TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT

BIOINFORMAATIKA ÕPPETOOL

Antimikroobse resistentsuse tuvastamine ja ennustamine Illumina sekveneerimisandmetest

Bakalaureusetöö

12 EAP

Heleri Savolainen

Juhendaja teadur Reidar Andreson

TARTU 2025

INFOLEHT

Antimikroobse resistentsuse tuvastamine ja ennustamine Illumina sekveneerimisandmetest

Antimikroobne resistentsus (AMR) on kujunenud ülemaailmseks rahvatervise probleemiks. Kiire ja täpne resistentsuse tuvastamine on oluline, et alustada sobivat ravi varakult ning vähendada empiirilise ravi ajaakna suurust. Terve genoomi sekveneerimine (WGS) on saanud oluliseks bioinformaatiliseks tööriistaks, võimaldades määrata resistentsusprofiile tõhusalt.

Käesoleva töö põhieesmärk oli tuvastada kahe bioinformaatilise tööriistaga AMR-i, millest esimene põhineb genoomipõhisel naabertüpiseerimisel (RASE) ning teine tugineb resistentsusgeenide ära tundmisel (ResFinder). Täpsemalt ennustati resistentsust ja tundlikkust *Klebsiella pneumoniae* ja *Escherichia coli* isolaatidest ja simuleeritud metagenoomsetest proovidest. Leiti, et ResFinder oli täpsem resistentsuse ennustamisel isolaadipõhistes proovides, teisalt RASE osutus usaldusväärsemaks tundlikkuse ennustamisel. Mõlema tööriista täpsus langes märgatavalt simuleeritud metagenoomsete andmete töötlemisel.
Märksõnad: antibiootikumid, antimikroobne resistentsus, WGS
CERCS: B110 Bioinformaatika, meditsiiniline informaatika, biomatemaatika, biomeetria

Detection and prediction of antimicrobial resistance from Illumina sequencing data

Antimicrobial resistance (AMR) has become a global public health concern. Rapid and accurate resistance detection is essential to initiate appropriate treatment early and shorten the empirical treatment window. Whole genome sequencing (WGS) is a valuable tool for efficiency determining resistance profiles.

This thesis aimed to evaluate two bioinformatic tools, RASE is based on genomic neighbour typing and ResFinder for detecting resistance genes. The study focused on predicting susceptibility and resistance in *Klebsiella pneumoniae* and *Escherichia coli* isolates and simulated metagenomic samples. Results showed that ResFinder was more accurate in predicting resistance in isolates, whereas RASE was more reliable for predicting susceptibility. Both tools showed decreased accuracy when applied to simulated metagenomic data.
Keywords: antibiotics, antimicrobial resistance, WGS
CERS: B110 Bioinformatics, medical informatics, biomathematics, biometrics

SISUKORD

INFOLEHT **2**
KASUTATUD LÜHENDID **4**
SISSEJUHATUS **5**
1. KIRJANDUSE ÜLEVAADE **6**
1.1. Antibiootikumide taust 6
1.2. Empiiriline diagnoosimine ja piirangud antibiootikumide kasutamisel 6
1.3. Antimikroobse resistentsuse (AMR) tuvastamine 7
1.3.1. AMR bioloogilised mehhanismid 9
1.3.2. *Klebsiella pneumoniae* ja *Escherichia coli* resistentsus antibiootikumidele 9
1.3.3. AMR tuvastamine WGS meetodil 10
1.4. Ülevaade bioinformaatilistest tööriistadest AMR-i ennustamiseks ja tuvastamiseks 11
1.4.1. Kover 12
1.4.2. ResFinder 14
1.4.3. Resistance-associated sequence elements (RASE) 16
2. EKSPERIMENTAALOSA **19**
2.1. Töö eesmärgid 19
2.2. Materjal ja metoodika 19
2.2.1. Tarkvara ja andmebaasid 19
2.2.2. Töövoog 20
2.2.3. RASE programmi töövoog 21
2.2.4. ResFinderi programmi töövoog 22
2.3. Tulemused 23
2.3.1. Simuleeritud tausta-DNA resistentsuse ja tundlikkuse ennustamine erinevatele antibiootikumidele RASE ja ResFinderi tarkvarade põhjal 23
2.3.2. RASE ja ResFinderi tarkvarade AMR-i ennustused *Klebsiella pneumoniae* näitel 24
2.3.3. RASE ja ResFinderi tarkvarade AMR-i ennustused *Escherichia coli* näitel 27
2.3.4. RASE ja ResFinderi töövoogude võrdlus ajakulu alusel 30
2.4. Arutelu 32
KOKKUVÕTE **36**
SUMMARY **37**
KIRJANDUSE LOETELU **38**
KASUTATUD VEEBIAADRESSID **42**
LISAD **43**
Lisa 1. 43
Lisa 2. 47
Lisa 3. 50
Lisa 4. 51
Lisa 5. 52
Lisa 6. 53
Lisa 7. 54
Lisa 8. 55
Lisa 9. 56
LIHTLITSENTS **57**

KASUTATUD LÜHENDID

AMR – antimikroobne resistentsus (*antimicrobial resistance*)

ARG – antimikroobne resistentsusgeen (*antimicrobal resistance gene*)

AST – antimikroobse tundlikkuse testimine (*antimicrobal susceptibility testing*)

BLAST – homoloogiaotsingu tarkvara (*basic local alignment search tool*)

BV-BRC – bakterite ja viiruste bioinformaatika andmekeskus (*Bacterial and Viral Bioinformatics Resource Center*)

CGE – genoomne epidemioloogia keskus (*Center for Genomic Epidemiology*)

ENA – Euroopa Nukleotiidarhiiv (*European Nucleotide Archive*)

FASTA – tekstipõhine failiformaat nukleotiidsete või aminohappeliste järjestuste salvestamiseks

FASTQ – tekstipõhine failiformaat talletamaks nukleotiidset järjestust ja sellele vastavaid kvaliteediskoore

KMA – k-meeri joondamine (*k-mer alignment*)

MIC – minimaalne inhibeeriv kontsentratsioon (*minimal inhibitory concentration*)

MDR – mitme ravimi resistentsus (*multi-drug resistance*)

NGS – teise põlvkonna sekveneerimine (*next-generation sequencing*)

PATRIC – patosüsteemide integreerimiskeskus (*Pathosystems Resource Integration Center*)

PBP – penitsilliini siduv valk (*penicillin-binding protein*)

PCR – polümeraasi ahelreaktsioon (*polymerase chain reaction*)

RASE – resistentsusega seotud järjestuselemendid (*resistance-associated sequence elements*)

RT-PCR – reaalaja polümeraasi ahelreaktsioon (*real-time PCR*)

SCM – genoomseid mustreid tuvastav algoritm (*set covering machine*)

WGS – terve genoomi sekveneerimine (*whole genome sequencing*)

SISSEJUHATUS

Antimikroobne resistentsus (AMR) on tänapäeval laialdaselt levinud rahvatervise probleem. Laialdane AMR-i levik piirab antibiootikumide efektiivsest kasutamist bakteriaalsete nakkushaiguste ravimisel. Mikroorganismide resistentsuse teke antibiootikumide suhtes on otseselt seotud suremuse kasvuga maailmas. Traditsioonilised resistentsuse määramise meetodid on usaldusväärsed, kuid nõuavad mikroorganismi eelnevat kasvatamist, mis muudab need ajakulukaks. Kriitilisematel juhtudel võib resistentse tüve hiline avastamine viia ebaefektiivse ravini, mis omakorda võib põhjustada tüsistusi.

Viimaste aastakümnetega on üha enam kasutusele võetud terve genoomi sekveneerimisel (WGS) põhinevaid meetodeid, mis võimaldavad resistentsusprofiile kiiremini tuvastada. Selleks on loodud mitmeid tarkvaralisi lahendusi, millest üks on resistentsusgeenidel põhinev ResFinder ning teine on resistentsusega seotud järjestuselementidel (RASE) põhinev programm. Mõlema bioinformaatilise tööriista kasutamine on perspektiivikas, ent nende täpsus ja sobivus erinevates olukordades pole veel lõplikult välja selgitatud.

AMR ennustustööriistade järele on suur vajadus ning oluline on hinnata olemasolevate lahenduste töökindlust erinevate andmestikega. Rakendatavate tarkvarade kasutamisel on oluline mõista, kuidas programmid käituvad keerukate proovidega, näiteks inimese verega. Lisaks sellele on oluline tööriistade tundlikkus määrata resistentseid ja tundlikke proove kasutades erinevaid andmebaase.

Käesoleva töö üldeesmärk on hinnata kahe eelmainitud antimikroobse resistentsuse ennustamise tarkvara toimimist teise põlvkonna sekveneerimisandmete (NGS) põhjal. Praktilises töös keskenduti *Klebsiella pneumoniae* ja *Escherichia coli* resistentsuse ja tundlikkuse määramisele isolaadipõhistest proovidest ning simuleeritud metagenoomsetest proovidest. Töö teoreetilise osa eesmärk oli anda ülevaade antimikroobse resistentsuse tuvastamise meetoditest ning tutvustada erinevaid bioinformaatilisi AMR tööriistu.

1. KIRJANDUSE ÜLEVAADE

1.1. Antibiootikumide taust

Antibiootikumid on ravimid, mille eesmärk on bakterite levikut tõkestada. Inimesed elavad mikroobimaailma keskkonnas, kus osad mikroobid on inimesele vajalikud teised mitte. Selle tagajärjel on evolutsiooni käigus tekkinud inimkonnale naturaalne immuunsus, kuid sellest ainult ei piisa – vajadus ka millegi kolmanda järgi (Pidcock ja Piotrowski, 2023). Viimase 100 aasta jooksul on antibiootikumide areng olnud hüppeline. Kui 1900. aastatel surid inimesed Ameerika ühendriikides peamiselt tuberkuloosi, kopsupõletikku ja düsenteeriasse, siis 100 aastat hiljem on kaks peamist surmapõhjustajat vähk ning südame ja veresoonkonna haigused (Nguyen et al., 2020).

Inimese eluiga on pikenenud ligi 20 aastat tänu antimikroobsete ravimite laialdasele kasutusele, kuid antud ravimite edasine arendus on alates 1950. aastatest olnud üsna tagasihoidlik. Lisaks suurendab ravimite väärkasutamine antimikroobse resistentsuse teket, mis on tõusev trend üle maailma (Hutchings et al., 2019). Resistentsete bakteritüvede infektsioonide tagajärjel sureb maailmas ligi 700 000 inimest aastas (Hyun et al., 2020).

Pärast penitsilliini avastamist muutus antibiootikumide kasutus märkimisväärselt. Sünteesiti uusi baktereid tõkestava levikuga ravimeid, näiteks aktinomütseete, streptomütsiine, β-laktaame jne (Hutchings et al., 2019). Laiahaardeline antibiootikumide kasutamine on aidanud kontrolli all hoida erinevaid bakteriaalseid infektsioone, mille vastu naturaalse immuunsusega ei saa (Pidcock, ja Piotrowski, 2023).

1.2. Empiiriline diagnoosimine ja piirangud antibiootikumide kasutamisel

Empiiriline diagnoosimine on kliinilises meditsiinis oluline etapp, mis täidab lünga kiire ravi alustamise ja laboridiagnostika tulemuste vahel. See tugineb arstide kliinilisele kogemusele ja tõenäosuslikele hinnangutele, võimaldades määrata ravi enne, kui patogeeni kohta on saadaval täpsemad andmed (MacFadden et al., 2014). Näiteks, ravimi määramisel saab lähtuda varasemalt kogutud andmetele. Kui täheldatakse, et konkreetne patogeen on mõjutatud ühe kindla antibiootikumi poolt, kasutatakse viimast uute patsientide ravimiseks (Břinda et al., 2020). Teisalt saab kasutada infektsioonide ravimisel mitme antibiootikumi kombinatsiooni: trimetoprim ja sulfametoksasool. Foolhappe süntees blokeeritakse kahes järjestikuses etapis, mille kaudu takistatakse DNA sünteesi ja bakterite paljunemist (Tenover, 2006).

