kopējo īpatsvaru minēto organismu eksoproteomā (attiecīgi, 60%, 64% un 58% sekvenču; 12. Tab.).

***P. syringae*, *H. pylori* un *Z. mobilis* eksoproteomu *secretomeP* olbaltumvielu klasificēšanas rezultāts (sekvenču skaits).**

a Prognozējošie mainīgie multiplai-4-grupu diskriminantanalīzei (7. Att.)

b Prognozējošie mainīgie lineārai grupu pāru diskriminantanalīzei (8. Tab)

Prediction result (number of sequences) for the classification of the *secretomeP* proteins from the exoproteomes of *P. syringae*, *H. pylori* and *Z. mobilis*.

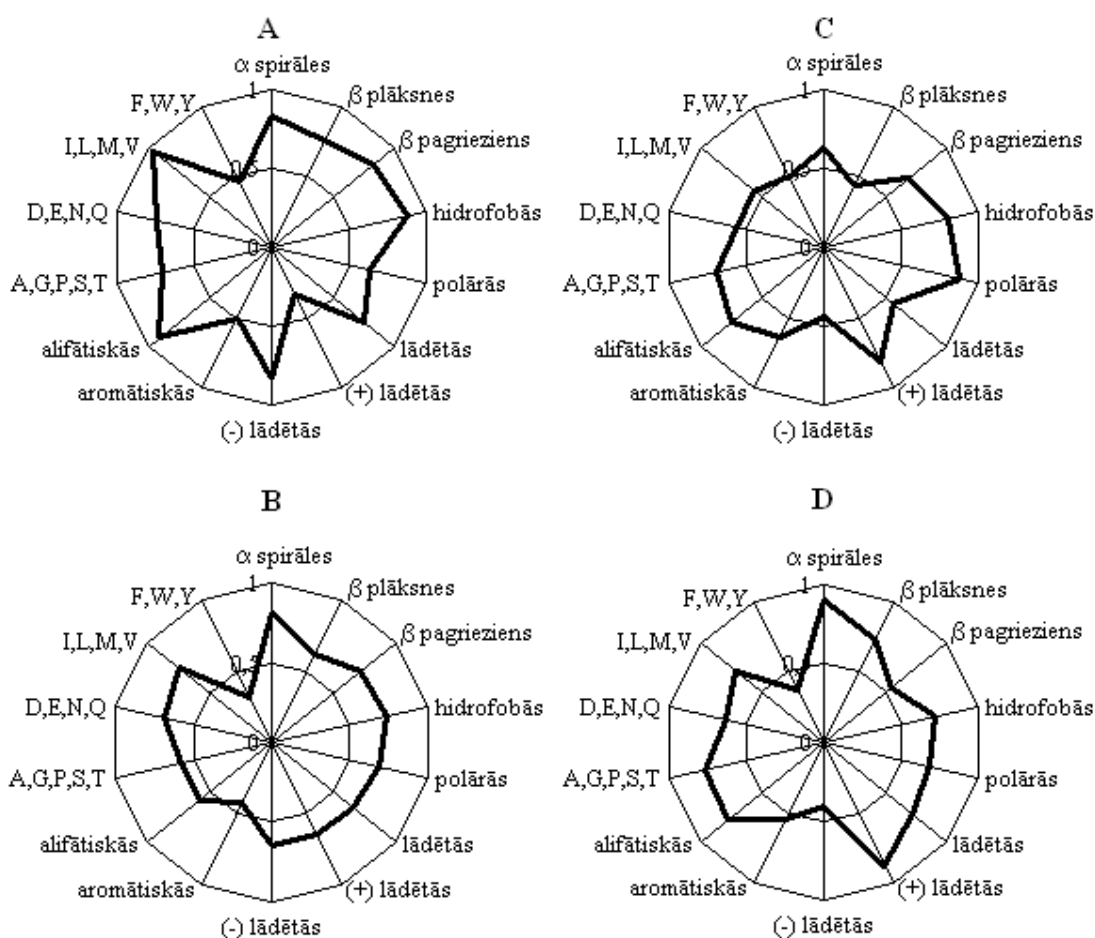
b The predictor variables selected from codon frequencies (Fig 7)

c The predictor variables selected from summarised frequencies of synonymous codons for ten amino acids (Table 8)

Skrēcijas tips <i>Type of secretion</i>	Organisms		
	<i>Organism</i>		
	<i>P. syringae</i>	<i>H. pylori</i>	<i>Z. mobilis</i>
Multiplā-4-grupu diskriminantanalīze ^a			
T1SS	4	0	10
T3SS	9	0	10
T4SS	0	16	0
T6SS	14	2	16
T1SS-T6SS	6	0	4
T3SS-T4SS	4	7	1
T3SS-T6SS	6	0	7
Lineārā grupu pāru diskriminantanalīze ^b			
T1SS	1	0	1
T3SS	2	5	2
T4SS	2	2	0
T6SS	2	0	3
Klasificēšanas rezultāts kopumā			
T1SS	5	0	11
T3SS	11	5	12
T4SS	2	18	0
T6SS	16	2	19

3.13. Sekvenču terminālo fragmentu salīdzināšana

Jau sākotnēji ierobežotajā paraugkopā apskatītajām T1SS un T3SS skrēcijas substrātu grupām (42 sekvences) novērotas sekvenču terminālajiem fragmentiem (50 aminoskābju atlikumi) prognozēto fizikāli-ķīmisko un struktūras īpašību (polaritāte, β -plāksnes) profilu (*pattern*) potenciālās iespējas atšķirt skrēcijas substrātu grupas (2., 3. Tab. Publikācijā I). Līdzīgi, turpmāk izvērtējot T1SS, T3SS, T4SS un T6SS skrēcijas substrātu paplašinātas kopas sastāva un struktūras īpatnības sekvenču C- un N- terminālajos fragmentos (64 aminoskābju atlikumi abos gadījumos) kopīgu īpašību aminoskābju grupām (Pasquier *et al.* 1998), tika novērotas atšķirības to periodiskajā izvietojumā (12. Att.; Publikācija II).



12. Att. Dažādu kopīgu īpašību aminoskābju kopu izvietojuma periodisko īpašību relatīvā pārstāvība T1SS (A), T3SS (B), T4SS (C) un T6SS (D) sekrēcijas substrātu sekvenču C-terminālajos fragmentos, kas aprēķināta, normējot sekvenču, kuru aminoskābju izvietojumam novērots vismaz viens būtisks periods, pret sekvenču skaitu katrā olbaltumvielu grupā.

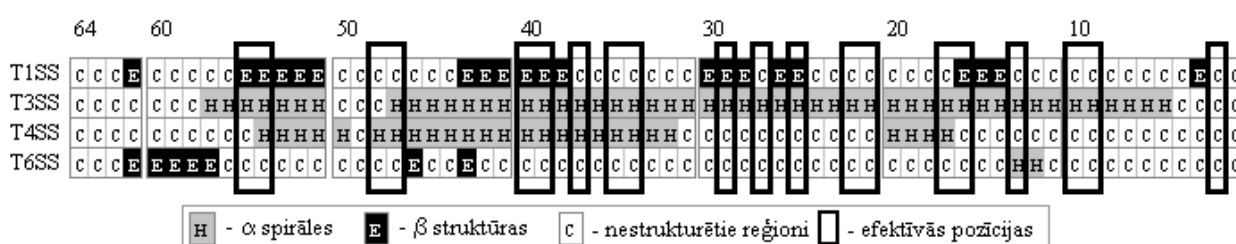
Fig. 12 The relative number of sequences exhibiting at least one significant periodicity (normalized per number of sequences within the respective protein group) for the property-based predefined groups of amino residues in C-termini of T1SS (A), T3SS (B), T4SS (C) and T6SS (D) secreted proteins from proteobacteria.

Izmantojot efektīvo mainīgo kopu, kas izvēlēta no aromātisko (His, Phe, Trp, Tyr) un alifātisko (Val, Lys, Ile, Ala) aminoskābju periodicitātes profiliem (datu vektors 38x111), ir iespējams adekvāti klasificēt 96.4% T1SS un T3SS sekrēcijas substrātu sekvenču (LOOCV; 3. Tab. un 3.Att. Publikācijā II), tāpat arī multiplai-4-grupu diskriminācijai izvēlēto efektīvo mainīgo kopa uzrādīja augstu prognozēšanas precizitāti (LOOCV; 79.6% ; Q=147.87).

T1SS sekrēcijas substrātiem novērotais statistiski lielākais ($p < 0.05$, Frīdmana tests) būtisko periodu relatīvais daudzums β plāksnes un pagriezienus veidojošo, kā arī hidrofobo, lādēto un alifātisko aminoskābju kopu gadījumā (12. Att.), tāpat arī savstarpēji neatšķirama periodu relatīvā

pārstāvība T3SS, T4SS un T6SS sekrēcijas substrātu grupās ($p>0.05$), ir atbilstoša mazākai LOOCV precizitātei multiplās-4-grupu diskriminācijas gadījumā. Šie rezultāti norāda uz T1SS sekrēcijas substrātu ievērojami izteiktākām otrējās struktūras atšķirībām.

Ir likumsakarīgi, ka arī sekrēcijas substrātu sekvenču C- terminālajos fragmentos (64 aminoskābju atlikumi) prognozēto (3D-PSSM metode, Kelley *et al.* 2000) otrējās struktūras elementu izvietojums konsensus (t.i., statistiski vispārinātās) sekvencēs norādīja noteiktus apgabalus kā iespējamās diskriminatīvās pazīmes olbaltumvielu grupu atšķiršanai (13. Att.; Publikācija III).



13. Att. T1SS, T3SS, T4SS un T6SS sekrēcijas substrātiem C- terminālo fragmentu 64 aminoskābju atlikumiem prognozēto otrējās struktūras elementu konsensus sekvences.

Fig. 13 Consensus sequences obtained for the subsets of T1SS, T3SS, T4SS and T6SS secreted protein C- terminal fragments of 64 amino acids after transcription in terms of predicted structural elements.

Prognozēto otrējās struktūras elementu profili sekrēcijas substrātu grupās tika pārkodēti skaitliskos vektoros (attiecīgi: 1, 2 un 3 α spirāļu, β struktūru (*strand*) un nestrukturēto reģionu (*coil*) gadījumā), un izmantoti efektīvo mainīgo atlasei T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu diskriminācijai. Iegūtās diskriminantfunkcijas precizitāte (LOOCV; 74.8%; Q=146.82) izrādījās samērojama ar multiplās-3-grupu (T1SS, T3SS, T4SS) diskriminantanalīzes (Publikācija III) veikspēju (LOOCV; 89.9%, Q=128.10) .