Tänapäeval on leitud võimalusi, kuidas kiirendada haigustekitaja tuvastamist (MacFadden et al., 2014). Diagnostika ja ravivaliku vahele jääb empiiriline ajaaken, kus arstid tuginevad tõenäosuslikele hinnangutele, enne kui on täpsemad tulemused uuritava mikroobi tundlikkuse kohta saadaval (MacFadden et al., 2014). Statistilise analüüsi abil kogutakse andmeid sümptomite ja ravitulemuste kohta, mis võimaldab leida korduvaid mustreid ja hinnata ravi efektiivsust (TechTarget, 2024). Statistilise analüüsi tulemused on kasulikud ka patsiendile, kuna kiire patogeeni (mitte)määramine võib osutuda patsiendile eluliselt tähtsaks (MacFadden et al., 2014).

Empiirilise diagnoosimise üheks peamiseks puuduseks on antibiootikumi resistentsuse tuvastamise keerukus, kuna sageli puuduvad usaldusväärsed andmed patogeenitüve resistentsusprofiili kohta (Carroll et al., 2025). Nakkushaiguste leviku uurimine molekulaarsel tasandil (näiteks sekveneerimine) aitab määrata kõrge riskitasemega mikroobi tüvesid (Břinda et al., 2020). Ühe näitena võib tuua pneumokokk-nakkuste ravi β-laktaamidega. Pneumokokid omavad antud antibiootikumi klassi suhtes laialdast resistentsust. Teadlased leidsid meetodi, mis määrab pneumokokkide resistentsuse penitsilliini siduvate valkude ehk PBP tüüpide uurimise kaudu. Uuringud näitasid tugevat seost resistentsuse ja PBP-de vahel. Kuna bakter kasutab antibiootikumi sidumiseks PBP-sid, siis leiti, et PBP tüüpide järjestuste analüüs võimaldas ennustada 98% täpsusega madalaimat antibiootikumi kogust (MIC), mida on vaja, et takistada konkreetse bakteritüve kasvu (Kowalska-Krochmal ja Dudek-Wicher, 2021; Li et al).

Molekulaarsed tööriistad on kriitilise tähtsusega olukordades, kus otsesed meetodid patogeeni tundlikkuse määramiseks puuduvad (Břinda et al., 2020). Seetõttu on molekulaarsed meetodid, nagu näiteks sekveneerimine, hädavajalikud, et tagada antibiootikumide tõhus kasutamine. Uuringud on näidanud, et antibiootikumi resistentsust saab ennustada ka siis, kui teadaolevad resistentsusgeenid puuduvad, rakendades bakteripopulatsiooni struktuuride uurimist või leides tüvede geneetilist sarnasust (Břinda et al., 2020; Nguyen et al., 2020).

1.3. Antimikroobse resistentsuse (AMR) tuvastamine

Antimikroobse resistentsuse ehk AMR-i üheks põhjustajaks on bakteri enda molekulaarsed mehhanismid, mis võimaldavad bakteril arendada resistentsust antibiootikumide vastu. Rakusisesed protsessid hõlmavad näiteks bakteri ülekantavaid AMR-geene, resistentsusgeenide ülesreguleerimist, poriinide modifitseerimist või kadumist (Ellington et al., 2017). AMR võib tekkida ka mitme teguri koostoimel, kuid üheks kriitiliseks põhjuseks on vajalikkus kasutada laiaspektriga antibiootikumi, et infektsiooni ohjeldada. Sellepärast võib nakkushaiguse põhjustaja

evolutsioneeruda resistentseks, mis tekitab probleemi nii arstidele kui ka patsiendile endale (Pidcock ja Piotrowski, 2023).

Antimikroobse resistentsuse tuvastamine algab antibiootikumi paneeli tegemisest. Selleks on vajalik proovi olemasolu, et organismi fenotüüpi määrata (Nguyen et al., 2020). Eelmainitud proovi tuleb näiteks polümeraasi ahelreaktsiooni (PCR) abil amplifitseerida, et uuritavat DNA-d saaks kindlate markerite vastu võrrelda (Anjum et al., 2017).

Nii PCR-il kui ka reaalaja polümeraasiahelreaktsioonil (RT-PCR) on väga oluline koht antimikroobse resistentsuse testimisel. Tegemist on väga levinud meetodiga, mida kasutatakse ka Euroopa epidemioloogilistel uuringutel, kui 2015. aasta novembris leiti Hiinas plasmiidne kolistiini resistentsusgeen *Escherichia coli* tüves. Seda detekteeriti 2015. aasta detsembris ka Taanis, kus inimene oli nakkunud pärast kanaliha söömist (Anjum et al., 2017; Hasman et al., 2015). Epidemioloogilist riski uuriti lisaks PCR-ile ka üle-genoomsete meetoditega, sh ka ResFinderi ja PlasmidFinderiga (Hasman et al., 2015). PCR-i rakendamine on siiamaani ülemaailmselt kasutusel ja vajalik, kuigi edaspidiselt mainitud meetod võib tulevikus olla kuluefektiivsem (Anjum et al., 2017). Üheks võimalikuks lähenemiseks on kasutada bioinformaatilisi tööriistu AMR-i tuvastamiseks (Nguyen et al., 2020).

Peamine väljakutse AMR-i tuvastamisel bioinformaatiliselt on usaldusväärsete andmebaaside leidmine. Varasemalt dokumenteeritud resistentsusfenotüübid, mis on tuvastatud näiteks PCR-i või DNA mikrokiibi abil, on võimalik otseselt siduda kindla geeniga. Selliste geenide lisamine sihtandmebaasi on märkimisväärselt lihtsam. Paraku iseloomustab resistentsusmehhanisme suur geneetiline varieeruvus, mida pole võimalik ainult eelmainitud meetoditega tuvastada. Seetõttu on vaja põhjalikke andmebaase, mis hõlmaksid erinevaid geneetilisi variatsioone sealhulgas punktmutatsioone ja seni kirjeldamata mehhanisme (Anjum et al., 2017).

AMR-i kindaks tegemiseks rakendatakse viimastel aastatel lisaks klassikalistele bioinformaatilistele meetoditele ka masinõppe mudeleid, mis aitavad tuvastada resistentsust. Kasutatakse treeningandmestikke, mis on tehtud näiteks sekveneeritud täisgenoomide järjestuste (WGS) või pangenoomide alusel (Nguyen et al., 2020). Hyun ja kolleegide poolt tehtud AMR uuringus analüüsiti pangenoome ning tuvastati 45 varem teadaoleva AMR geeni ja resistentsusfenotüübi seost (Hyun et al., 2020). Treeningandmestike eesmärgiks on masinale „õpetada“, millised tunnused viitavad resistentsuse olemasolule ja seejärel ennustada resistentsuse olemasolu juba päris proovides (Nguyen et al., 2020).

1.3.1. AMR bioloogilised mehhanismid

Bakteri resistentsus mingi aine vastu tekib siis, kui bakteritüvi suudab paljuneda tingimustes, kus aine tavaliselt bakteri hävitaks. Tavaliselt on resistentsus tingitud bakteris aset leidnud geneetilisest muutusest, mille biokeemiline mehhanism on teada (Cloete, 2003). Bakterite resistentsust antimikroobse aine vastu võivad põhjustada ka spontaansed mutatsioonid geneetilises koodis. Erinevad bakterid saavad mõjutada valkude sihtmärklaudu, reguleerida nii poriinide kui ka pumpade aktiivsust viies antimikroobset ainet rakust välja (Elías-López et al., 2024; Tenover, 2006). Euroopas enim kasutatud antibiootikumide rühmadeks olid 2023. aasta andmetel Euroopas penitsilliinid, makroliidid, tsefalosporiinid (sh teised β-laktaamid), tetratsükliinid ja kinoloonid (ECDC, 2023).

Eelmainitud antibiootikumide rühmadel on erinevad mehhanismid kasutamaks ära bakterite elutsüklit. Neli peamist toimemehhanismi on valgusünteesi-, nukleiinhappe- ja metaboolsete radade inhibitsioon ning rakukesta sünteesi häirimine (Hussein et al., 2024; Tenover, 2006). Valgusünteesi inhibeerides saab kasutada ära bakteri 30S või 50S ribosomaalset subühikut, mida teevad nii makroliidid kui ka aminoglükosiidid (Drlica ja Zhao, 1997; Tenover, 2006). 30S ribosoomi subühik osaleb mRNA tõlkimisel valkudeks, kusjuures 50S subühik aitab ühendada kasvavat peptiidahelat. Nukleiinhappe sünteesi võivad inhibeerida kinoloonid, kasutades ära DNA güraasi tööd. Täpsemalt on güraas vajalik DNA lahti pakkimiseks DNA replikatsiooni ajal ning selle pärssimisel takistatakse edasist DNA sünteesi (McManus, 1997; Tenover, 2006). β-laktaamsed antibiootikumid, mille alla kuuluvad nii penitsilliinid, tsefalosporiinid ja karbapeneemid, lüüsivad rakumembraani, kus kinnitutakse peptidoglükaanile inhibeerides selle edasist sünteesi (Tenover, 2006).

1.3.2. *Klebsiella pneumoniae* ja *Escherichia coli* resistentsus antibiootikumidele

Tulenevalt oma laialdasest levikust ja mitmeravimiresistentsusest (MDR) on *Klebsiella pneumoniae* sobiv näide, millega käsitleda antimikroobse resistentsuse mehhanisme. See bakter on resistentne peaaegu kõigile β-laktaam antibiootikumidele, suutes neid neutraliseerida (Li et al., 2023). Üheks oluliseks mehhanismiks on blaOXA-1 geeni amplifikatsioon, mis suudab koos välismembraani poriinide (nt OmpK36) inaktiveerumisega viia resistentsuseni eri β-laktaam antibiootikumide vastu (Shropshire et al., 2021). Teine oluline mitmeravimiresistentsuse mehhanism on efflukspumpade modifitseerimine, mis võimaldavad tavaolukorras antibiootikumil bakterisse tungida (Hussein et al., 2024). AcrAB-TolC ja MdtK on efflukspumbad, mille kaudu on

kujunenud resistentsus kinoloonidele ja tetratsükliinidele (Swedan ja Aldakhily, 2024). OmpK-35 ja OmpK-36 on poriinid, mille kaudu pääsevad rakku tsefalosporiinid ja karbapeneemid, kuid nende valkude modifitseerimine aitab kaasa resistentsuse tekkele *K. pneumoniae*'s (Elías-López et al 2024).

Escherichia coli on teine näide mitmeravimiresistentsest bakterist, millesse nakatumise suremuseprotsent on arengumaades endiselt kõrge (Kosek et al., 2003; Morita ja Kuroda, 2025; Qadri et al., 2005). Tavaolukordades ei pea *E. coli* põhjustatud düsenteeriat ravima antibiootikumidega, kuid on soovitatav kasutada äärmuslikel juhtudel. Kui trimetoprimmi ja sulfametoksasooli sai kasutada enterotoksigeense *E. coli* puhul paarkümmend aastat tagasi, siis tänapäeval ei mõju antud kombinatsioon teatud piirkondades tõhusalt (Ericsson, 2003; Morita ja Kuroda, 2025). Enim kasutatavamad antibiootikumiklassid on makroliidid (erütromütsiin, asitromütsiin), tetratsükliinid (doksütsükliin) ja fluorokinoloonid (tsiprofloksatsiin, norfloksatsiin), millest viimase klassi antibiootikume kasutatakse esmase ravivalikuna (Ericsson, 2003; Morita ja Kuroda, 2025; Qadri et al., 2005).

1.3.3. AMR tuvastamine WGS meetodil

WGS ehk terve genoomi sekveneerimine on organismi genoomi kõigi nukleotiidide järjendamine, mille abil on võimalik antibiootikumide resistentsust määrata juba geenitasandil (CDC, 2024; Illumina, 2024). WGS-i kasutatakse tänapäeval laialdaselt, kuna genoomi DNA annab pea kõikehõlmava ülevaate organismist (Hendriksen et al., 2019). Antud meetodi kasutamise eelis on võimalus tuvastada väga mitut AMR geeni või mutatsiooni samaaegselt, kuid terve genoomi sekveneerimisel on isolaatide uurimisel probleemiks selle ajakulu, kus esmane bakterite kasvatamine võib osutuda ajamahukaks, et saada analüüsiks piisav kogus materjali (Anjum et al., 2017; Hendriksen et al., 2019). Ühtlasi peab bakteritelt eraldatav DNA olema saastevaba, et suurendada uuritava geneetilise informatsiooni usaldusväärsust (Anjum et al., 2017).

Viimase aja genoomiuuringud on esile toonud selle, et WGS-iga tekib suurel hulgal andmeid võrreldes traditsioonilise Sangeri sekveneerimistulemustega (Anjum et al., 2017; Moeckel et al., 2024). Teise põlvkonna sekveneerimismeetodid (NGS) toodavad palju lühikesi DNA lõike, mis on sageli resistentsusgeenidest lühemad. Tervikliku geeni kokkupanemiseks on tarvis järjestuste ülekatmist pikemate lugemitega (Anjum et al., 2017). Sekveneerimise tulemusena saadud ulatuslik andmehulk on sageli probleemiks, kuna vähesed lugemid on otseselt seotud huvipakkuva

fenotüübiga. Samas võivad teatud resistentsuse tunnused hõlmata suuremaid genoomi piirkondi kui ainult üksik geen, sestap on hädavajalik kasutada suuri andmekogusid (Břinda et al., 2020).

Andmete analüüsimiseks on vaja algoritme ning üheks võimalikuks analüüsimeetodiks on k-meeride sageduste uurimine, kus k-meerid on kindla pikkusega järjestikused nukleotiidi- või aminohappejärjestused. Selle meetodi kasutusala on viimastel aastatel suurenenud, kuna arvutusprotsessid võivad selle abil olla ajaliselt vähem nõudlikud kui mõne muu meetodiga (Moeckel et al., 2024). Siiski tuleb meeles pidada, et mida pikem on k-meeri pikkus, seda keerulisem ja mahukam on arvutustöö (Moeckel et al., 2024). K-meeride analüüsi abil on võimalik resistentsust tuvastada ning selle jaoks on ühe näitena välja töötatud platvorm Kover, mis põhineb genoomis olevate biomarkerite ülesotsimisel (Drouin et al., 2016a).

1.4. Ülevaade bioinformaatilistest tööriistadest AMR-i ennustamiseks ja tuvastamiseks

Antimikroobse resistentsuse tuvastamiseks on laboratoorsete meetodite kõrval võimalik kasutada ka bioinformaatilisi lähenemisi (Seoane et al., 2020). Kui laboris kasutatavad antibiogrammid ja MIC-id on enamasti täpsed, siis nende puuduseks on bakterikultuuride kasvatamine (Drouin et al., 2016b). Antibiogrammid annavad ülevaate mikroorganismist ja tema tundlikkusest eri antibiootikumide suhtes, mille andmed võib koguda haiglatest ja profiili koostatakse edasi laboris (Pretzel, 2015). Peale selle võib aja jooksul kindel fenotüüp välja kujundada resistentsuse kindlale antibiootikumile. Selleks on genoomsed meetodid paremad (Drouin et al., 2016b).

Lisades antimikroobse tundlikkuse testimisele (AST) juurde ka WGS-i, võib uuritav tulemus tulla kiirem ja täpsem. WGS-i puhul on puuduseks sekveneerimisest tulenev suurenenud arvutusressursside kasutus, kuid uuemad resistentsuse analüüsi tarkvarad on selle tarbeks optimeeritud (Seoane et al., 2020). Teisalt on sekveneerimisest tulevad andmed korduvkasutatavad, kui avastatakse veel uurimata jäänud AMR geene (Bortolaia et al., 2020).

Antimikroobse resistentsuse ennustamiseks genoomsetest andmetest on võimalik kasutada masinõppe meetodeid. Masinõppe algoritmid peaksid suutma tuvastada ka varem tundmatuid geenimustreid, kuid see eeldab tõhusaks töötamiseks suurte treeningandmestike olemasolu – mida rohkem on andmeid, seda tõhusamalt suudab algoritm ennast arendada (Coursera, 2025; Pesesky et al., 2016). Näidisandmetega treenimisel peaks masin jõudma välja liigitussüsteemini, mida saaks

kasutada edaspidises töös tõese ja väära tulemuse määramiseks. Üldisem eesmärk on õpetada masin andmeid analüüsima, et jõuda usaldusäärsete tulemusteni (Larrañaga et al., 2006).

Teine võimalus on kasutada reeglipõhist lähenemist, kus AMR tuvastatakse uuritava proovi ja viitekataloogi võrdlusel. Siinkohal on oluline viitekataloogi kvaliteet, kuna vastasel juhul saadakse tulemuseks valed AMR põhjustavad geenid või geenivariandid (Hu et al., 2024). AMR-i kontekstis peaksid reeglipõhised meetodid olema kasutatavad nii väikeste kui ka suurte andmestikega ning vigu peaks olema üsna lihtne tuvastada (Pesesky et al., 2016).

Alternatiivina on pakutud välja lähenemine, kus resistentsuse ennustamisel ei tugineta teadaolevatele AMR-geenidele, vaid kasutatakse konserveerunud geenide alamhulka. Sellise lähenemisega on võimalik saada arvestatav tulemus ka mittetäielike genoomide korral. Nguyen ja kolleegide arvates oleks vaja teha edasiarendusi metagenoomsete kontiigide tuvastamisel, et ennustustulemusi veel parandada (Nguyen et al., 2020).

Antimikroobse resistentsuse määramisel on oluline veenduda, et sekveneeritud andmed pärinevad õigelt mikroorganismilt. Selleks kasutatakse taksonoomilise liigitamise tööriistu, millest levinuim on Kraken. Tegemist on bioinformaatilise tööriistaga, mis on loodud metagenoomsete DNA järjestuste kiireks ja täpseks taksonoomiliseks liigitamiseks. Varasemad lahendused, nagu BLAST, pakkusid kõrget täpsust, kuid seda kiiruse arvelt. Kraken ühendab täpsuse ja töökiiruse, et võimaldada efektiivset analüüsi ka suurte andmehulkade juures (Wood ja Salzberg, 2014).

AMR-i tuvastamine on tänu bioinformaatikale muutunud kiiremaks ja täpsemaks, seda tänu masinõppele ja reeglipõhistele lähenemistele. Sealhulgas on oluline koht ka traditsioonilistel AMR-i tuvastamise meetoditel. Programmid nimedega Kover, ResFinder ja RASE mängivad olulist rolli, et tuvastada antimikroobset resistentsust eri sorti andmetest (Břinda et al., 2020; Drouin et al., 2016b; Zankari et al., 2012).

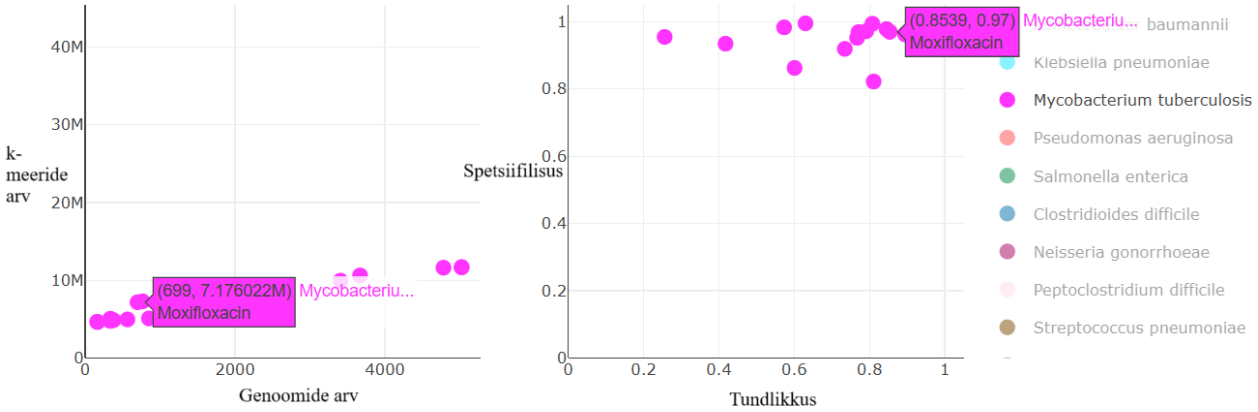
1.4.1. Kover

Kover platvorm on loodud selleks, et ennustada AMR-i ja leida seosed uuritava proovi ja antibiootikumi resistentsuse vahel. Resistentsuse analüüsi tulemustele lisatakse ka mudelite visuaalne ülevaade, et paremini leide mõista. Kover kasutab andmebaasina patosüsteemide integreerimiskeskust (PATRIC), kuhu on tänaseks kogutud üle 200 000 bakterigenoomi (BV-BRC, 2025). Koveri üheks eesmärgiks oli luua hõre (*sparse*) mudel, mis suudaks võimalikult väheste andmepunktidega fenotüüpi määrata kuluefektiivsemalt (Drouin et al., 2016a).

Koveri platvorm kasutab genoomseid mustreid tuvastavat algoritmi (SCM), millega saab luua reeglipõhiseid mudeleid, mis kasutavad „jah/ei“ vormis reegleid (Drouin et al., 2016b). Algoritmile on vaja sisendit tulemuse loomiseks ning k-meerid ja SCM seotakse omavahel kokku. Otsitava k-meeri olemasolul annab algoritm väljundiks „jah“ ja vastupidi. Kover kasutab k-meere, et muuta bakterite genoom masinõppe algoritmi jaoks sobivamaks ja nende kasutamise eelis seisneb genoomi struktuuris, kus paljud k-meerid võivad üheaegselt esineda ja toetab sarnasuste leidmist (Drouin et al., 2016a).

SCM algoritmi saab kasutada suurte andmemahtude juures, kuna selle tööefektiivsus kasvab lineaarselt genoomide (<1000) ja k-meeride arvuga (<1 000 000) (Drouin et al., 2016a ja 2016b). Vähese mälumahuga suudab algoritm õppida andmebaasis k-meere ära tundma vahemikus 33 sekundit kuni mõni tund (Drouin et al., 2016a). Seda üsna täpselt, kuid mõnikord tundlikkuse arvelt. Algses artiklis leiti, et 36 andmestikust ületas SCM 80% spetsiifilisuse läve 33 erinevas andmestikus, kuid ülejäänud kolmes mitte. See tähendab, et valepositiivseid tulemusi esines harva ehk tundlikke baktereid ei määratud ekslikult resistentseks. Tundlikkuse osas saavutati 36 andmestikust vähemalt 80% tulemus 25 andmestikus ning üheksas ülejäänus oli spetsiifilisus madalam, mis viitab, et mudel jättis osa tõeliselt resistentseid tüvesid tuvastamata. See näitab, et SCM kaldub pigem kõrge spetsiifilisuse kui kõrge tundlikkuse suunas (Drouin et al., 2016b).