Precizējot otrējās struktūras elementu diskriminatīvo pazīmju kapacitāti, tika izvēlētas efektīvo mainīgo kopas sekrēcijas substrātu grupu pāru lineārai diskriminācijai (4. Tab. Publikācijā III), un augstākā precizitāte (LOOCV; 100%) sasniegta, klasificējot T4SS un T6SS sekrēcijas substrātus, kas varētu norādīt arī uz iespējamo sekrēcijas signālu saturošo apgabalu struktūras būtiskajām atšķirībām minētajās olbaltumvielu grupās.

Multiplās-4-grupu diskriminantanalīzes efektīvo mainīgo kopa patiesi klasificēja 47.6% no neatkarīgās olbaltumvielu sekvenču kopas, proti, 59% no *Salmonella typhimurium* T3SS substrātu un nevienu no proteobaktēriju lipāzēm (T1SS). Rezultātu precizēšanai izmantojot atbilstošo sekrēcijas tipu efektīvo mainīgo kopas pāru lineārā diskriminācijā klasifikācijas rezultāti sasniedza 70-82% un 50-100%, attiecīgi, T3SS un T1SS substrātu kopās.

4. Diskusija

T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu multiplai diskriminantanalīzei tika izvēlētas divas efektīvo mainīgo kopas – kodonu izlietojuma un aminoskābju sinonīmo kodonu frekvences (7. Att.) ar savstarpēji samērojamu LOOCV krosvalidācijas kļūdu intensitāti (4. Tab.; Publikācija IV). Izvēlētie 24 kodoni pārstāv 16 aminoskābes, kurām atbilst 51 kodons, tādējādi 53% no attiecīgo aminoskābju kodoniem netika izvēlēti par efektīvajiem mainīgajiem, bet efektīvo kodonu kopas aizstāšana ar 16 atbilstošo aminoskābju sinonīmo kodonu kopu vai 51 individuāla kodona frekvencēm būtiski samazina diskriminantanalīzes precizitāti ($p < 0.05$, Mak Nemara kritērijs). Sešpadsmit no 24 efektīvajiem kodoniem ir pārstāvēti deviņu no 10 efektīvo aminoskābju sinonīmo kodonu kopā. Šo „kopīgo komponentu” kopas (7. Att.) diskriminantfunkciju precizitāte ir būtiski zemāka nekā efektīvo 24 kodonu kopai (LOOCV; 86.5%; $p < 0.05$, Mak Nemara kritērijs), bet samērojama ar 10 efektīvo aminoskābju kopas veiktspēju ($p > 0.05$, Mak Nemara kritērijs). Tas, savukārt, norāda uz „kopīgo” kodonu izšķirīgo lomu attiecīgo sinonīmo kodonu izlietojumā, priekšnosakot efektīvo aminoskābju, tādējādi arī olbaltumvielu otrējo struktūru nemainību (*preservation*). Sinonīmo kodonu asimetriskā izlietojuma ietekmi uzsver arī samazināta, lai arī samērojama, diskriminantfunkciju precizitāte (LOOCV; 88.3%; $p > 0.05$, Mak Nemara kritērijs), aizstājot 10 efektīvo aminoskābju sinonīmo kodonu kopu ar atbilstošajiem 26 individuālajiem kodoniem.

Abu klasifikatoru (7. Att.) precizitātes samērojamība (4. Tab.) ne vien uzsver to komplementāru izmantošanas iespēju, pat nepieciešamību, bet arī norāda, ka nevienai no metodēm nav tiešu priekšrocību. Gadījumos, kad olbaltumvielai nav zināma tās sekvenca un / vai atbilstošo kodonu secība, pirmējai raksturošanai (*preliminary screening*) var būt piemērota efektīvo mainīgo kopa, kas izvēlēta no eksperimentāli noteiktā atbilstošo aminoskābju satura. Tā, iegūto diskriminantfunkciju veiktspēja tika pārbaudīta šādam iespējamajam praktiskajam pielietojumam, iekļaujot vispārzināmos ierobežojumus aminoskābju sastāva analīzē (aspargīna un asparģīnskābes, kā arī glutamīna un glutamīnskābes saturu pārstāvējot to frekvenču summām). Efektīvo aminoskābju frekvenču kopā aizstājot Asp un Asn, kā arī Glu un Gln individuālās frekvences frekvenču summu, efektīvo aminoskābju kopas (7. Att.) diskriminantfunkciju precizitāte saglabājas.

Tieši pēdējā dekādē, pat piecgadē, strauji audzis zināšanu un kopējās informācijas apjoms par neklasiskās sekrēcijas mehānismiem, kas tieši atspoguļojas arī promocijas darba izstrādē. Proti, kopš sākotnējo rezultātu publicēšanas (Publikācija I), pētījumā izmantoto olbaltumvielu grupu un sekvenču skaits ir vairāk kā dubultojies (Publikācija IV), līdztekus pieaugot eksperimentāli raksturoto ekstracellulāro olbaltumvielu un definēto neklasiskās sekrēcijas tipu skaitam.

Vienu ekoloģisko nišu pārstāvošiem organismiem ir līdzības to olbaltumvielu sekvenču aminoskābju sastāvā, bet kopīgas kodonu izlietojuma tendences pamatā mēdz novērot funkcionāli tuvām olbaltumvielām (Goodarzi *et al.* 2008). Tādēļ, lai pēc iespējas mazinātu kādas olbaltumvielu grupas dominanti, pētījumā izmantotas sekvences, kuras pārstāv dažādas olbaltumvielu funkcionālās klases, t.sk. augiem, dzīvniekiem un cilvēkiem patogēnu proteobaktēriju efektorās olbaltumvielas, kas iedarbojas citosolā vai starpšūnu telpā. Likumsakarīgi, ka sekrēcijas substrātu sekvenču grupām novērots nevienmērīgs atšķirību profils (1. Tab.) un raksturīga augsta kodonu un aminoskābju kompozicionālā variabilitāte (6. Att.). Tomēr kodonu gadījumā novērotā augstākā informatīvā kapacitāte (5. Att.; 6. Att. C) nenodrošina augstāku diskriminantfunkciju precizitāti efektīvo kodonu kopai (7. Att.; 4. Tab.), joprojām uzsverot, ka no kodoniem un aminoskābju sinonīmajiem kodoniem atlasīto efektīvo mainīgo kopas ir komplementāri izmantojamas iespējami precīzākā diskriminācijas rezultāta iegūšanai. Kā izņēmums jānorāda, ka, lai arī iegūtie neviennozīmīgie rezultāti var tikt precizēti ar atbilstošām lineārās diskriminantanalīzes efektīvo mainīgo kopām (8. Tab.), T3SS un T6SS klasifikācijas neviennozīmīgu rezultātu gadījumā varētu tikt dota priekšroka rezultātiem, kuri iegūti izmantojot efektīvo kodonu kopas diskriminantfunkcijas (2. Tab.; 6. Tab.; 9. Tab.). Multiplās diskriminantanalīzes gadījumā var ņemt vērā arī klasifikācijas rezultātos norādīto t.s. otru atbilstošāko grupu (*2nd highest group*), un gadījumos, kad ar vienu efektīvo mainīgo kopu (7. Att.) kā otra atbilstošākā ir norādīta otra klasifikatora prognozētā grupa (un *vice versa*), šāda prognozēšanas rezultātu saskaņa viennozīmīgi izslēdz atlikušās divas neklasiskās sekrēcijas substrātu grupas no sākotnēji potenciāli iespējamā četru grupu klāsta.

Abu klasifikatoru (7. Att.) augsto precizitāti (2. - 5. Tab.) pamato novērotās būtiskās atšķirības aminoskābju sastāvā un kodonu izlietojuma raksturlielumos T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupās (1. Tab.; 1. Tab. Publikācijā IV). Zināma rezultātu nesakritība kodonu adaptācijas indeksa (CAI; Sharp un Li 1987) un efektīvo kodonu skaita (Nc; Wright 1990) gadījumos (1. Tab.), visticamāk, ilustrē atšķirības abu kodonu izlietojuma indeksu

aprēķinā. Proti, Nc raksturo kodonu izlietojumu paša gēna ietvaros, tad CAI ir vairāk attiecināms uz gēna un atsauces genoma (*reference set*) kodonu izlietojuma atšķirībām.

Efektīvo un neizvēlēto kodonu un aminoskābju sinonīmo kodonu sastāva būtiskās atšķirības to saistībā ar olbaltumvielu sekvenču un struktūras raksturlielumiem (8. - 10. Att.; 10. Tab.; 2. Att. un 3. Piel. Publikācijā IV) labi saskaņojas ar vispārpieņemto priekšstatu, ka bāžu saturs un kodonu izlietojums nosaka aminoskābju sastāvu (Foster *et al.* 1997, Paila *et al.* 2008, Knight *et al.* 2001; Sueoka 1961), tādējādi arī vispārējās olbaltumvielu fizikāli-ķīmiskās īpašības (Biro *et al.* 2003; Singer un Hickey 2000).

Efektīvo aminoskābju sinonīmo kodonu kopas CAI ir augstāks nekā pārējo (neatlasīto) aminoskābju kopai (CAI 0.683 ± 0.009 un 0.645 ± 0.009 ; $p < 0.01$, Vilkoksona kritērijs), kas norāda uz evolucionārā spiediena (*evolutionary force*) atspoguļojumu efektīvajās aminoskābēs. Tāpat arī T1SS, T3SS, T4SS un T6SS sekrēcijas substrātos novērotā augstākā ($p < 0.02$, Vilkoksona kritērijs) efektīvo kodonu izlietojuma asimetrija (CAI 0.678 ± 0.008) salīdzinājumā ar neizvēlētajiem kodoniem (CAI 0.651 ± 0.010). Tas uzsvēr noteiktu, katra olbaltumvielu sekrēcijas tipa substrātiem raksturīgu, kodonu izlietojumu (*type specific signature*) un zināmā mērā sasauca ar vispārējā sekrēcijas ceļa (GSP) substrātu signālpeptīdiem raksturīgo asimetrisko kodonu izlietojumu (Zalucki *et al.* 2009). Jāmin, ka GC saturs un dažu aminoskābju (īpaši serīna) izlietojuma atšķirības ir izmantotas kā pazīmes T3SS sekrēcijas substrātu prognozēšanai baktērijās *Salmonella typhimurium* un *Pseudomonas syringae* (Samudrala *et al.* 2009), un ir sastopami samērā pretrunīgi novērojumi par cisteīna izteikti samazināto saturu T1SS sekrēcijas substrātu sekvencēs (Angkawidjaja un Kanaya 2006). Neklasiskās sekrēcijas substrātu C-terminālo fragmentu (T3SS gadījumā, N-terminālo) detalizētāka izpēte tādējādi varētu atvieglot sekrēcijas signāla novietojuma precizēšanu. Promocijas darbā iegūti sākotnējie rezultāti (dati nav pārstāvēti), kur T1SS sekrēcijas substrātu C-terminālā apgabala (13÷47 aminoskābju atlikumi) kompozicionālā vecuma vienības (Sobolevsky un Trifonov 2005) ir būtiski zemākas nekā pilnām sekvencēm, tā norādot uz atšķirīgu aminoskābju sastāvu postulētajā sekrēcijas signālu saturošajā apgabalā.

Saskaņā ar vispārpieņemto nostādni, ka izmaiņas kodonu un aminoskābju sekvencēs kopumā atspoguļo evolucionāro attīstību (Sorimachi un Okayasu 2008), iegūtās efektīvo mainīgo kopas (7. Att.) un sekrēcijas substrātu grupu atšķirības GC satura, kodonu izlietojuma (CAI, Nc), metaboliskās dārdzības un kompozicionālā vecuma raksturlielumus (1. Tab.; Publikācija IV)

varētu norādīt arī uz atšķirīgu T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu attīstību (Sorimachi un Okayasu 2008, Sharp un Li 1987; Wright 1990; Akashi un Gojobori 2002; Sobolevsky un Trifonov 2005; Rocha *et al.* 2006). Ir būtiski, ka maksimālais grupu pāru atšķirību skaits novērots, salīdzinot tās pēc glicīna frekvencēm (1. Tab.; 5. Att.; $p < 0.05$, Vilksona kritērijs, Kruskala-Valisa tests), kuras tiek uzskatītas par pašu vienkāršāko kompozicionālā vecuma mēru (*glycine clock*; Trifonov 1999). Tomēr horizontālās gēnu pārneses (HGT) līdzdalība ir apstiprināta visu četru sekrēcijas sistēmu attīstībā (Kelley *et al.* 2007, Ma *et al.* 2006; Frank *et al.* 2005, Boyer *et al.* 2009), kas, savukārt, var veicināt maldinošus secinājumus par olbaltumvielu sekvenču evolucionāro diverģenci (Koonin *et al.* 2000). Tiešām, ievērojama daļa analizēto olbaltumvielu (11. Tab.; Publikācija IV) ir raksturojamas kā horizontāli pārnestu vai nezināmas (*alien*) izcelsmes gēnu kodētas sekvences (Garcia-Vallvé *et al.* 2000). Jāņem vērā, ka šie rezultāti ir vēl precizējami, jo *Serratia* sp., *Proteus* sp., *Azotobacter* sp. un *Edwardsiella* sp. genomu nepieciešamie vispārīnātie dati nav pārstāvēti HGT-Datubāzē (Garcia-Vallvé *et al.* 2003), tādēļ svešās izcelsmes gēnu tuvinātai identificēšanai promocijas darbā izmantota filoģenētiski iespējami tuvu baktēriju genomu informācija (Publikācija IV).

Kopumā svešās izcelsmes sekvences raksturojamas ar zemāku GC- saturu un mazāk izteiktu kodonu izlietojuma asimetriju, tātad arī mazāku konservativitāti, salīdzinājumā ar laterālās izcelsmes sekvenčām (4. Piel. Publikācijā IV). Svešās izcelsmes gēnu izslēgšana no analizējamo olbaltumvielu kopas ievērojami samazina sekvenču raksturlielumu izkliedi grupu ietvaros (4. Piel. Publikācijā IV). Tomēr augsta to variabilitāte saglabājās laterālās izcelsmes T3SS sekrēcijas substrātu grupā (3. Tab. Publikācijā IV), kas atbilst vispārzināmai sekvenču daudzveidībai T3SS sekrēcijas substrātos (McCann un Guttman 2008). Tajā pat laikā nav vērojama atšķirīgas izcelsmes gēnu nevienmērīgā sadalījuma (augstākais un zemākais īpatsvars T4SS un T6SS substrātu grupās) ietekme uz iegūto efektīvo mainīgo kopu (7. Att.) diskriminantfunkcijām. Līdzīgi, diskriminantfunkcijās neatspoguļojas arī atšķirīgais α , β , γ un δ/ξ proteobaktēriju pārstāvju atšķirīgais sadalījums sekrēcijas substrātu grupās, kas saskan ar viedokli, ka GC- saturs (16S rDNS) neatspoguļo proteobaktēriju subklašu (α , β , γ , δ/ξ) diverģences atšķirības (Haywood-Farmer un Otto 2003), lai gan ir ziņots par proteobaktēriju laterālu attīstību, γ proteobaktērijiem kā pēdējām diverģējot no kopīgā senča (Gupta 2000). Savukārt, pēc abu efektīvo mainīgo kopu (7. Att.) diskriminantfunkcijām 10 kļūdaini klasificēto sekvenču efektīvo kodonu 2. pozīcijā, atšķirībā no patiesi klasificēto olbaltumvielu kopai novērotā, ir samazināts guanīna saturs, tā norādot, ka diskriminantfunkciju ierobežojumi ir

saistāmi ar atšķirīgas evolucionārās ģenēzes atspoguļošanas kļūdaini klasificēto olbaltumvielu sekvenču efektīvajos kodonos un aminoskābēs (Knight *et al.* 2001).

Atbilstoši vispārējam priekšstatam, sekrēcijas substrāti un sekrēcijas kompleksi ir koevolucionējuši to savstarpējas saderības un sekrēcijas signālu atpazīšanas optimizēšanai (Economou 2002, Bendtsen *et al.* 2005). Tas, savukārt, nozīmē, ka vienas sekrēcijas sistēmu komponentiem un substrātiem ir jāsaturs kādas kopīgas iezīmes. Būtiski samazinātās T4SS sekrēcijas substrātu kompozicionālā vecuma vienības (1. Tab.; 1. Tab. Publikācijā IV) un līdzīgas sakarības metaboliskās dārdzības un kodonu izlietojuma raksturlielumos (Sobolevsky un Trifonov 2005; Trifonov *et al.* 2006; Akashi un Gojobori 2002; Sharp un Li 1987; Wright 1990) ir saskanīgas ar postulēto evolucionāri vēlāku T4SS izveidi kā mozaīkveida sekrēcijas sistēmu, tai adaptējot komponentus no citām olbaltumvielu sekrēcijas sistēmām (Cao un Saier 2001). No otras puses, T4SS un T6SS sekrēcijas substrātiem novēroti būtiski atšķirīgi, pat pretēji kodonu izlietojuma un aminoskābju satura raksturlielumi (1. Tab.; 1. Tab. Publikācijā IV), kas zināmā mērā, vismaz substrātu līmenī (Economou 2002, Bendtsen *et al.* 2005), ir pretrunā ar ziņojumu par T6SS izveidošanos, adaptējot T1SS, T3SS un T4SS komponentus (Shrivastava un Mande 2008). Savukārt, T1SS sekrēcijas substrātu sekvencēm novērotais vienkāršāku aminoskābju augstākais īpatsvars, kā arī augstās CAI un zemās Nc vērtības, ir saskaņā ar postulēto ABC transportsistēmu operonu agrīno diverģenci (Tomii un Kanehisa 1998).

Promocijas darba izstrādē iegūti arī sākotnējie rezultāti (dati nav pārstāvēti), salīdzinot T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu un transportsistēmu aminoskābju satura raksturlielumus. Baktērijām, kuru sekrēcijas substrātu sekvences iekļautas diskriminantfunkciju izstrādes kopā, T1SS sekrēcijas komponentus pārstāvēt TolC un tai paralogām olbaltumvielām, T3SS – HrpA, T4SS – VirB, un T6SS gadījumā – Vgr grupas olbaltumvielām. Ņemot vērā, ka tikai daļai no pētījumā iekļautajiem sekrēcijas substrātiem bija iespējams piemeklēt attiecīgā organismā pārstāvētas atbilstošās sekrēcijas sistēmas komponentu sekvences, šie rezultāti šobrīd nav vispārināmi, bet tomēr pietiekami skaidri apstiprina vispārējo tendenci, ka sekrēcijas sistēmu komponentu sekvences ir visumā konservatīvākas, salīdzinot ar sekrēcijas substrātiem (McCann un Guttman 2008). Proti, sekrēcijas sistēmu komponenti kopumā izrādījās raksturojami ar zemākām kompozicionālā vecuma vienībām (Trifonov *et al.* 2006). Jāuzsver, ka T1SS, T3SS un T4SS sekrēcijas sistēmu substrātu un komponentu olbaltumvielu vidējais kompozicionālais vecums veido tieši proporcionālu sakarību ($r=0.9998$, $N=3$), ko izjauc T6SS sekrēcijas tipa pārstāvju pievienošana ($r=0.960$, $N=4$, $p>0.05$). Pamatojoties uz nostādni, ka komponentiem un

substrātiem ir jāsaturs noteiktas kopīgas iezīmes, pārbaudītas sekrēcijas komponentu grupu atšķiršanas iespējas ar sekrēcijas substrātu klasificēšanai izvēlēto efektīvo aminoskābju sinonīmo kodonu kopas (7. Att.) diskriminantfunkcijām. Tās izrādījās piemērotas T6SS komponentu sekvenču atpazīšanai, kļūdaini klasificējot lielāko vairumu no pārējo sekrēcijas sistēmu komponentiem. Šāda T6SS komponentu precīza klasifikācija, iespējams, parāda, ka izvēlētajās transportsistēmu komponentu sekvenču pārstāv tos T6SS komponentus, kas vienlaikus tiek uzskatīti arī par T6SS sekretētajām efektorajām olbaltumvielām (Tseng *et al.* 2009). Tanī pat laikā T1SS, T3SS un T4SS komponentu sekmīga diskriminācija (LOOCV; 98.3%) ir iespējama, no aminoskābju sinonīmo kodonu frekvencēm izvēloties citu efektīvo mainīgo kopu (dati nav pārstāvēti).