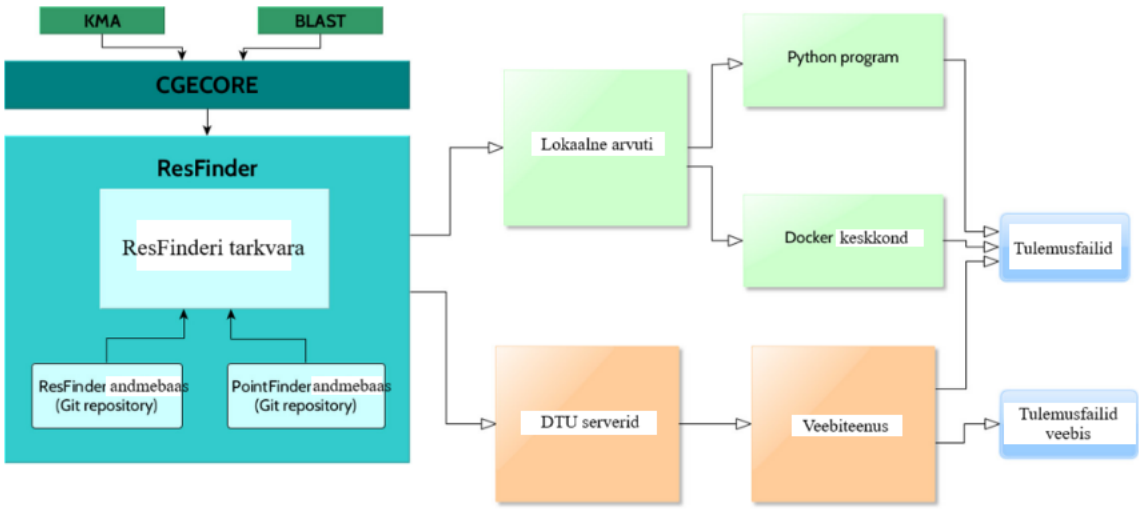
Koveri platvorm pakub põhjalikku SCM algoritmi tulemuste visualisatsiooni. SCM-i algoritmiga treenitud mudel kasutab k-meere genoomide esindamiseks, mis on osutunud kasulikuks AMR-i mehhanismide tuvastamisel (joonis 1), kus ennustatakse moksiofloksantiini resistentsust *Mycobacterium tuberculosis*’es 96% täpsusega (Drouin et al., 2016b). Leiti, et ühes uuritavas k-meeris puudus DNA güraasi A subühikus (gyrA) 6 aminohapet, mis annab bakterile fluorokinoloonide resistentsuse. Platvormil leidub ka teisi antibiootikume, näiteks amoksitsilliin, tsükloseriin ja kanamütsiin. (Chien et al., 2016; Drouin et al., 2016b)



Joonis 1. Kover-AMR-platvormi visualisatsioon *Mycobacterium tuberculosis*'e ja moksiofloksantiini näitel. Vasakpoolsel graafikul on näidatud genoomide arv (699) ja nende genoomide andmed SCM-is, mida esindavad k-meerid (7,17 miljonit). Parempoolsel graafikul on näidatud mudeli tundlikkust (0,8539) ja spetsiifilisust (0,97). (Kohandatud Kover – AMR Platform, 2017 järgi)

1.4.2. ResFinder

ResFinder on bioinformaatiline tööriist, millega saab tuvastada antimikroobseid resistentsusgeene (ARG) ja seda kasutajasõbralikult alates 2012. aastast (Zankari et al., 2012). Tööriista saab kasutada nii veebipõhiselt kui ka lokaalses arvutis (joonis 2). Nimelt on ResFinderit hea kasutada suurte andmemahtude analüüsimisel ja isolaatide resistentsuse määramisel (Seoane et al., 2020). ARG-ide tuvastamisele suudab ResFinder genereerida ka arvutipõhiseid antibiogramme, mis aitavad resistentsusprofiile paremini tõlgendada (Bortolaia et al., 2020). 2021. aastal tehti ResFinderiga üle 800 000 tuhande otsingu, mis näitab programmi suurt populaarsust ja nõudlust (Florensa et al., 2022).



Joonis 2. ResFinder platvormi töövoog. ResFinder koosneb kolmest Git versioonihaldussüsteemist (*Git repository*), sisaldades ResFinderi tarkvara, ResFinderi ja PointFinderi andmebaasi. Genoomse epidemioloogia keskuse CORE (CGECORE) on tarkvara keskne komponent, mis integreerib k-meeri joondamise (KMA) ja homoloogiaotsingu (BLAST) algoritmid. Tarkvara allalaadimisel on võimalik ResFinderit jooksutada lokaalses arvutis ja isoleeritud Docker keskkonnas. DTU serverid hoiustavad veebiteenuse keskkonda. Mõlema variandi jooksutamisel tekivad tulemusfailid, mida edasi analüüsida. (Kohandatud Florensa et al., 2022 järgi)

ResFinder on aastate jooksul täienenud. Esialgses ResFinderi versioonis keskenduti peamiselt horisontaalsele resistentsusgeeni ülekandele, kuid uuendatud programmis kasutatakse ka mutatsiooni põhiseid resistentsusgeene (PointFinder). Erisusi on veelgi, kusjuures varasemalt tugines ResFinder homoloogiaotsingu tarkvarale (BLAST), siis 2020. aasta versioonis muudeti ResFinderit enamgi (Bortolaia et al., 2020; Zankari et al., 2012). Varasem programmi versioon oli kirjutatud Perlis, kuid alates ResFinder 3.0 kasutati koodi kirjutamiseks Pythonit (Florensa et al., 2022). Täpsemalt tehti muudatusi programmi koodis, et võimaldada sekveneerimise toorandmete kasutamist. Sekveneerimise toorandmete kasutust kontrolliti ca 14 000 vaatlusega, kus vaid 0,2% juhtudest ilmnesid erisused sellest, kas kasutada FASTQ või FASTA formaate (Bortolaia et al., 2020).

Varasemates versioonides kasutati *de novo* assambleerimist programmiga Velvet ja hiljem SPAdes, kuid mõlemad olid sekveneerimisvigadele vastuvõtlikud. Probleem esines siis, kui analüüsitavad ARG-id moodustasid vähem kui 1% lugemitest. Takistuse ületamiseks töötati välja k-meeride

joondamine (KMA), mis tugineb ConClave algoritmil, et joondada toorjärjestusi otse ResFinderi andmebaasi järjestustele ilma *de novo* assambleerimiseta (Florensa et al., 2022).

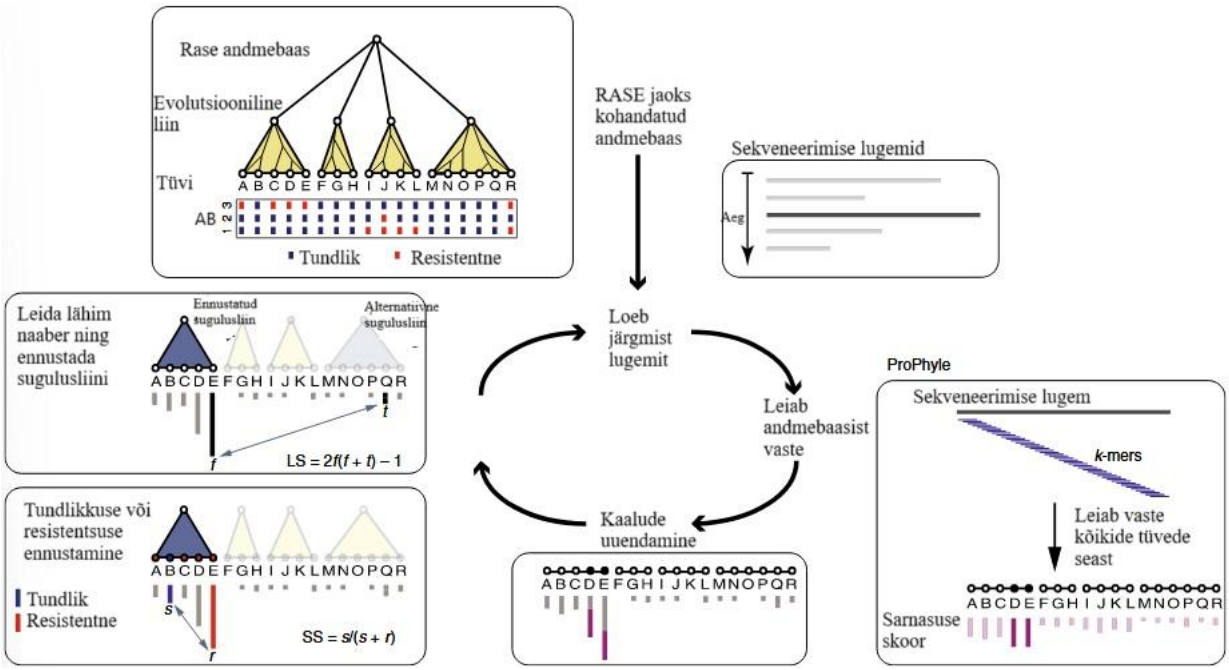
KMA loodi ResFinderi jaoks, et lahendada NGS-i andmetest tuleva redundantse andmebaasi probleemi (Florensa et al., 2022). Redundantsus (*redundancy*) ehk kõdusus viitab väga paljudele sarnastele järjestustele või korduvatele andmetele andmebaasis, mis võivad arvutustööd märkimisväärselt raskendada (Sikic ja Carugo, 2010). Toorandmete joondamine redundantse andmebaasi vastu osutub keeruliseks, kuna puudub garantii, et uuritav lugem kataks referentsi ainulaadse osa. Selle tulemusena võivad mitmed vasted olla võrdselt sobivad, mis raskendab edasist joondamist. Kuid KMA suudab erinevalt teistest joondamismeetoditest (BowTie2, BWA-MEM) kiirelt ja täpselt lugemeid kaardistada, mis annab meetodile eelise suurte andmebaaside ees (Clausen et al., 2018).

AMR-i ennustusprogrammide võrdluses uuriti WGS-põhiseid antibiootikumi resistentsuse ennustustööriistu, kus ResFinder 4.1 osutus üldiselt täpsemaks kui Kover-AMR patogeensete *E. coli* tüvede puhul (Verschuuren et al., 2022). Täpsus varieerus olenevalt kasutatavast antibiootikumist, kus piperatsilliini ja tasobaktaami kombinatsiooni puhul oli ResFinderi ennustuste kooskõla laboratoorsete tulemustega ligikaudu 80%, mis on madalam võrreldes gentamütsiini ja tsiprofloksatsiini korral täheldatud 96-98% täpsusega (Verschuuren et al., 2022). Madalama täpsuse põhjuseks olid nii valepositiivsed tulemused, kus tundlikud tüved ennustati resistentseteks ja valenegatiivsetest, kus resistentsed tüved ennustati tundlikeks (PAHO, 2025).

ResFinder on suure osakaaluga bioinformaatika huviliste ning teadlaste elus. Tänu mitmetele eelmainitud uuendustele on tarkvara nii veebipõhiselt kui ka lokaalselt muutunud efektiivsemaks ning täpsemaks. Tegu on ühe juhtiva tööriistaga genoomipõhise resistentsuse uurimisel.

1.4.3. Resistance-associated sequence elements (RASE)

Genoomipõhine naabertüpiseerimise (*genomic neighbour typing*) meetodiga on võimalik tuvastada antimikroobset resistentsust. Antud meetodi rakendamiseks on tarvis RASE tarkvara, mis põhineb resistentsusega seotud järjestuselementides (joonis 3). Patogeeni proovi sekveneeritakse reaalajas ning saadud järjestusi võrreldakse genoomide andmebaasiga. „*Genomic neighbour*“ tuleb sellest, et tuvastada kõige lähedasem sugulane uuritavale proovile. See meetod võimaldab ennustada patogeeni tüve minutitega ning on väga perspektiivikas ka molekulaarse epidemioloogia arendamises (Břinda et al., 2020).



Joonis 3. RASE töövoog. Esimesena laetakse RASE andmebaas mällu. Sekveneerimise lugemeid võrreldakse modifitseeritud ProPhyle algoritmi abil k-meeride andmebaasiga. Edasi arvutatakse k-meeride vastavusprotsent, mida uuendatakse kaalude kumulatiivse skooriga. Tundlikkuse skoor (SS) aitab ennustada resistentsust ja tundlikkust. Sugulusliini (LS) skoor aitab ennustada sugulusliini. Väljundina esitatakse tulemusfailid. (Kohandatud Břinda et al., 2020 järgi)

RASE programmi tundlikkuse ja resistentsuse ennustamine põhineb skooridel (joonis 3). Täpsemalt arvutatakse pärast k-meeride vastavusprotsenti kumulatiivne skoor, mis aitab leida sarnasemaid referentsliine ja tüvesid. Tundlikkuse ja resistentsuse ennustamisel võetakse iga uuritav antibiootikum eraldi arvesse, kus tuleb samuti skoore rakendada. Tundlikkuse skoor (SS) on vahemikus 0-1, kus 0 viitab loetud järjestuste resistentsusele ja 1,0 tundlikkusele. Sugulusliini skoori (LS) puhul näitab 0 nii ennustatud kui alternatiivse sugulusliini mõlemale sobivusele ja 1,0 viitab ainult ennustatud sugulusliinile. Nii tundlikkuse skoori (SS) kui ka sugulusliini skoori (LS) puhul on 0,6 kõrgemad tulemused kõrge usaldusväärsusega, kuna madalamad tulemused viitavad evolutsiooniliselt hiljem tekkinud tundlikkusele või resistentsusele, mida ei saa RASE usaldusväärselt eristada (Břinda et al., 2020).