Pamatā, neklasiskās sekrēcijas substrātu (t.sk. efektoro olbaltumvielu) atpazīšanas metožu izveidē izmanto šauru baktēriju spektru (Samudrala *et al.* 2009; Löwer un Schneider 2009, Vinatzer *et al.* 2005) ar izņēmumu metodes *EffectiveT3* gadījumā, kur izmantotas pat četru proteobaktēriju ģinšu (*Chlamydia*, *Salmonella*, *Yersinia*, *Escherichia*) T3SS efektorās olbaltumvielas (Arnold *et al.* 2009). Šāda pieeja ir piemērota, lai klasificētu attiecīgo vai tiem filoģenētiski tuvu organismu gēnu produktus, taču liedz pietiekami vispārināt izmantotās metodes. Promocijas darbā iegūto diskriminantfunkciju izstrādei izvēlētajā 111 sekvenču paraugkopā pēc iespējas tika samazināta kāda organisma vai olbaltumvielu funkcionālās grupas dominānce, tomēr iegūtās diskriminantfunkcijas izrādījās piemērotas, lai klasificētu arī atsevišķa organisma neklasiskās sekrēcijas substrātu grupu. Tā, ar augstu precizitāti iespējams prognozēt baktērijas *Salmonella typhimurium* T3SS sekrēcijas substrātus (7. Tab.), lai gan tikai trīs šīs baktēriju sugas sekvenču kopas (Samudrala *et al.* 2009) ir pārstāvētas diskriminantfunkciju izstrādes sekvenču kopā (*training set*).

Izvēloties izstrādes sekvenču kopas olbaltumvielu datubāzēs, arī izdarot jebkādas slēdzienus par olbaltumvielu pazīmēm, ir jāņem vērā to kļūdainu anotāciju iespējamība. Tā, dažu funkcionālo klašu fermentu kopai neprecīzas anotācijas ir konstatētas pat 85% sekvenču (Schnoes *et al.* 2009). Salīdzinoši mazā daļa eksperimentāli anotētu olbaltumvielu kopējā neklasiski sekretējamā eksoproteomā likumsakarīgi norāda, ka anotēto sekvenču klāstā un tādēļ arī izstrādes kopā (1. Piel. Publikācijā IV) nav pilnā mērā pārstāvēta atbilstoši sekretējamu olbaltumvielu funkcionālo klašu daudzveidība, kas liedz pārvērtēt jebkuru prognozēšanas metožu, t.sk. promocijas darba izstrādē iegūto efektīvo mainīgo kopu (7. Att.; 8. Tab.) diskriminantfunkciju, iespējas klasificēt šādas olbaltumvielas. Tomēr ir pierādīts, ka savstarpēji homologo sekvenču

izslēgšana no izstrādes kopas būtiski ceļ attiecīgo klasifikatoru precizitāti (Hung *et al.* 2010.). Līdzīgi, arī promocijas darba rezultātu gadījumā apstiprinājās, ka prognozējošie mainīgie, kas sākotnēji izdalīti T1SS un T3SS sekrēcijas substrātu kopu (42 sekvenses) atšķiršanai, saglabā savas diskriminatīvās īpašības un augsto precizitāti arī pilnajā (73 sekvenses) T1SS un T3SS sekvenču kopā (8. Tab.; 3. Tab. Publikācijā I).

Abu klasifikatoru samērojamā kļūdu intensitāte, atpazīstot izstrādes kopas T1SS, T3SS, T4SS un T6SS sekrēcijas substrātus, arī neatkarīgas datu kopas olbaltumvielas (4. Tab.; 7. Tab.), pamato iegūto diskriminantfunkciju izmantošanas iespējas *secretomeP* olbaltumvielu (Bendtsen *et al.* 2005) precīzākai raksturošanai. Tās novērtētas baktēriju *Pseudomonas syringae* B728a, *Helicobacter pylori* 26695 un *Zymomonas mobilis* ZM4 eksoproteomos, klasificējamo sekvenču kopā iekļaujot olbaltumvielas, kuru garums ir vismaz 161 aminoskābes atlikumi (īsākā sekvence izstrādes kopā). Ar efektīvo mainīgo kopām, kas iegūtas no aminoskābju un kodonu izlietojuma frekvencēm (7. Att., 8. Tab.), ir saskanīgi klasificēta, attiecīgi, 60%, 64% un 58% sekvenču piederība kādai noteiktai sekrēcijas substrātu grupai (12. Tab.). Jāuzsver, ka šīs baktērijas pārstāv dažādas proteobaktēriju subklases, un savstarpēji samērojama klasifikācijas rezultāts netieši apstiprina, ka nevienmērīgais proteobaktēriju subklašu sadalījums izstrādes sekvenču kopā (*training set*) būtiski neietekmē efektīvo mainīgo kopu diskriminantfunkcijas (11. Att.). Tomēr kopumā 9% pārbaudīto eksoproteomu sekvenču nav atšķiramas trīs grupu pāru ietvaros (T1SS:T6SS; T3SS:T4SS; T3SS:T6SS) (12. Tab.). Tas netieši atspoguļo izstrādes kopā novēroto, ka visām T1SS substrātu sekvencēm diskriminantanalīzē ar abām efektīvo mainīgo kopām (7. Att.) kā otra atbilstošākā grupa tiek norādīta T6SS, kamēr T3SS grupas sekvencēm – T4SS vai T6SS. Turklāt līdzīgas sekrēcijas substrātu grupu atšķirību tendences kopumā novērotas, salīdzinot vispārējos aminoskābju un kodonu izlietojuma frekvenču parametrus (1. Tab.; 1. Tab. Publikācijā IV).

Baktēriju *Z. mobilis* proteomā netika identificēts neviens T4SS komponents, tāpat baktērijām *H. pylori* ir prognozēta tikai divu T1SS komponentu pārstāvība (*Genome Atlas* datubāze; <http://www.cbs.dtu.dk/>), kas ir saskanīgi ar to, ka klasifikatori (7. Att.) neprognozē T4SS un T1SS substrātu klātbūtni attiecīgo baktēriju sekvenču kopās (12. Tab.). Neviennozīmīgi pronozēto *P. syringae*, *H. pylori* un *Z. mobilis* olbaltumvielu vērā ņemamais apjoms (12. Tab.) var tikt skaidrots ar to, ka daļa šo sekvenču var būt atbilstošas olbaltumvielu klasei, kas ir nepietiekami (nemaz) pārstāvēta izstrādes sekvenču kopā. Piemēram, analogi pārbaudītas klasifikatoru (7. Att.) iespējas klasificēt vienas fermentu klases olbaltumvielas, kurām nav

novērojams atšķelamais N- terminālais signālpeptīds. Klasifikācijai izmantotas 12 levānsaharāžu sekvences, (sacb_psesh, sacb_psesg, sacb_rahaq, sacb_erwam, q93fu9_9psed, q9lhx1_acexy, q2g754_novad, q4ttt0_9sphn, q2nbk2_erylh, q883p5_psesm, q4zum6_pseu2, sacb_zymmo) un tām konstatētas T6SS sekrēcijas substrātu grupu īpašības. Šo rezultātu būs iespējams izvērtēt, līdzko kādai no šīm sekvencēm eksperimentāli tiks noteikts atbilstošais sekrēcijas tips. Tomēr jau tagad jāņem vērā, ka diskriminantfunkciju izstrādes sekvenču kopā nav pārstāvēta neviena hidrolāze. Atkārtoti jāuzsver, ka klasifikatoru (7. Att.; 8. Tab.) iespējas nav vispārināmas bez papildus novērtējuma ar citām *in silico* metodēm. Piemēram, novērtējot atbilstošās kodējošās sekvences lokalizāciju genomā, lai arī, *vice versa*, olbaltumvielas raksturošanai sekvences lokalizācijas genomā noteikšana vien ir nepietiekama. Vienkāršotā gadījumā, būtu maz ticams vienas sekrēcijas sistēmas substrāta gēna izvietojums t.s. patogenitātes salīnā vienkopus ar citas sistēmas komponentiem. Tomēr, lai arī baktēriju *Burkholderia mallei* virulences faktora TssM gēna transkripcijas regulācija norit vienlaicīgi ar turpat blakus lokalizēto T6SS komponentu gēniem, olbaltumvielas sekrēcija nav atkarīga no attiecīgās T6SS (Shanks *et al.* 2009).

Visas izstrādes kopas olbaltumvielu sekvences ir sākotnēji pārbaudītas ar vispārējām prognozēšanas metodēm (Psort, SignalP, SecretomeP). *Psort* metode prognozē olbaltumvielu lokalizāciju (Gardy *et al.* 2005), savukārt *SignalP* (Bendtsen *et al.* 2004b) ir novērtēta kā precīzākā no 13 metodēm Sec- atkarīgās sekrēcijas signālpeptīdu un to atšķelšanas vietu (*cleavage site*) atpazīšanai (Choo *et al.* 2009). Apstiprinājās, ka izvēlētajās sekvencēs nav N-terminālo signālpeptīdu, ar izņēmumu *Yersinia enterocolitica* olbaltumvielas YopR (q7brz4_yeren) gadījumā, kas eksperimentāli anotēta kā T3SS sekrēcijas substrāts (Lee un Schneewind 1999; Blaylock *et al.* 2008). Dažas citas T3SS efektorās olbaltumvielas ar *SecretomeP* metodi (Bendtsen *et al.* 2005) tiek klasificētas kā atkarīgas no vispārējā sekrēcijas ceļa, lai gan *SignalP* metode to sekvencēs nenorāda signālpeptīdu klātbūtni, un *vice versa*. Šie pretrunīgie prognozēšanas metožu rezultāti var būt saistāmi ar T3SS sekrēcijas signālus saturošo sekvenču N- terminālo fragmentu individuālajām īpatnībām (Ghosh 2004; Arnold *et al.* 2009). Turklāt jāņem vērā, ka *SecretomeP* metodes izstrādes sekvenču kopa ir Sec- atkarīgas olbaltumvielas, jo tolaik baktērijām eksperimentāli anotētu neklasiskās sekrēcijas substrātu skaits bija vairāk kā nepietiekams (Bendtsen *et al.* 2005). Tādējādi var būt iespējams, ka pārbaudīto eksoproteomu *secretomeP* olbaltumvielu kopa (12. Tab.) nepārstāv visas neklasiski sekretējamās šo organismu olbaltumvielas, vai arī to vidū varbūtēji ietilpst netipisku T2SS un / vai T5SS substrātu sekvences. Ne velti *LocateP* metode (Zhou *et al.* 2008) ir postulēta kā piemērotāka

(Desvaux *et al.* 2009), lai arī *SecretomeP* metode ir paredzēta arī Gram-pozitīvo baktēriju neklasiski sekretējamo olbaltumvielu atlasei. Savukārt, izmantojot eikariotiskas izcelsmes *secretomeP* olbaltumvielas kā izstrādes sekvenču kopu, ir iegūta diskriminatīvu pazīmju kombinācija neklasiskās sekrēcijas substrātu atlasei, sasniedzot augstu (85%), bet ne maksimālo precizitāti (Hung *et al.* 2009). Tas zināmā mēra ilustrē vispārējo olbaltumvielu pazīmju prognozēšanas metožu individuālo nepietiekamību, tātad nepieciešamību savstarpēji salīdzināt ar atšķirīgām *in silico* pieejām iegūtos rezultātus.

Ierobežotā eksperimentāli anotēto neklasiskās sekrēcijas substrātu sekvenču kopa liedz pārbaudīt pieņēmumu, vai šobrīd raksturoto sekrēcijas tipu paveidu substrāti var būt savstarpēji atšķirami arī attiecīgo tipu ietvaros. Turklāt tādā pat mērā ir iespējama vēl citu, šobrīd nezināmu, sekrēcijas sistēmu klātbūtne proteobaktērijās. Šādu hipotētiski iespējamu sekrēcijas sistēmu ietekme nav izslēgta arī pārbaudītajos *P. syringae*, *H. pylori* un *Z. mobilis* eksproteomos, tādējādi papildus palielinot arī neviennozīmīgi klasificēto sekvenču apjomu (29%; 12. Tab.).

Analogu sekrēcijas sistēmu pārstāvība arī citās baktēriju grupās (Tseng *et al.* 2009) padara iespējamu krietni plašāku šo klasifikatoru pielietojumu, t.sk. arī ārpus proteobaktēriju robežām. Piemēram, esam ziņojuši par T1SS, T3SS un T4SS diskriminācijai izvēlēto prognozēto otrējās struktūras elementu izkārtējuma pozīciju kopas (Publikācija III; Publikācija II) piemērotību klasificēt kā atbilstošas T3SS substrātu grupai trīs *Bacillus subtilis* olbaltumvielas (fla_bacsu, flgk_bacsu, flid_bacsu), kurām postulēta sekrēcijas iespēja pēc T3SS līdzīga mehānisma (Tjalsma *et al.* 2004). Tāds pats prognozēšanas rezultāts, ar izņēmumu flid_bacsu gadījumā, sasniegts, izmantojot efektīvo kodonu un aminoskābju sinonīmo kodonu kopu (7. Att.) diskriminantfunkcijas, tādējādi apstiprinot iegūto klasifikatoru iespējas pietiekami precīzi klasificēt T1SS, T3SS, T4SS vai T6SS sekrēcijas sistēmām homologu vai līdzīgu sekrēcijas sistēmu substrātus arī no Gram-pozitīvajām baktērijām. Tas varētu veidot pamatu atsevišķam pētījumam, izmantojot Gram-pozitīvu baktēriju olbaltumvielu sekvences ar zināmu un atbilstoši raksturotu sekrēcijas mehānismu. Jo sevišķi tādēļ, ka šai baktēriju grupai, salīdzinājumā ar proteobaktērijām, ir prognozēts mazāks Sec- atkarīgā sekretoma apjoms (Song *et al.* 2009) un *Genome Atlas* datubāzē 33% no 49 Firmicutes sugu (82 celmi) kopējā eksproteoma ir prognozēti kā *secretomeP* olbaltumvielas (Bendtsen *et al.* 2005), kas ir lielāks potenciāli neklasiski sekretējamo olbaltumvielu apjoms, salīdzinot ar proteobaktērijām (2. Att.; Bendtsen *et al.* 2005). Līdzīgā veidā ar iegūtajiem klasifikatoriem (7. Att.; Publikācija III; Publikācija II) ir prognozēta baktērijās *B. subtilis* ekspresēta rekombinanta interleikīna-1 (o43645_human)

sekvences (Bellini *et al.* 1995) iespējamā sekrēcija pa T3SS, t.i., šīs olbaltumvielas īpašības ir atbilstošas proteobaktēriju T3SS sekrēcijas substrātu īpašībām. Analogi prognozēta arī vairāku citu rekombinantu olbaltumvielu iespējamā neklasiskā sekrēcija, no kurām γ interferonam (ifng_human) prognozēta atbilstība T4SS, adiponektīnam (gfp_aeqvi) – T1SS, bet prolaktīnam (prl_human) – T3SS sekrēcijas substrātu īpašībām (dati nav pārstāvēti).

T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu C- terminālo fragmentu strukturālās īpatnības atspoguļojas to fizikāli ķīmisko īpašību un otrējās struktūras elementu pozicionālā izvietojuma, kā arī aromātisko un alifātisko aminoskābju periodisko īpašību profilu būtiskajās atšķirībās (13. Att.; 12. Att.; Publikācija III; Publikācija II Publikācija I). Olbaltumvielu otrējās struktūras elementu prognozēšanai izmantotas dažādas metodes, t.sk. “Phyre” (Kelley un Sternberg 2009), kur iegūti samērojami pāru diskriminantanalīzes rezultāti, bet zemāka (LOOCV; 67.6%) multiplās diskriminantfunkcijas precizitāte nekā 3D-PSSM gadījumā (LOOCV; 74.8%), kas, domājams, atspoguļo atšķirīgās pieejas sekvenču otrējās struktūras elementu prognozēšanas metodikas izstrādē.

Randomizētiem C- terminālajiem fragmentiem prognozētie otrējās struktūras elementu profili uzrādīja būtiski zemāku T1SS, T3SS un T4SS sekrēcijas substrātu (89 sekvences) grupu atšķiršanas precizitāti (Publikācija III), tādējādi papildus uzsverot aminoskābju pozicionālā izkārtojuma nozīmi. Kopumā šie rezultāti sasaucas ar hipotēzi par vienas klases olbaltumvielu struktūru tendenci saglabāt hidropātiskās īpašības ģeometriski ekvivalentās sekvenču pozīcijās, mūsu gadījumā par „olbaltumvielu klasi” uzskatot atšķirīgu sekrēcijas sistēmu substrātu „sekrēcijas signāla klasi” (Krissinel 2007). Tomēr būtiski augstāku T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu klasifikācijas precizitāti iespējams sasniegt ar efektīvo mainīgo kopām, kas izvēlētas no kodonu izlietojuma un aminoskābju sinonīmo kodonu frekvencēm (7. Att.). To uzsver arī neatkarīgas sekvenču kopas klasifikācijas rezultāti (7. Tab.). Līdzīgi kā sākotnēji izvēlētās T1SS un T3SS sekrēcijas substrātu sekvenču kopas gadījumā (42 sekvences), kad novērtētas grupu diskriminācijas iespējas pēc aminoskābju izlietojuma sekvenču C- un N-terminālajos fragmentos (50 aminoskābju atlikumi), iegūstot diskriminantfunkcijas ar zemāku LOOCV krosvalidācijas precizitāti nekā salīdzinot aminoskābju izlietojumu pilnās olbaltumvielu sekvencēs (4. Tab. Publikācijā I).

Promocijas darba rezultāti papildina pētījumus par sekvenču īpašību izmantošanu dažādu olbaltumvielu, t.sk. baktēriju eksoproteomu raksturošanai, savstarpējai atšķiršanai un

klasifikācijai (Emmanuelson *et al.* 2007, Gromiha un Suwa 2004, Bendtsen *et al.* 2005). Pirmo reizi apstiprināts, ka olbaltumvielu aminoskābju sastāvs, kodonu izlietojums, un olbaltumvielu C-terminālo fragmentu prognozētās otrējās struktūras elementu izvietojums var veidot efektīvas neatkarīgo mainīgo kombinācijas sekrēcijas substrātu grupēšanai, lai prognozētu to piederību kādam no šobrīd definētajiem olbaltumvielu neklasiskās sekrēcijas tipiem proteobaktērijās. Promocijas darba izstrādē apstiprinātās olbaltumvielu klasifikācijas / prognozēšanas iespējas ir uzskatāmas par līdzvērtīgām un papildinošām metodēm ar noteiktām priekšrocībām neklasiskās sekrēcijas substrātu identifikācijai, jo šim nolūkam izmantotās funkcionāli daudzveidīgās olbaltumvielu sekvences pārstāv 30 proteobaktēriju ģintis. Gadījumos, kad attiecīgā organisma genoms nav sekvencēts vai pietiekami raksturots, iegūtie klasifikatori var ievērojami atvieglot sekretējamas olbaltumvielas īpašību eksperimentālas verifikācijas plānošanu. Gluži tāpat kā citas metodes (Emmanuelson *et al.* 2007, Samudrala *et al.* 2009, Vinatzer *et al.* 2005), arī neklasiskās sekrēcijas substrātu klasifikatorus jāturpina pārbaudīt un precizēt kopsolī ar eksperimentāli anotētu sekrēcijas substrātu skaita pieaugumu un joprojām augošo izpratni par olbaltumvielu sekrēcijas mehānismiem.

Secinājumi

1. Proteobaktēriju neklasiskās sekrēcijas T1SS, T3SS, T4SS un T6SS substrātu grupas var atšķirt pēc olbaltumvielu aminoskābju un kodējošo sekvenču sastāva, īpaši pēc to α spirāles veidojošo aminoskābju īpatsvara un nukleotīdu satura kodonu 3. pozīcijā.
2. T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupām ir atšķirīgs aminoskābju un prognozēto otrējās struktūras elementu pozicionālais izkārtojums sekvenču C-terminālajos fragmentos.
3. T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupu olbaltumvielu aminoskābju un kodējošo sekvenču sastāvs atspoguļo sekvenču ģenēzes atšķirības, ar evolucionāri jaunāku un enerģētiski “dārgāku” aminoskābju lielāku īpatsvaru T4SS substrātos.
4. T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu kodējošajās sekvencēs ir atšķirīgi pārstāvēti svešas izcelsmes gēni.
5. Multiplā diskriminantanalīzē T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupu klasifikācijai izdalītas 24 kodonu frekvenču un 10 aminoskābju sinonīmo kodonu frekvenču kopas, kas nodrošina augstu un savstarpēji samērojamu izšķiršanas spēju, sasniedzot, attiecīgi, 97.3% un 92.8% precizitāti.
6. Abi klasifikatori uzrāda samērojamu izšķiršanas spēju un precizitāti, patiesi prognozējot 81.0% sekvenču piederību neatkarīgā T1SS un T3SS sekrēcijas substrātu kopā.
7. Asimetriskā α , β , γ , δ un ξ proteobaktēriju subklašu pārstāvība un svešas izcelsmes sekvenču nevienmērīgais sadalījums T1SS, T3SS, T4SS un T6SS sekrēcijas substrātu grupās būtiski neietekmē diskriminantfunkciju klasifikācijas/prognozēšanas iespējas.
8. Abi klasifikatori ar augstu precizitāti diskriminē T1SS un T4SS sekrēcijas substrātus, un to salīdzinoši ierobežotākās iespējas klasificēt T3SS un T6SS sekrēcijas substrātus ir skaidrojamas ar atšķirībām kļūdaini klasificēto sekvenču ģenēzē un, attiecīgi, šo olbaltumvielu īpašībās.