Erinevalt teistest traditsioonilistest resistentsuse määramise meetoditest, mis kasutavad näiteks BLAST-i, ei otsi RASE otseselt resistentsust määravaid geene, vaid lähtub populatsioonistruktuuri fenotüüpilistest seostest. Kõrge riskiga tüvede tuvastamiseks tuginetakse bakterite geneetilisele sugulusele. Sellel eeldusel, et bakteri fenotüüp – antud juhul resistentsus – on tugevalt seotud

bakteri populatsioonistruktuuriga. See lähenemine on kasulik kliinilises kontekstis, kuna võimaldab resistentsust ennustada ka siis, kui resistentsusprofiil pole täielikult teada. Praegu määratakse patogeeni täpne sugulusliin alles pärast oluliste kliiniliste otsuste langetamist. Genoomipõhine naabertüpiseerimine võiks seda protsessi kiirendada, isegi kui DNA eraldamise protsess võtab neli tundi aega (Břinda et al., 2020).

RASE laialdane kasutus oleks epidemioloogia uuringutes kasulik. Fenotüüpiliste testide tegemine ja kultuuride kasvatamine võib võtta kaua aega, siis RASE tarkvaraga reaalajas sekveneerimine aitaks arstidel kiiremini langetada otsuseid. Sealhulgas oleks võimalik uurida ka resistentsuse levikut haiglasiseselt või haiglatevaheliselt (Břinda et al., 2020).

Teisalt rõhutavad Břinda ja kolleegid, et RASE-l on mitmeid piiranguid, millest üks on andmebaasi mitmekesisus. Usaldusväärse ennustuse jaoks on tarvis andmebaasi, mis sisaldab erinevaid bakteriaalseid tüved, millega võib kohalikus haiglas kokku puutuda. Muuseas mõjutab täpsust ka RASE meetodi omapära, mis tugineb populatsioonistruktuuridele. Kui antibiootikumi resistentsusgeen kandub üle horisontaalse geeniülekandega, siis meetodi täpsus võib veelgi väheneda. See võib viia vale sugulusliini määramiseni. Kui nakkustekitaja pole teada või proovis esineb mitme erineva mikroorganismi kombinatsioon, võivad k-meerid kattuda RASE andmebaasis eri patogeenidega. See on oluline, kuna peab kasutama õiget andmebaasi, et elimineerida eksitavate tulemuste teke (Břinda et al., 2020).

Carroll ja kolleegide poolt tehtud uuringus leiti, et RASE antimikroobse resistentsuse ennustamisel *K. pneumoniae* puhul olid kõrge sugulusliini (LS) skooriga tulemused mitme antibiootikumi osas kliiniliselt usaldusväärsed, sealhulgas gentamütsiin, piperatsilliin ja tasobaktaam ning tseftriaksoon. Eelmainitud antibiootikumide tundlikuks ennustamised olid kooskõlas laboratoorsete tulemustega 90% juhtudest. Meropeneemi puhul näitasid laboratoorsed tulemused 100% tundlikkust, kuna võrdlusandmebaasis ei olnud ühtegi resistentset tüve. Seega ei saanud hinnata, kui täpselt RASE suudab tuvastada meropeneemiresistentsust (Carroll et al., 2025)

RASE põhineb populatsioonistruktuuri seoste otsimisel, et ennustada antimikroobset resistentsust. Antud lähenemine erineb teistest AMR-i meetoditest, pakkudes kiiremat ja potentsiaalselt laiemalt kasutatavat alternatiivi. Kuid lähenemisel on ka eelmainitud piiranguid, mis viitavad vajadusele RASE tarkvara edasiarendusele.

2. EKSPERIMENTAALOSA

2.1. Töö eesmärgid

Töö põhieesmärk oli võrrelda RASE ja ResFinderi AMR tuvastamisprogrammide tööefektiivsust. Täpsemad eesmärgid olid:

1) Selgitada välja, kuidas erinevad RASE ja ResFinderi tulemused kasutades sama uuritavat andmestikku.
2) Analüüsida, kuidas suudavad mõlemad programmid tuvastada resistentsust olukorras, kus uuritavale proovile on lisatud simuleeritud inimese vere metagenoomne taust.
3) Uurida, kui tundlikud on RASE ja ResFinderi programmid erinevate simuleeritud lahjendusastmete puhul ning millise lahjenduse juures suudavad programmid ennustada AMR-i usaldusväärselt.

Praktilises töös ei kasutatud algselt plaanitud Koveri platvormi, kuna seda tarkvarapaketti ei õnnestunud tehniliste probleemide tõttu tööle saada.

2.2. Materjal ja metoodika

2.2.1. Tarkvara ja andmebaasid

Bioinformaatiliste tarkvaraplatvormidena kasutasin AMR-i tuvastamiseks RASE ja ResFinderit (versioon 4.0). Andmete esmaseks puhastamiseks ja struktureerimiseks kasutasin Microsoft Excelit. Andmete visualiseerimiseks sh jooniste tegemiseks kasutasin R-i (versioon 4.4.3).

Andmebaasidena kasutasin RASE platvormi puhul *Klebsiella pneumoniae* ja *Escherichia coli* isolaatide andmekogu, mis isoleeriti uropatogeensetest bakteritest. Antud andmebaasid olid üles ehitatud RASE tarkvara spetsifikatsioonidele (Carroll et al., 2025). Kasutatud andmebaasid laadisin alla veebist (Carroll et al., 2024a, 2024b).

RASE *Klebsiella pneumoniae* andmebaas sisaldas järgmisi antibiootikume, mille vastu resistentsust/tundlikkust tuvastati: fosfomütsiin (FOS), amikatsiin (AMK), amoksitsilliin (AMOX), tsefepiim (CFPM), tseftriaksoon (CTX), tseftasidiim (CAZ), tsiprofloksatsiin (CIP), kolistiin (COL), ertapeneem (ETP), gentamütsiin (GEN), meropeneem (MEM), piperatsilliin/tasobaktaam (PIPTAZ) ja tigetsükliin (TIG) (Carroll et al., 2024b). Antud RASE andmebaas sisaldas 1511 sekveneeritud genoomi koos teadaolevate resistentsusfenotüüpidega.

Selle põhjal tuvastati proovi geneetiliselt lähimad sugulased ja nende alusel omakorda resistentsus või tundlikkus (Břinda et al., 2020; Carroll et al., 2024b).

RASE *Escherichia coli* andmebaas sisaldas järgmisi antibiootikume: gentamütsiin (GENT), amoksitsilliin/klavulaanhape (AMCL), ampitsilliin (AMP), tsefasoliin (CEFAZ), tseftriaksoon (CEFTRIAX), tsiprofloksatsiin (CIPRO), meropeneem (MERO), nitrofurantoiin (NITRO), piperatsilliin/tasobaktaam (PIPTAZ), trimetoprim/sulfametoksasool (SXT). Võrreldes *K. pneumoniae* RASE andmebaasiga, sisaldas *E. coli* andmebaas 148 genoomi, mille põhjal ennustati resistentsust ja tundlikkust (Carroll et al., 2024a).

ResFinderi tööks vajalikud andmebaasid (ResFinder, PointFinder, DisinFinder) laadisin eraldi alla ametlikest Git versioonihaldussüsteemidest (*Git repository*) (Bortolaia et al., 2020). ResFinderi andmebaas sisaldas resistentsusgeene, PointFinder sisaldas punktmutatsioone (ResFinder bitbucket, 2020). DisinFinder otsis resistentsust erinevatele desinfitseerimisvahenditele (ResFinder, 2025).

Praktilise töö vältel kasutasin töögrupi Linux serverit ja töövoo käivitasin käsurealt.

2.2.2. Töövoog

Esimese sammuna oli vaja saada nii RASE kui ka ResFinderi programmid tööle Linuxi keskkonnas. Selleks kasutasin Git versioonihaldussüsteeme (RASE, 2020; ResFinder bitbucket, 2020). Edasine töö toimus veebiliidesel bakteriaalse ja viiruste bioinformaatika andmekeskuse (BV-BRC) kodulehel, kust otsisin antibiootikumide resistentseid või tundlikke andmeid *K. pneumoniae* ja *E. coli* kohta. Täpsemalt leidsin kasutatud proovid AMR fenotüüpide alt. *K. pneumoniae* puhul otsisin tüvesid, mis oleksid BV-BRC andmetel 1) laboratoorselt kinnitatud; 2) kontrollitud kasutades *broth dilution* meetodit; 3) näitaksid antibiootikumi vastu, kas tundlikkust või resistentsust (BV-BRC, 2025). *E. coli* puhul otsisin tüvesid, mis oleksid sarnaselt *K. pneumoniae*'le laboratoorselt kinnitatud ning antibiootikumi vastu resistentsed või tundlikud. BV-BRC etapp oli oluline, et hiljem kontrollida, kui täpselt ennustasid RASE ja ResFinder resistentsust või tundlikkust antibiootikumidele.

Algsed proovide fenotüübid leidsin BV-BRC-st, lugemid laadisin serverisse seejärel Euroopa Nukleotiidiarhiivist (ENA), mis hoiustab sekveneerimisandmeid igast maailmanurgast (ENA, 2025). *K. pneumoniae* ja *E. coli* proovide sekveneerimine viidi läbi Illumina platvormil, kasutades

terve genoomi sekveneerimist. Sekveneerimisandmed olid esitatud paarislugemitena ning lugemipaaride arv varieerus 800 000 aluspaarist kuni 4 miljoni aluspaarini.

Tulemuste usaldusväärsuse tagamiseks kaasasin analüüsi ainult need tüved, mille LS väärtus oli RASE tarkvara rakendamisel vähemalt 0,5. Antud piiri rakendasin, kuna LS näitab kui tugevalt kattub uuritava tüve järjestus parima andmebaasis leitud tüvega. LS=0,5 viitab sellele, et analüüsitud lugemid jagunesid võrdselt esimese ja teise vaste vahel (Břinda et al., 2020). Sellest tulenevalt rakendasin antud praktilises töös LS ≥ 0,5 lävendit, et tagada RASE tarkvara piisav eristusvõime.

RASE tarkvara LS ≥ 0,5 lävendi puhul sain *K. pneumoniae* 115 sekveneerimistulemuse analüüsist 25 usaldusväärset vastet ning *E. coli* 143 sekveneerimislugemist 21 usaldusväärset proovi (lisa 1 ja 2).

Otsitud proovidele oli tarvis teha simuleeritud lahjendused, et uurida, millisest lahjendusastmest tarkvara õigesti resistentsust või tundlikkust ennustas. Simuleeritud lahjenduste tegemiseks kasutasin inimese tausta-DNA proovi, mis sisaldas lisaks inimese DNA-le ka teiste bakterite geene, mida leidub tavaliselt inimese veres (lisa 3). Antud tausta-DNA fail valmis töögrupi teise liikme poolt. Tausta-DNA failiformaadiks oli sarnaselt uuritavate proovidega FASTQ. Simulatsiooni eesmärgiks oli jäljendada inimese verd, millest edaspidi saaks ennustada, kas inimest nakatanud patogeen on resistentne või tundlik antibiootikumile, sh kuidas valitud tarkvarad tulevad toime lisaks peamisele uuritavale tüvele ka inimese „verega“.

Lahjendused tegin valmis töögrupi skriptiga Linuxi keskkonnas käsureal, mille väljundfailides esitatakse tulemused kümnendastmetes (Bioinformaatika töögrupp, 2025). Simuleeritud lahjendused algasid 10^0 tasemelt ja ulatusid kuni 10^6-ni. 10^0 lahjendus sisaldas vaid ühe lugemirea uuritavast tüvest, samas kui 10^6 lahjenduses oli lisaks simuleeritud inimese verele ka 1 000 000 rida uuritavat tüve. Ühe uuritava tüve kohta sain seitse lahjendust, mis olid simuleeritud inimese verega kokku segatud. Täiendavalt viisin läbi kontrollkatse lahjendamata tüvega, mida ei lahjendatud inimese simuleeritud vere andmetega. Edasi toimus lahjendatud proovide analüüs RASE ja ResFinderi tarkvaradega.

2.2.3. RASE programmi töövoog

Simuleeritud vere ja uuritava tüve lahjendused koos lahjendamata tüvega lisasin Linuxis eraldi RASE keskkonda (*environment*), et alustada analüüsi. Tüvede analüüsiks kulus koos lahjendatud

tüvedega pea 3-4 päeva, viidates tarkvara põhjalikkusele. Tarkvara jooksutamisel analüüsisin sekveneerimisandmeid lugemi haaval, mis võis olla aeganõudev.