Pateicības

Vislielākā pateicība promocijas darba vadītājam Dr.habil.biol. Pēterim Zikmanim par nenogurstošu atbalstu, palīdzību un konsultācijām, izstrādājot promocijas darbu.

Tāpat, paldies kolēģēm LU MBI Dr.biol. Lilijai Auziņai, Dr.biol. Dagnijai Upītei, Mag. biol. Ļaisanai Šakirovai un Dr.biol. Olgai Muterei par morālo atbalstu promocijas darba izstrādes laikā.

Paldies korp! Gundega filistrēm Dr.chem. Rudītei Kalniņai un Sibillai Korulei par palīdzību celt manuskripta kvalitāti.

Promocijas darba izstrādāšanai **saņemts Eiropas Sociālā Fonda projekta** Nr. 2009/0138/1DP/1.1.2.1.2/09IPI/VIAA/004 "Atbalsts doktora studijām Latvijas Universitātē" atbalsts (līgums Nr. 09/142/38).

Literatūras saraksts

1. Agarraberes F.A., Dice J.F. 2001. Protein translocation across membranes. – *Biochimica et Biophysica Acta*, 1513:1-24.
2. Ahn J.H., Pan J.G., Rhee J.S. 1999. Identification of the *tliDEF* ABC transporter specific for lipase in *Pseudomonas fluorescens* SIK W1. – *Journal of Bacteriology*, 181(6):1947-1852.
3. Akashi H., Gojobori T. 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. – *Proceedings National Academy of Science USA*, 99:3695-3700.
4. Alfano J.R. 2009. Roadmap for future research on plant pathogen effectors. – *Molecular Plant Pathology*, 10(6):805-813.
5. Alfano J.R., Collmer A. 2004. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. – *Annual Reviews in Phytopathology*, 42:385-414.
6. Angkawidjaja C., Kanaya S. 2006. Family I.3 lipase: bacterial lipases secreted by the type I secretion system. – *Cellular and Molecular Life Sciences*, 63:2804-2817.
7. Arnold R., Brandmaier S., Kleine F., Tischler P., Heinz E., Behrens S., Niinikoski A., Mewes H.W., Horn M., Rattei T. 2009. Sequence-based prediction of type III secreted proteins. – *PLoS Pathogens*, 5(4):e1000376.
8. Abfalg J., Gong J., Kriegel H.P., Pryakhin A., Wei T., Zimek A. 2007. Supervised ensembles of prediction methods for subcellular localization. – *Journal of Bioinformatics and Computational Biology*, 7(2):269-85.
9. Bellini A.V., Galli G., Luccheze G., Grandi G. 1995. Expression and secretion of mature human beta interleukin-1 in *Bacillus subtilis* and means and methods for its achievement, US Patent, 5437988.
10. Bendtsen J.D., Jensen L.J., Blom N., Von Heijne G., Brunak S. 2004a. Feature-based prediction of non-classical and leaderless protein secretion. – *Protein Engineering, Design*

and Selection, 17(4):349-356.

11. Bendtsen J.D., Nielsen H., von Heijne G., Brunak S. 2004b. Improved prediction of signal peptides: Signal P 3.0. – *Journal of Molecular Biology*, 340:783-795.
12. Bendtsen J.D., Kiemer L., Fausboll A., Brunak S. 2005. Non-classical protein secretion in bacteria. – *BMC Microbiology*, 5:58.
13. Bendtsen J.D., Wooldridge K.G. 2009. Non-classical secretion. – In: K.G. Wooldridge (ed.), *Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis.*, UK, Caister Academic Press:225-235.
14. Bingle L.E.H., Bailey C.M., Pallen M.J. 2008. Type VI secretion: a beginner's guide. – *Current Opinion in Microbiology*, 11(1):3-8.
15. Biro J.C., Benyó B., Sansom C., Szlávecz Á., Förös G., Micsik T., Benyó Z. 2003. A common periodic table of codons and amino acids. – *Biochemical and Biophysical Research Communications*, 306:408-415.
16. Birtalan S.C., Phillips R.M., Ghosh P. 2002. Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. – *Molecular Cell*, 9(5):971-80.
17. Blaylock B., Sorg J.A., Schneewind O. 2008. *Yersinia enterocolitica* type III secretion of YopR requires a structure in its mRNA. – *Molecular Microbiology*, 70(5):1210-22.
18. Bleasdale B., Lott P.J., Jaganathan A., Stevens M.P., Birtles R.J., Wigley P. 2009. The *Salmonella* pathogenicity island (SPI)-2-encoded type III secretion system is essential for survival of *Salmonella enterica* serovar Typhimurium in free-living amoebae. – *Applied and Environmental Microbiology*, 75(6):1793-1795.
19. Bönemann G., Pietrosiuk A., Diemand A., Zentgraf H., Mogk A. 2009. Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. – *The EMBO Journal*, 28(4):315-25.

20. Boyer F., Fichant G., Berthod J., Vandenbrouck T., Attree I. 2009. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? – *BMC Genomics*, 10:104.
21. Bu L., Xiao J., Lu L., Xu G., Li J., Zhao F., Li X., Wu J. 2009. The repertoire and evolution of ATF-binding cassette systems in *Synechococcus* and *Prochlorococcus*. – *Journal of Molecular Evolution*, 69(4):300-310.
22. Cambronne E.D., Roy C.R. 2006. Recognition and delivery of effector proteins into eukaryotic cells by bacterial secretion systems. – *Traffic*, 7:929-939.
23. Cao T.B., Saier M.H.Jr. 2001. Conjugal type IV macromolecular transfer systems of Gram-negative bacteria: organismal distribution, structural constraints and evolutionary conclusions. – *Microbiology*, 147:3201-3214.
24. Cascales E. 2008. The type VI secretion toolkit. – *EMBO Reports*, 9(8):735-741.
25. Cascales E., Christie P.J. 2003. The versatile bacteria Type IV secretion systems. – *Nature Reviews Microbiology*, 1:137-149.
26. Chan Y.H. 2005. *Biostatistics 303*. Discriminant analysis. – *Singapore Medicine Journal*, 46:54-62.
27. Chandran V., Fronzes R., Duquerroy S., Cronin N., Navaza J., Waksman G. 2009. Structure of the outer membrane complex of a type IV secretion system. – *Nature*, 462:1011-1015.
28. Chen S.L., Lee W., Hottes A.K., Shapiro L., McAdams H.H. 2004. Codon usage between genomes is constrained by genome-wide mutational processes. – *Proceedings of National Academy of Science*, 101(10):3480-3485.
29. Chervaux C., Holland I.B. 1996. Random and directed mutagenesis to elucidate the functional importance of helix II and F-989 in the C-terminal secretion signal of *Escherichia coli* hemolysin. – *Journal of Bacteriology*, 178:1232-1236.

30. Chibucos M.C., Tseng T.-T., Setubal J.C. 2009. Describing commonalities in microbial effector delivery using the gene ontology. – *Trends in Microbiology*, 17(7):312-319.
31. Christie P.J., Cascales E. 2005. Structural and dynamic properties of bacterial type IV secretion systems. – *Molecular Membrane Biology*, 22(1-2):51-61.
32. Choo K.H., Tan W.T., Ranganathan S. 2009. A comprehensive assessment of N-terminal signal peptides prediction methods. – *BMC Bioinformatics*, 10(Suppl 15): S2.
33. Chou K.C., Zhang C.T. 1995. Prediction of protein structural classes. – *Critical Reviews in Biochemistry and Molecular Biology*, 30:275-349.
34. Chung C.W., You J., Kim K., Moon Y., Kim H., Ahn J.H. 2009. Export of recombinant proteins in *Escherichia coli* using ABC transporter with an attached lipase ABC transporter recognition domain (LARD). – *Microbial Cell Factories*, 8:11.
35. Cochrane G., Akhtar R., Bonfield J., Bower L., Demiralp F., Faruque N., Gibson R., Hoad G., Hubbard T., Hunter C., Jang M., Juhos S., Leinonen R., Leonard S., Lin Q., Lopez R., Lorenc D., McWilliam H., Mukherjee G., Plaister S., Radhakrishnan R., Robinson S., Sobhany S., Hoopen P.T., Vaughan R., Zalunin V., Birney E. 2008. Petabyte-scale innovations at the European Nucleotide Archive. – *Nucleic Acids Research*, 37(Database issue):D19-25.
36. Coombes B.K. 2008. Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. – *Trends in Microbiology*, 17(3):89-94.
37. Cornelis G.R. 2002. The *Yersinia* Ysc-Yop ‘type III’ weaponry. – *Nature Reviews in Molecular Cell Biology*, 3:742-752.
38. Cover T.M., Thomas J. A. 1991. *Elements of information theory*. New York, John Wiley & Sons, Inc.
39. Dayhoff M.O., Schwartz R.M., Orcutt B.C. 1978. A model of evolutionary change in proteins. – In: M.O.Dayhoff (ed.) *Atlas of protein sequence and structure*. Vol. 5, suppl. 3, Silver Spring (MD): National Biomedical Research Foundation:345-352.

40. DeBuck E., Anné J., Lammertyn E. 2007. The role of protein secretion systems in the virulence of the intracellular pathogen *Legionella pneumophila*. – *Microbiology*, 153:3948-3953.
41. DePaz H.D., Sangari F.J., Bolland S., Garcia-Lobo J.M., Dehio C., de la Cruz F., Llosa M. 2005. Functional interactions between type IV secretion systems involved in DNA transfer and virulence. – *Microbiology*, 151:3505-3516.
42. Desvaux M., Hébraud M., Talon R., Henderson I.R. 2009. Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. – *Trends in Microbiology*, 17:139-145.
43. Economou A. 2002. Bacterial secretome: the assembly manual and operating instructions. – *Molecular Membrane Biology*, 19:159-169.
44. Emmanuelson O., Brunnak S., von Heijne G., Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. – *Nature Protocols*, 2:953-971.
45. Evdokimov A.G., Tropea J.E., Routzhan K.M., Copeland T.D., Wangh D.S. 2001. Structure of the N terminal domain of *Yersinia pestis* YopH at 20Å resolution. – *Acta Crystallographica Section D: Biological Crystallography*, 57:793-799.
46. Fernández L.A., deLorenzo V. 2001. Formation of disulphide bonds during secretion of proteins through the periplasmic-independent type I pathway. – *Molecular Microbiology*, 40(2):332-346.
47. Filloux A., Hachani A. Bleves S. 2008. The bacterial type VI secretion machine: yet another player for protein transport across membranes. – *Microbiology*, 154:1570-1583.
48. Fielding A.H. Bell J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – *Environmental Conservation*, 24(1):38-49.
49. Flintoft L. 2009. Microbial ecology: type III secretion systems: jumping into the reservoir. – *Nature Reviews Microbiology*, 7:177.

50. Foster P.G., Jermin L.S., Hickey A.A. 1997. Nucleotide comparison bias affects amino acid content in proteins coded by animal mitochondria. – *Journal of Molecular Evolution*, 44:282-288.
51. Frank A.C., Alsmark C.M., Tholleson M., Andersson S.G.E. 2005. Functional divergence and horizontal transfer of the type IV secretion systems. – *Molecular Biology and Evolution*, 22:1325-1336.
52. Furutani A., Takaoka M., Sanada H., Noguchi Y., Oku T., Tsuno K., Ochiai H., Tsug S. 2009. Identification of novel type III secretion effectors in *Xanthomonas oryzae* pv. *oryzae*. – *Molecular Plant-Microbe Interactions*, 22(1):96-106.
53. Garcia-Vallvé S., Romeu A., Palau J. 2000. Horizontal gene transfer in Bacterial and Archaeal complete genomes. – *Genome Research*, 10:1719-1725.
54. Garcia-Vallvé S., Guzman E., Montero M.A., Romeu A. 2003. HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. – *Nucleic Acids Research*, 31(1):187-189.
55. Gardy J. L., Laird M.R., Chen F., Rey S., Walsh C.J., Ester M., Brinkman F.S.L. 2005. PSORTb v.2.0: Expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. – *Bioinformatics*, 21(5):617-623.
56. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. – *Nucleic Acid Research*, 31(13):3784-3788.
57. Gentschev I., Fensterle J., Schmidt A., Potapenko T., Troppmair J., Goebel W., Rapp U.R. 2005. Use of a recombinant *Salmonella enterica* serovar Typhimurium strain expressing C-Raf for protecting against C-raf induced lung adenoma in mice. – *BMC Cancer*, 5:15.
58. Gerlach R.G., Hensel M. 2007. Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. – *International Journal of Medical Microbiology*, 297:401-415.

59. Ghosh P. 2004. Process of protein transport by the type III secretion system. – *Microbiology and Molecular Biology Reviews*, 68(4):771-795.
60. Goodarzi H., Torabi N., Najafabadi H.S., Archetti M. 2008. Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons. – *Gene*, 407(1-2):30-41.
61. Gophna U., Ron E.Z., Graur D. 2003. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. – *Gene*, 312:151-163.
62. Goudenège D., Avner S., Lucchetti-Miganeh C., Barloy-Hubler F. 2010. CoBaltDB: complete bacterial and archaeal orfomes subcellular localization database and associated resources. – *BMC Microbiology*, 10:88.
63. Gouridis G., Karamanou S., Gelis I., Kalodimos C.G., Economou A. 2009. Signal peptides are allosteric activators of the protein translocase. – *Nature*, 462(7271):363-367.
64. Grantham R. 1974. Amino acid difference formula to help explain protein evolution. – *Science*, 185:863-864.
65. Gromiha M.M., Suwa M. 2005. A simple statistical method for discriminating outer membrane proteins with better accuracy. – *Bioinformatics*, 21:961-968.
66. Gupta R.S. 2000. The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. – *FEMS Microbiology Reviews*, 24(4):367-402.
67. Haywood-Farmer E., Otto S.P. 2003. The evolution of genomic base composition in bacteria. – *Evolution*, 57(8):1783-1792.
68. Holland I.B. 2004. Translocation of bacterial proteins – an overview. – *Biochimica et Biophysica Acta*, 1694:5-16.
69. Holland I.B., Schmitt L., Young J. 2005. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). – *Molecular Membrane Biology*, 22(1-2):29-39.

70. Hueck C.J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. – *Microbiology and Molecular Biology Reviews*, 62(2):379-433.
71. Hung C.H., Huang H.L., Hsu K.T., Ho S.J., Ho S.Y. 2010. Prediction of non-classical secreted proteins using informative physicochemical properties. – *Interdisciplinary Science and Computational Life Science*, 2:263-270.
72. Izadi-Pruneyre N., Wolff N., Redeker V., Wandersman C., Delepierre M., Lecroisey A. 1999. NMR studies of the C-terminal secretion signal of the haem-binding protein, HasA. – *European Journal of Biochemistry*, 261:562-568.
73. Jenewein S., Holland B., Schmitt L. 2009. Type I bacterial secretion systems. – In: K.Wooldridge (ed.) *Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis.*, UK, Caister Academic Press:45-66.
74. Jitprasutwit S., Thaewpia W., Muangsombut V., Lulitanond A., Leelayuwat C., Lertmemongkolchai G., Korbsrisate S. 2010. Effect of acidic pH on the invasion efficiency and the type III secretion system of *Burkholderia thailandensis*. – *The Journal of Microbiology*, 48(4):526-532.
75. Kelley L.A., MacCallum R.M., Sternberg M.J.E. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. – *Journal of Molecular Biology*, 299:499-520.
76. Kelley L.A., Stenberg M.J.E. 2009. Protein structure prediction on the web: a case study using the Phyre server. – *Nature Protocols*, 4(3):363-371.
77. Kelley S.T., Cassirerb E.F., Weiserc G.C., Safaee S. 2007. Phylogenetic diversity of *Pasteurellaceae* and horizontal gene transfer of leukotoxin in wild and domestic sheep. – *Infection, Genetics and Evolution*, 7(1):13-23.
78. Klecka W.R. 1980. *Discriminant analysis, quantitative applications in social sciences.* Beverly Hills, Sage publications Inc.

79. Knight R.D., Freeland S.J., Landweber L.F. 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. – *Genome Biology*, 2(4):research0010.
80. Koonin E.V., Aravind L., Kondrashov A.S. 2000. The impact of comparative genomics on our understanding of evolution. – *Cell*, 101:573-576.
81. Koronakis V., Hughes C., Koronakis E. 1991. Energetically distinct early and late stages of HlyB/HlyD-dependent secretion across both *Escherichia coli* membranes. – *The EMBO journal*, 10(11):3263-3272.
82. Krissinel E. 2007. On the relationship between sequence and structure similarities in proteomics. – *Bioinformatics*, 23(6):717-723.
83. Kyte J., Doolittle R.F. 1982. A simple method for displaying the hydropathic character of a protein. – *Journal of Molecular Biology*, 157:105-132.
84. Lacroix B., Kozlovsky S.V., Citovsky V. 2008. Recent patents on *Agrobacterium*-mediated gene and protein transfer, for research and biotechnology. – *Recent Patents of DNA and Gene Sequences*, 2(2):69-81.
85. Leary R.H. , Rosen J.B., Jambeck P. 2004. An optimal structure discriminative index for protein fold recognition. – *Biophysical Journal*, 86:411-419.
86. Lee V.T., Schneewind O. 1999. Type III machines of pathogenic yersiniae secrete virulence factors into the extracellular milieu. – *Molecular Microbiology*, 31(6):1619-1629.
87. Leiman P.G., Basler M., Ramagopal U.A., Bonanno J.B., Sauder J.M., Pukatzki S., Burley S.K., Almo S.C., Mekalanos J.J. 2009. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. – *Proceedings of National Academy of Science USA*, 106(11):4154-4159.
88. Levitt M. 1978. Conformational Preferences of Amino Acids in Globular Proteins. – *Biochemistry*, 17:4277-4285.

89. Li C.M. 2007. Type III secretion system of phytopathogenic bacterium *Pseudomonas syringae*: from gene to function. Thesis. Finland. University of Helsinki. 64pp.
90. Li H. 2001. Thermoresponsive secretion of the extracellular enzyme levansucrases from *Pseudomonas syringae*. Thesis. Germany. Fachbereich Biologie der Phillips-Universität Marburg. 134pp.
91. Locher M., Lehnert B., Krauss K., Heesemann J., Groll M., Wilharm G. 2005. Crystal structure of the *Yersinia enterocolitica* Type III secretion chaperone SycT. – Journal of Biological Chemistry, 280(35):31149-31155.
92. Lory S. 1992. Determinant of extracellular protein secretion in Gram-negative bacteria. – Journal of Bacteriology, 174(1):3423-3428.
93. Lotter H., Rüssmann H., Heesemann J., Tannich E. 2004. Oral vaccination with recombinant *Yersinia enterocolitica* expressing hybrid type III proteins protects gerbils from amebic liver abscess. – Infection and Immunology, 72(12):7318-7321.
94. Löwer M., Schneider G. 2009. Prediction of type III secretion signals in genomes of Gram-negative bacteria. – PLoS ONE, 4(6):e5917.
95. Ma W., Dong F.F.T., Stavrinides J., Guttman D.S. 2006. Type III effector diversification via both pathoadaptation and horizontal transfer in response to a coevolutionary arms race. – PLoS Genetics, 2(12):e209.
96. Mann P., Goebel E., Barbarich J., Pilione M., Kennett M., Harvill E. 2007. Use of a genetically defined double mutant strain of *Bordetella bronchiseptica* lacking adenylate cyclase and type III secretion as a live vaccine. – Infection and Immunology, 75(7):3665-72.
97. Marteyn B., West N.P., Browning D.F., Cole J.A., Shaw J.G., Palm F., Mounier J., Prévost M.C., Sansonetti P., Tang C.M. 2010. Modulation of *Shigella* virulence in response to available oxygen *in vivo*. – Nature, 465:355-358.
98. Masi M., Wandersman C. 2010. Multiple signals direct the assembly and function of a type 1 secretion system. – Journal of Bacteriology, 192(15):3861-3869.