RASE väljundfailid jaotati kolmeks: sobivusfailid (*matching*), R-graafikud (*plots*) ja ennustustulemused (*prediction*). Praktilises töös kasutasin R-graafikuid ja ennustusi kahel eesmärgil: esmalt hinnata, kas programm ennustas AMR-i usaldusväärselt (LS≥ 0,5) ning teisalt lihtsustamaks tulemuste tõlgendamist. R-i graafikute failid jagunesid neljaks, millest kolm kajastasid erinevatel ajahetkedel ennustatud sugulusliine (LS) ning viimane andis ülevaate, milliste antibiootikumide suhtes on uuritav tüvi resistentne või tundlik (lisa 4). Kuigi RASE tarkvara on algselt loodud kolmanda põlvkonna sekveneerimisandmete jaoks, siis on see rakendatav ka NGS andmetele (Břinda et al., 2020; Carroll et al., 2025). Praktilises töös kasutasin ennustusfaili (.tsv) tulemuste interpretatsiooniks, keskendudes sekveneerimise viimase ajapunkti andmetele.

2.2.4. ResFinderi programmi töövoog

ResFinderi tarkvara kasutasin sarnaselt RASE tarkvaraga, kus proovide analüüsimiseks tegin eraldi keskkonna Linuxi käsureal ning analüüsisin samu proove, mis RASE-gagi. ResFinderi kasutamise omapäraks oli pidev andmebaaside uuendamine igal sisselogimisel, et võimalikult täpsed tulemused saada. ResFinderi rakendamisel kasutasin kõikide lahjenduste ja lahjendamata tüvede puhul samu parameetreid, mida oleks olnud võimalik vajadusel kohandada, kuid antud töö raames osutusid need sobivaks, millest tähtsamad olid pikkuse (-l) ja läve (-t) parameetrid.

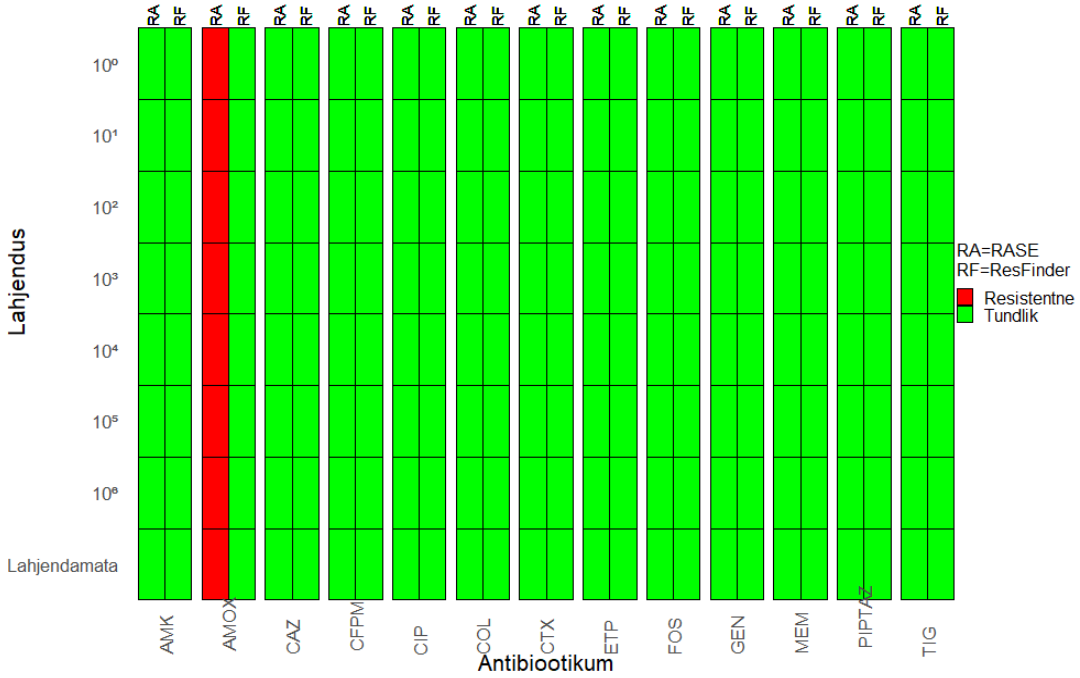
Käesolevas töös kasutasin l=0,6, mis tähendab, et geen loetakse tuvastatuks ainult siis, kui vähemalt 60% geenipikkusest on kaetud järjestusandmetega. Lävend -t määras, kui sarnane peab uuritav järjestus olema ResFinderi andmebaasis oleva resistentsusgeeniga, et see vastavaks loetaks. Antud juhul oli -t väärtuseks 0,8, mis tähendab, et tuvastatuks loeti need järjestused, mille sarnasus vähemalt teadaoleva resistentsusgeeniga oli vähemalt 80% (ResFinder bitbucket, 2020).

ResFinderi analüüsist tekkinud tulemusfailid jaotati kolmeks: tekstifailid (.txt), järjestusfailid (.fasta) ja masinloetavad failid (.json). Igaüks omas erinevat funktsiooni, millest antud töös analüüsisin ResFinderi andmebaasist saadud tekstifaile.

2.3. Tulemused

2.3.1. Simuleeritud tausta-DNA resistentsuse ja tundlikkuse ennustamine erinevatele antibiootikumidele RASE ja ResFinderi tarkvarade põhjal

Simuleeritud lahjenduse faili põhjal hindasin mõlema tarkvara ennustusi resistentsuse ja tundlikkuse osas. Antud faili analüüs mõlema tarkvaraga pidi kinnitama, et tegemist oli antibiootikumidele tundliku fenotüübiga. See oli oluline edasisteks analüüsideks, et uuritavate tüvede põhjal oleks võimalik usaldusväärselt ennustada nende resistentsust ja tundlikkust konkreetsete antibiootikumide suhtes. Simuleeritud taustaproov näitas RASE *K. pneumoniae* andmebaasis resistentsust amoksitsilliinile (joonis 4). *E. coli* andmebaasis leidis RASE resistentsust ampitsilliini suhtes (lisa 6). Edasistes analüüsides ma ei kasutanud eelmainitud antibiootikume.



Joonis 4. Simuleeritud inimese vere DNA proovi AMR-i ennustustulemused *Klebsiella pneumoniae* RASE andmebaasi ja ResFinderi andmebaaside põhjal. Graafiku x-teljel on erinevad antibiootikumid, mida leidus RASE andmebaasis. Y-teljel on lahjendamata proov, mis on ainult tausta-DNA ning edasi 10^0-10^6 on lahjendatud tausta-DNA proovi fenotüübid. RASE andmebaasis leidus 13 antibiootikumi, millest 12 olid tundlikud ja üks resistentne. ResFinderi andmebaasides ei leitud ühtegi resistentsusgeeni antud antibiootikumide suhtes.

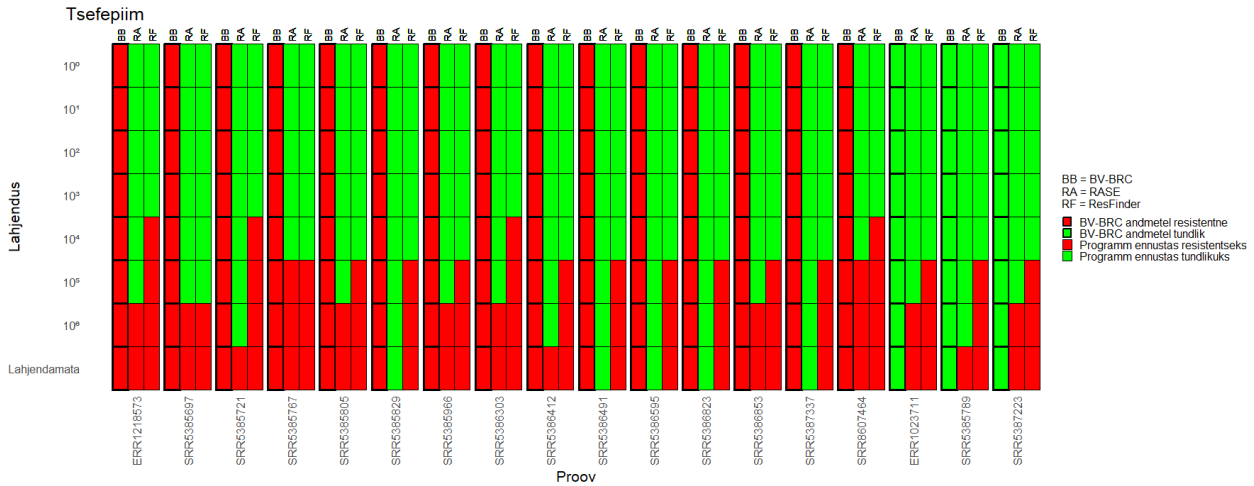
Simuleeritud inimese vere DNA proovi analüüsimisel andis RASE andmebaas *Klebsiella pneumoniae* ja *Escherichia coli* puhul LS skoorideks vastavalt 0,042 ja 0,186 (lisa 5). Tegu oli ootuspärase tulemusega, kuna tegemist polnud puhta isolaadiga, vaid keerukama metagenoomse taustaga. RASE tarkvaras kasutatav LS skoor viitab kui tugevalt vastab uuritav proov mõne andmebaasis oleva tüvega (Břinda et al., 2020). Eesmärk oli leida metagenoomse tausta seest kindel uuritava tüve liik, siis isolaadi tüve LS skoor peab olema võimalikult kõrge, kuid teiste liikide LS skoor võimalikult madal. See näitas, et simuleeritud tausta-DNA proov ei sarnane oluliselt ühegi resistentsusprofiiliga RASE andmebaasis.

2.3.2. RASE ja ResFinderi tarkvarade AMR-i ennustused *Klebsiella pneumoniae* näitel

Klebsiella pneumoniae antibiootikumi resistentsuse tulemuste põhjal valisin tulemuste tõlgendamiseks välja järgmised antibiootikumid: tsefepiim, gentamütsiin, meropeneem ning piperatsilliin ja tasobaktaam. Valik tugineb kirjanduse ülevaatele, mille kohaselt on *K. pneumoniae* sageli resistentne β-laktaamidele sh meropeneemile (karbapeneemide klass) ja tsefepiimile (tsefalopsoriinide klass) (Li et al., 2023; DrugBank, 2025). Gentamütsiin lisati, kuna see kuulub aminoglükosiidide klassi ning piperatsilliin koos tasobaktaamiga esindab β-laktaami ja β-laktamaasi inhibiitori kombinatsiooni, uurides kuidas suudavad nii RASE kui ka ResFinder tuvastada kombinatsioonantibiootikume (Gin et al., 2007; DrugBank, 2025).

Esimesena analüüsisin tsefepiimi suhtes resistentsust ja tundlikkust kasutades RASE ja ResFinderi tarkvara (joonis 5). Resistentseid proove oli 15 ja tundlikke proove kolm. *K. pneumoniae* resistentsuse hindamisel vaatlesin täpsust eelkõige lahjendusastmetel vahemikus lahjendamata kuni 10^4, kuna lahjemates proovides (10^0-10^3) ei tuvastanud kumbki RASE ega ResFinderi andmebaas usaldusväärselt resistentsusprofiile. Tsefepiimi kõikide proovide õigesti ennustatud tulemuste protsent oli resistentsuse puhul vastavalt RASE-l 17% ja ResFinderil 40% (joonis 5).

Lahjendamata proovide osas tuvastasid RASE ja ResFinder resistentsust täpsusega 67% (10/15) ja 100% (15/15). Tundlikkust ei suutnud kumbki programm õigesti tuvastada. Lahjendamata kuni 10^4 lahjendusastme piires oli RASE puhul resistentsuse täpsus 33% (20/60) ja ResFinderil 80% (48/60). Tundlikkuse täpsus oli vastavalt 58% (7/12) ja 50% (6/12). Kui uurisin ainult lahjendusastmeid, langes resistentsuse määramine RASE-l 10%-ni (10/105) ja ResFinderil 31%-ni (33/105). Tundlikkuse määramine oli lahjendusastetel kõrgem 90% (19/21) ja 71% (15/21), kuid kogu andmestiku põhjal langes vastavalt 79%-ni (19/24) ja 62%-ni (15/24) (joonis 5).



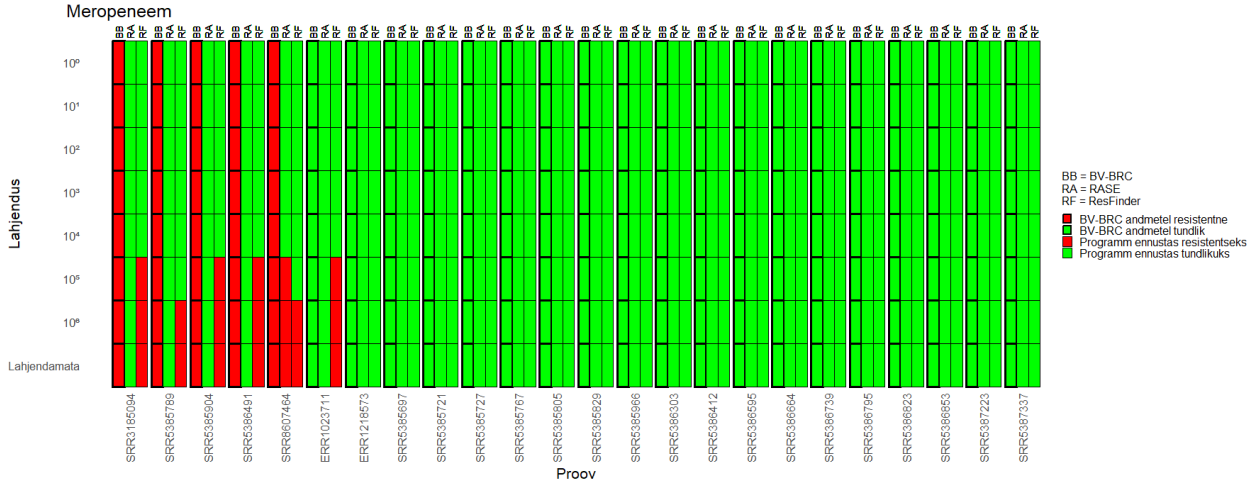
Joonis 5. Tsefepiimi resistentsusennustuste võrdlus BV-BRC (BB) andmetega *K. pneumoniae* näitel. Graafiku X-teljel on esitatud erinevad *K. pneumoniae* proovid koos nende nimetusega ENA andmebaasist. Y-teljel on lahjendusastmed: esimene tase tähistab lahjendamata proovi, sellele järgnevad lahjendused vahemikus 10^0-10^6 lahjendused, mis koosnevad proovi ja tausta-DNA segust. Iga proovi esimene tulp kajastab laboratoorselt kinnitatud resistentsust või tundlikkust (BB). Teine ja kolmas tulp näitavad vastavalt RASE (RA) ja ResFinderi (RF) ennustusi. Punase värviga on tähistatud resistentsus ja rohelisega tundlikkus.

Analüüsi teine antibiootikum oli gentamütsiin, mille algsest 25 proovist oli laboratoorselt kinnitatud 24 proovi (lisa 7). Kokku oli 14 resistentset ja 10 tundlikku proovi, mistõttu oli tundlikke proovide osakaal suurem kui tsefepiimi puhul. RASE-l kui ka ResFinderil oli kõigi proovide ennustamise osakaal resistentsuse määramisel vastavalt 7% ja 32% (lisa 7). Programmide ennustatud tundlikkus oli madalam võrreldes tsefepiimiga: RASE 53% ja ResFinder 56%.

Kolmanda antibiootikumina analüüsisin meropeneemi, mis kuulub sarnaselt tsefepiimiga β-laktaam antibiootikumide klassi, kuid tsefalosporiinide alamklassi. Käesolevas valimis oli resistentsete ja tundlikke proovide jaotus selgemalt eristuv (joonis 6). BV-BRC andmetel oli 24 proovist laboratoorselt resistentseks kinnitatud vaid viis ning ülejäänud 19 loeti tundlikeks. Õigesti ennustatud tulemuste osakaal oli tundlikkuse ennustamisel kõrge – kõikidest tulemustest ennustas RASE õigesti 100% ja ResFinder 98%. Resistentsuse tulemused olid vastavalt 8% ja 33% (joonis 6).

Lahjendamata proovide osas tuvastasid RASE ja ResFinder resistentsust täpsusega 20% (1/5) ja 100% (5/5). Tundlikkust tuvastati õigesti vastavalt 100% (19/19) ja 95% (18/19). Lahjendamata kuni 10^4 lahjendusastme piires oli RASE puhul resistentsuse täpsus 15% (3/20) ja ResFinderil 65% (13/20). Tundlikkuse täpsus oli vastavalt 100% (76/76) ja 98% (149/152). Kui uurisin ainult

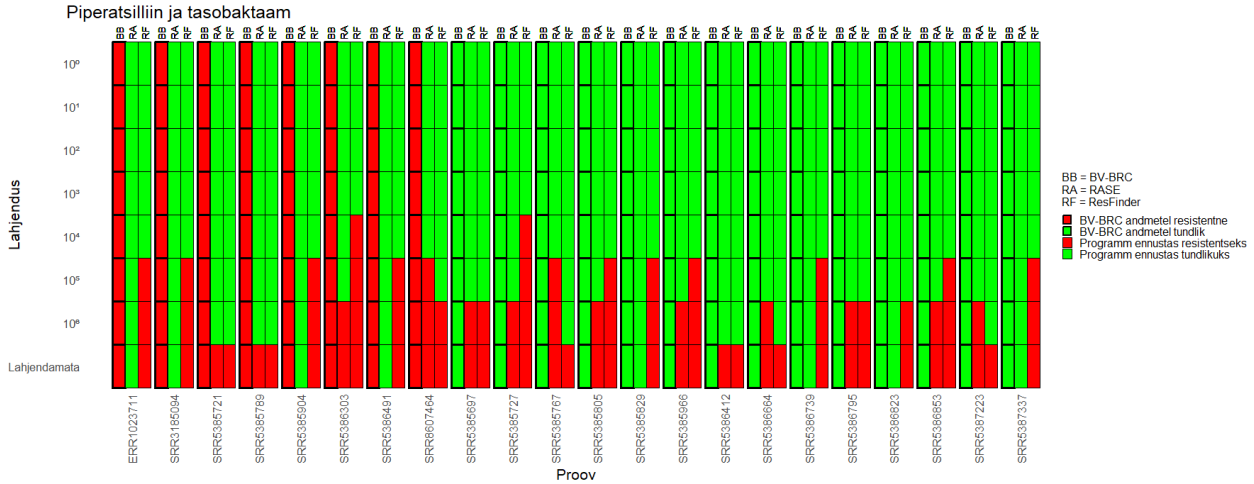
lahjendusastmeid, langes resistentsuse määramine RASE-l 4%-ni (1/28) ja ResFinderil 29%-ni (8/28), kuid tundlikkuse määr oli sama, mis lahjendamata kuni 10^4 lahjendusastme piires - 100% (133/133) ja 98% (131/133) (joonis 6).



Joonis 6. Meropeneemi resistentsusennustuste võrdlus BV-BRC (BB) andmetega *K. pneumoniae* näitel. Graafiku X-teljel on esitatud *K. pneumoniae* erinevad proovid koos nende nimetusega ENA andmebaasist. Y-teljel on lahjendusastmed: esimene tase tähistab lahjendamata proovi, sellele järgnevad lahjendused vahemikus 10^0-10^6 lahjendused, mis koosnevad proovi ja tausta-DNA segust. Iga proovi esimene tulp kajastab laboratoorselt kinnitatud resistentsust või tundlikkust (BB). Teine ja kolmas tulp näitavad vastavalt RASE (RA) ja ResFinderi (RF) ennustusi. Punase värviga on tähistatud resistentsus ja rohelisega tundlikkus.

Viimase antibiootikumina võrdlesin piperatsilliini ja tasobaktaami kombinatsiooni RASE ja ResFinderi tarkvaradega. Antud valimis leidus laboratoorselt kinnitatud proovidest kaheksa resistentset ja 14 tundlikku proovi. Kõikide proovide raames ennustas RASE õigesti resistentsust vaid 13% ja ResFinder 31% kõigist tulemustest (joonis 7). Tundlikkuse ennustamine oli parem, kus RASE ennustas õigesti 75% ja ResFinder 71% kõigist juhtudest. Tuleb tõdeda, et antud kombinatsiooni puhul ennustas ResFinder tundlikke proove enesekindlalt resistentseteks.

Lahjendamata proovide osas tuvastasid RASE ja ResFinder resistentsust täpsusega 5% (4/8) ja 100% (8/8). Tundlikkust tuvastati õigesti vastavalt 28% (4/14) ja 0% (0/14). Lahjendamata kuni 10^4 lahjendusastme piires oli RASE puhul resistentsuse täpsus 22% (7/32) ja ResFinderil 63% (20/32) ning tundlikkuse täpsus oli vastavalt 64% (36/56) ja 43% (24/56). Kui uurisin ainult lahjendusastmeid, langes resistentsuse määramine RASE-l 5%-ni (3/56) ja ResFinderil 21%-ni (12/56), kusjuures tundlikkuse määramine oli kõrgem 90% (88/98) ja 81% (80/98) (joonis 7).



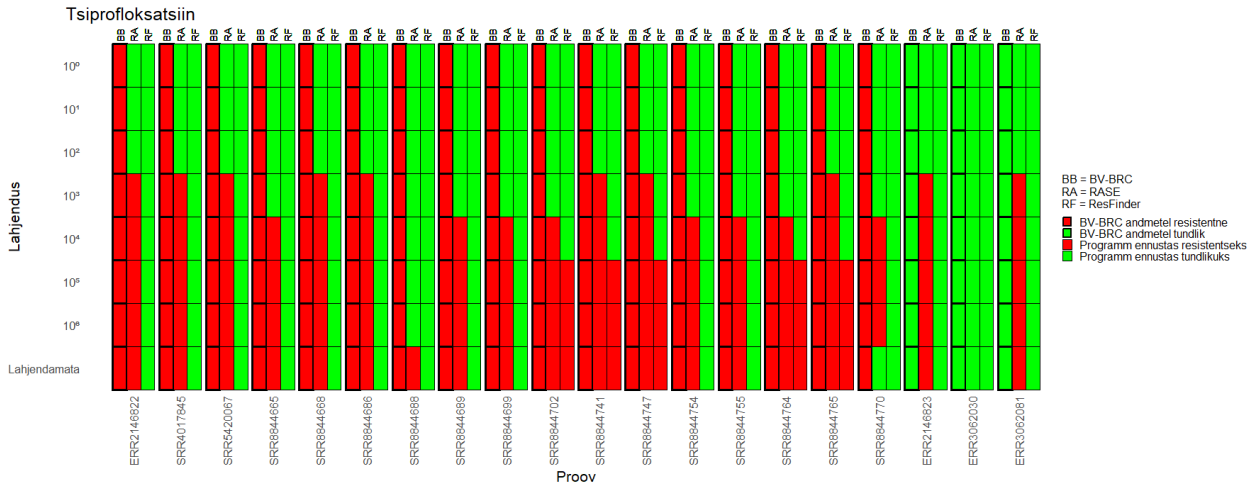
Joonis 7. Piperatsilliini ja tasobaktaami resistentsusennustuste võrdlus BV-BRC (BB) andmetega *K. pneumoniae* näitel. Graafiku X-teljel on esitatud erinevad *K. pneumoniae* proovid koos nende nimetusega ENA andmebaasist. Y-teljel on lahjendusastmed: esimene tase tähistab lahjendamata proovi, sellele järgnevad lahjendused vahemikus 10^0-10^6 lahjendused, mis koosnevad proovi ja tausta-DNA segust. Iga proovi esimene tulp kajastab laboratoorselt kinnitatud resistentsust või tundlikkust (BB). Teine ja kolmas tulp näitavad vastavalt RASE (RA) ja ResFinderi (RF) ennustusi. Punase värviga on tähistatud resistentsus ja rohelisega tundlikkus.

2.3.3. RASE ja ResFinderi tarkvarade AMR-i ennustused *Escherichia coli* näitel

Escherichia coli antibiootikumi resistentsuse tulemuste tõlgendamiseks valisin analüüsi järgmised neli antibiootikumi: trimetoprim ja sulfametoksasool, tseftriaksoon, gentamütsiin ja tsiprofloksatsiin. Need ravimid esindavad erinevaid antibiootikumide klasse ning valiku eesmärk oli hinnata, kuidas RASE ja ResFinder suudavad tuvastada resistentsust erinevate toimemehhanismidega antibiootikumide suhtes.