99. McCann H.C., Guttman D.S. 2008. Evolution of the type III secretion system and its effectors in plant-microbe interactions. – *New Phytologist*, 177:33-47.
100. McLachlan A.D. 1977. Analysis of periodic patterns in amino acid sequences: collagen. – *Biopolymers*, 16:1271-1297.
101. Mergulhão F.J., Summers D.K., Monteiro G.A. 2005. Recombinant protein secretion in *Escherichia coli*. – *Biotechnology Advances*, 23(3):177-202.
102. Michel G.P.F., Voulhoux R. 2009. The Type II secretory system (T2SS) in Gram-negative bacteria: a molecular nanomachine for secretion of Sec and Tat-dependent extracellular proteins. – In: K.G.Wooldridge (ed.), *Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis.*, UK, Caister Academic Press:67-92.
103. Paila U., Kondam R., Ranjan A. 2008. Genome bias influences amino acid choices: analysis of amino acid substitution and re-compilation of substitution matrices exclusive to an AT-biased genome. – *Nucleic Acids Research*, 36:6664-6675.
104. Pasquier C.M., Promponas V.I., Varvayannis N.J., Hamodrakas S.J. 1998. A web server to locate periodicities in a sequence. – *Bioinformatics*, 14:479-750.
105. Pugsley A. 1993. The complete general secretory pathway in Gram-negative bacteria. – *Microbiological Reviews*, 57(1):50-108.
106. Pukatzki S., Ma A.T., Sturtevant D., Krastins B., Sarracino D., Nelson W.C., Heidelberg J.F., Mekalanos J.J. 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. – *Proceedings of National Academy of Science USA*, 103(5):1528-1533.
107. Pukatzki S., Ma A.T., Revel A.T., Sturtevant D., Mekalanos J.J. 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. – *Proceedings National Academy of Science USA*, 104:15508-15513
108. Pukatzki S., McAuley S.B., Miyata S.T. 2009. The type VI secretion system: translocation of effectors and effector-domains. – *Current Opinion in Microbiology*, 12:11-17.

109. Rolán H.G., Xavier M.N., Santos R.L., Tsois R.M. 2009. Natural antibody contributes to host defense against an attenuated *Brucella abortus virB* mutant. – *Infection and Immunology*, 77(7):3004-3013.
110. Ren Q., Paulsen I.T. 2005. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. – *PLoS Computational Biology*, 1(3):e27.
111. Rice P., Longden I., Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. – *Trends in Genetics*, 16:276-277.
112. Rocha E.P.C., Touchon M., Feil E.J. 2006. Similar compositional biases are caused by very different mutational effects. – *Genome Research*, 16:1537-1547.
113. Saier M.H.Jr. 2006. Protein secretion systems in Gram-negative bacteria. – *Microbe*, 1(9):414-419.
114. Samudrala R., Heffron F., McDermott J.E. 2009. Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type III secretion systems. – *PLoS Pathogens*, 5:e1000375.
115. Sandkvist M. 2001. Type II secretion and pathogenesis. – *Infection and Immunology*, 69(6):3523-3535.
116. Sapriel G., Wandersman C., Delepelaire P. 2002. The N terminus of the HasA protein and the SecB chaperone cooperate in the efficient targeting and secretion of HasA via the ATP-binding cassette transporter. – *Journal of Biological Chemistry*, 277(8):6726-6732.
117. Schnoes A.M., Brown S.D., Dodevski I., Babbitt P.C. 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. – *PloS Computational Biology*, 5(12): e1000605.

118. Shanks J., Burtnick M.N., Brett P.J., Waag D.M., Spurgers K.B., Ribot W.J., Schell M.A., Panchal R.G., Gherardini F.C., Wilkinson K.D., Deshazer D. 2009. *Burkholderia mallei tssM* encodes a putative deubiquitinase that is secreted and expressed inside infected RAW 264.7 murine macrophages. – *Infection and Immunology*, 77(4):1636-48.
119. Sharp P.M., Li W.H. 1987. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. – *Nucleic Acids Research*, 15:1281-1295.
120. Shrivastava S., Mande S.S. 2008. Identification and functional characterization of gene components of type VI secretion system in bacteria genomes. – *PLoS ONE*, 3(8):e2955.
121. Singer G.A.C., Hickey D.A. 2000. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. – *Molecular Biology and Evolution*, 17(11):1581-1588.
122. Sleator R.D., Walsh P. 2010. An overview of in silico protein function prediction. – *Archives of Microbiology*, 192:151-155.
123. Snyder L.A.S., Loman N.J., Fütterer K, Pallen M.J. 2008. Bacterial flagellar diversity and evolution: seek simplicity and distrust it? – *Trends in Microbiology*, 17(1):1-5.
124. Sobolevsky Y., Trifonov E.N. 2005. Conserved sequences of prokaryotic proteomes and their compositional age. – *Journal of Molecular Evolution*, 61:591-6.
125. Song C., Kumar A., Saleh M. 2009. Bioinformatic comparison of bacterial secretomes. – *Genomics Proteomics Bioinformatics*, 7(1-2):37-46.
126. Sorimachi K., Okayasu T. 2008. Codon evolution is governed by linear formulas. – *Amino Acids*, 34:661-668.
127. Soto M.A., Tohac J. 1983. Conservation of physico-chemical amino acid properties during the evolution of proteins. – *Origins of Life*, 13:147-152.

128. Sperandio V., Mellies J.L., Nguyen W., Shin S., Kaper J.B. 1999. Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. – Proceedings of National Academy of Science USA, 96:15196-15201.
129. Sueoka N. 1961. Compositional correlation between deoxyribonucleic acid and protein. Cold Spring Harbor Symposia on Quantitative Biology, 26:35-43.
130. Stavrinides J., Ma W., Guttman D.S. 2006. Terminal reassortment drives the quantum evolution of type III effectors in bacterial pathogens. – PLoS Pathogens, 2(10):e104:0913-0921.
131. The UniProt Consortium. 2008. The Universal Protein Resource (UniProt). – Nucleic Acids Research, 36(Database issue):D190-D195.
132. Tjalsma H., Antelmann H., Jongbloed J.D.H., Braun P.G., Darmon E., Dorenbos R., Dubois J.Y.F., Westers H., Zanen G., Quax W.J., Kuipers O.P., Bron S., Hecker M., vanDijl J.M. 2004. Proteomics of protein secretion by *Bacillus subtilis*: separating the “secrets” of the secretome. – Microbiology and Molecular Biology Reviews, 68:207-233.
133. Tomii K., Kanehisa M. 1998. A comparative analysis of ABC transporters in complete microbial genomes. – Genome Research, 8(10):1048-1059.
134. Trček J., Wilharm G., Jacobi C.A., Heesemann J. 2002. *Yersinia enterocolitica* YopQ: strain-dependent cytosolic accumulation and post-translational secretion. – Microbiology, 148:1457-1465.
135. Trifonov E.N. 1999. Glycine clock: Eubacteria first, Archaea next, Protoctista, Fungi, Planta and Animalia at last. – Gene Therapy and Molecular Biology, 4:313-322.
136. Trifonov E.N., Gabdank I., Barash D., Sobolevsky Y. 2006. *Primordia vita*. Deconvolution from modern sequences. – Origin of Life and Evolution of Biospheres, 36:559-565.
137. Troisfontaines P., Cornelis G.R. 2005. Type III secretion: more systems than you think. – Physiology, 20:326-339.

138. Tseng T.T., Tyler B.M., Setubal J.C. 2009. Protein secretion systems in bacterial-host associations and their description in the gene ontology. – *BMC Microbiology*, 9: (Suppl 1):S2.
139. Turner D., Wooldridge K.G., Ala'Aldeen D. 2009. Protein secretion and pathogenesis in *Neisseria meningitidis*. – In: K.G.Wooldridge (ed.), *Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis.*, UK, Caister Academic Press:347-360.
140. Vilches S., Jimenez N., Tomás J.M, Merino S. 2009. *Aeromonas hydrophila* AH-3 type III secretion system expression and regulatory network. – *Applied and Environmental Microbiology*, 75(19):6382-6392.
141. Vinatzer B.A., Jelenska J., Greenberg J.T. 2005. Bioinformatics correctly identifies many type III secretion substrates in the plant pathogen *Pseudomonas syringae* and the biocontrol isolate *P. fluorescens* SBW25. – *Molecular Plant-Microbe Interactions*, 18:877-888.
142. Wilharm G., Lehmann V., Krauss K., Lehnert B., Richter S., Ruckdeschel K., Heesemann J., Trulzsch K. 2004. *Yersinia enterocolitica* Type III secretion depends on the proton motive force but not on the flagellar motor components MotA and MotB. – *Infection and Immunology*, 72(7):4004-4009.
143. Winans S.C. 2004. Reciprocal regulation of bioluminescence and type III protein secretion in *Vibrio harveyi* and *Vibrio parahaemolyticus* in response to diffusible chemical signals. – *Journal of Bacteriology*, 186(12):3674-3676.
144. Wolfgang M.C., Lee V.T., Glimore M.E., Lory S. 2003. Coordinate regulation of bacterial virulence genes by a novel adenylate cyclase-dependent signaling pathway. – *Cell*, 4(2):253-263.
145. Wooldridge K.G. 2009. Preface. – In: K.G.Wooldridge (ed.), *Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis.*, UK, Caister Academic Press:xii.

146. Wright F. 1990. The 'effective number of codons' used in a gene. – *Gene*, 87:23-29.
147. Xia X., Xie Z. 2002. Protein structure, neighbour effect, and a new index of amino acid dissimilarities. – *Molecular Biology and Evolution*, 19:58-67.
148. Yu X.J., McGourty K., Liu M., Unsworth K.E., Holden D.W. 2010. pH sensing by intracellular *Salmonella* induces effector translocation. – *Science*, 328(5981):1040-1043.
149. Zalucki Y.M., Beacham I.R., Jennings M.P. 2009. Biased codon usage in signal peptides: a role in protein export. – *Trends in Microbiology*, 17:146-150.
150. Zhang F., Greig D.I., Ling V. 1993. Functional replacement of the hemolysin A transport signal by a different primary sequence. – *Proceedings National Academy of Science USA*, 90:4211-4215.
151. Zhou M., Boekhorst J., Francke C., Siezen R.J. 2008. LocateP: genome-scale subcellular-location predictor for bacterial proteins. – *BMC Bioinformatics*, 9:173.
152. Zuo Y.C., Li Q.Z. 2010. Using *K*-minimum increment of diversity to predict secretory proteins of malaria parasite based on groupings of amino acids. – *Amino Acids*, 38(3):859-867.

Pielikums