Kirjanduse põhjal on tsiprofloksatsiin tänapäeval üks tõhusamaid ravivõimalusi, millega ravida *E. coli* infektsioone (Ericsson, 2003; Morita ja Kuroda, 2025; Qadri et al., 2005). Trimetoprimmi ja sulfametoksasooli kombinatsiooni kasutati laialdaselt 20-30 aastat tagasi kriitilisematel juhtudel, kuid resistentsuse tase on nende suhtes paljudes piirkondades kõrge (Ericsson, 2003; Morita ja Kuroda, 2025). Tseftriaksooni ja gentamütsiini lisasin analüüsi, et esindada vastavalt β-laktaam- ja aminoglükosiidide klassi.

Esimesena analüüsisin tsiprofloksatsiini suhtes resistentsust ja tundlikkust kasutades RASE ja ResFinderi tarkvara (joonis 8). *E. coli* resistentsuse hindamisel vaatlesin täpsust eelkõige lahjendusastmetel vahemikus lahjendamata kuni 10^3 proovini, kuna lahjemates proovides (10^0-10^2) ei tuvastanud RASE ega ResFinderi andmebaas usaldusväärselt resistentsusprofiile (joonis 8).



Joonis 8. Tsiprofloksatsiini resistentsusennustuste võrdlus BV-BRC (BB) andmetega *E. coli* näitel. Graafiku X-teljel on esitatud erinevad *E. coli* proovid koos nende nimetusega ENA andmebaasist. Y-teljel on lahjendusastmed: esimene tase tähistab lahjendamata proovi, sellele järgnevad lahjendused vahemikus 10^0-10^6 lahjendused, mis koosnevad proovi ja tausta-DNA segust. Iga proovi esimene tulp kajastab laboratoorselt kinnitatud resistentsust või tundlikkust (BB). Teine ja kolmas tulp näitavad vastavalt RASE (RA) ja ResFinderi (RF) ennustusi. Punase värviga on tähistatud resistentsus ja rohelisega tundlikkus.

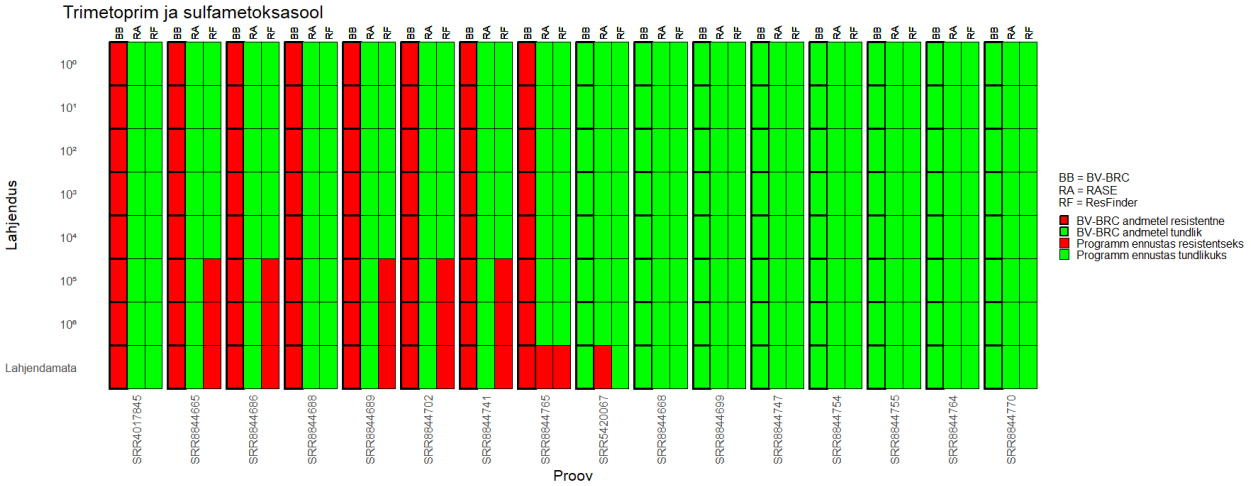
Tsiprofloksatsiini suhtes olid laboratoorselt kinnitatud andmete põhjal 17 proovi resistentsed ja vaid kolm tundlikku. Tähelepanuväärne oli RASE võimekus määrata resistentsust alates 10^3 lahjendusastmest, mis viitab antud juhul tarkvara potentsiaalile ka madalamatel bakteri kontsentratsioonidel resistentsust tuvastada. Kogu andmestiku (sh) kõigi lahjendusastmete põhjal tuvastas RASE õigesti 53% (72/136) ja ResFinder 12% (15/136) resistentsustulemustest. Tundlikkuse täpsus oli vastavalt 75% (18/24) ja 100% (24/24) (joonis 8).

Lahjendamata proovide osas tuvastasid RASE ja ResFinder resistentsust täpsusega 94% (16/17) ja 29% (5/17). Tundlikkust tuvastati õigesti 100%-lise (3/3) täpsusega. Lahjendamata kuni 10^3 lahjendusastme piires oli RASE puhul resistentsuse täpsus 85% (72/85) ja ResFinderil 18% (15/85), kuid tundlikkuse täpsus RASE-l langes 33%-ni (5/15), aga ResFinderil jäi samaks 100% (15/15). Kui uurisin ainult lahjendusastmeid, langes resistentsuse määramine RASE-l 49%-ni (57/119) ja ResFinderil 8%-ni (10/119), aga tundlikkuse määramine oli RASE-l kõrgem: 86% (18/21) ja ResFinderil 100% (21/21). (joonis 8).

Analüüsi teine antibiootikum oli trimetoprim ja sulfametoksasool, mille algsest 21 proovist oli laboratoorselt kinnitatud 16 proovi (joonis 9). Nendest kaheksa resistentset ja kaheksa tundlikku proovi, kus resistentsete ja tundlikke proovide arv oli tasakaalus. Nii RASE kui ka ResFinderi

kõikide tulemuste resistentsuse ennustused olid väga madalad vastavalt 2% ja 25% (joonis 9). Teisest küljest ennustati tundlikkust kõikide proovide raames 98% ja 100% täpsusega. Märkimisväärne oli tundlikkuse määramine 100% täpsusega ainult lahjendatud proovides.

Lahjendamata proovide osas tuvastasid RASE ja ResFinder resistentsust täpsusega 13% (1/8) ja 75% (6/8). Tundlikkust tuvastati õigesti vastavalt 88% (7/8) ja 100% (8/8). Lahjendamata kuni 10^3 lahjendusastme piires oli RASE puhul resistentsuse täpsus 3% (1/40) ja ResFinderil 40% (16/40). Tundlikkuse täpsus oli vastavalt 98% (39/40) ja 100% (40/40). Kui uurisin ainult lahjendusastmeid, siis langes resistentsuse määramine RASE-l 0%-ni (0/56) ja ResFinderil 18%-ni (10/56), kuid tundlikkuse määramine oli mõlemal programmil kõrgem - 100% (56/56) (joonis 9).

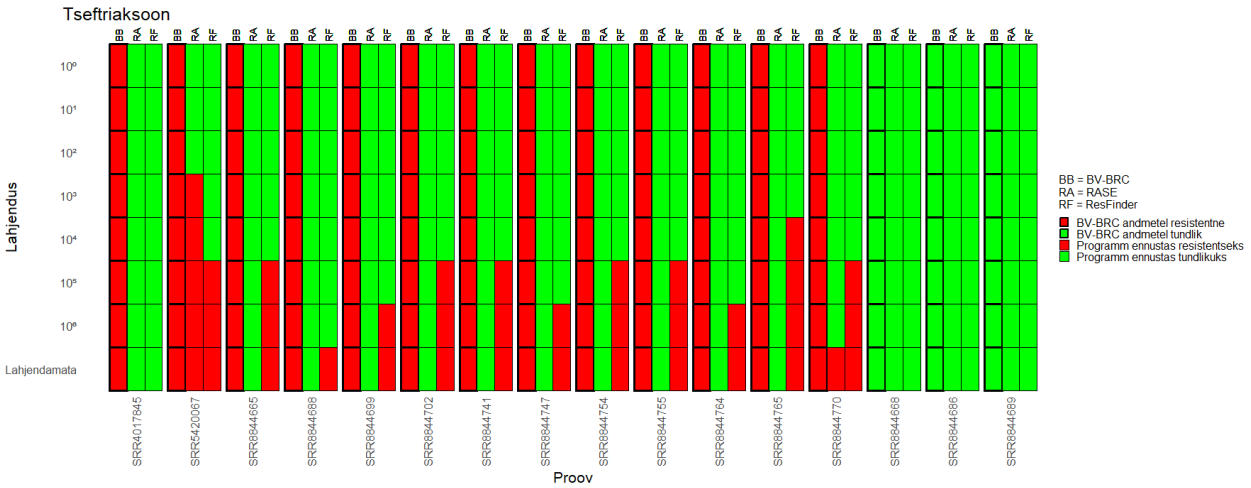


Joonis 9. Trimetoprim ja sulfametoksasooli resistentsusennustuste võrdlus BV-BRC (BB) andmetega *E. coli* näitel.. Graafiku X-teljel on esitatud erinevad *E. coli* proovid koos nende nimetusega ENA andmebaasist. Y-teljel on lahjendusastmed: esimene tase tähistab lahjendamata proovi, sellele järgnevad lahjendused vahemikus 10^0-10^6 lahjendused, mis koosnevad proovi ja tausta-DNA segust. Iga proovi esimene tulp kajastab laboratoorselt kinnitatud resistentsust või tundlikkust (BB). Teine ja kolmas tulp näitavad vastavalt RASE (RA) ja ResFinderi (RF) ennustusi. Punase värviga on tähistatud resistentsus ja rohelisega tundlikkus.

Kolmanda antibiootikumina analüüsisin gentamütsiini. BV-BRC andmetel oli 20 proovist kuus resistentsed ja 14 gentamütsiinile tundlikud. Nii RASE kui ka ResFinder ennustasid tundlikkust kõikide proovide raames pea 100% täpsusega (lisa 8). Resistentsust määrati vähem täpselt, kus kõikide tulemuste raames RASE ennustas õigesti 2% ja ResFinder 25% tulemustest. Tuleb tõdeda, et RASE ei tuvastanud ainult lahjenduste puhul resistentsust mitte üheski proovis, kuid ResFinder leidis resistentsusgeene 17% juhtudest (lisa 8).

Viimase antibiootikumina uurisin tseftriaksooni, mis kuulub β-laktaamide klassi. Antud valimis leidus 16 BV-BRC kinnitatud proovist 13 resistentset ja kolm tundlikku. Kõikide tulemuste lõikes tuvastasid RASE ja ResFinderi programmid resistentsuse õigesti vastavalt 6% ja 31% juhtudest, samas kui tundlikkust määrati mõlema programmi poolt õigesti 100% ulatuses (joonis 10).

Lahjendamata proovide osas tuvastasid RASE ja ResFinder resistentsust vastavalt 15% (2/13) ja 92% (12/13) täpsusega, kuid tundlikkust tuvastati 100%-lise täpsusega (3/3). Lahjendamata kuni 10^3 lahjendusastme piires oli RASE puhul resistentsuse täpsus 9% (6/65) ja ResFinderil 49% (32/65), aga tundlikkuse täpsus oli jällegi 100% (15/15). Kui uurisin ainult lahjendusastmeid, langes resistentsuse määramine RASE-l 4%-ni (4/91) ja ResFinderil 22%-ni (20/91) ning tundlikkuse määramine oli endiselt 100% (21/21) (joonis 10).



Joonis 10. Tseftriaksooni resistentsusennustuste võrdlus BV-BRC (BB) andmetega *E. coli* näitel. Graafiku X-teljel on esitatud erinevad *E. coli* proovid koos nende nimetusega ENA andmebaasist. Y-teljel on lahjendusastmed: esimene tase tähistab lahjendamata proovi, sellele järgnevad lahjendused vahemikus 10^0-10^6 lahjendused, mis koosnevad proovi ja tausta-DNA segust. Iga proovi esimene tulp kajastab laboratoorselt kinnitatud resistentsust või tundlikkust (BB). Teine ja kolmas tulp näitavad vastavalt RASE (RA) ja ResFinderi (RF) ennustusi. Punase värviga on tähistatud resistentsus ja rohelisega tundlikkus.

2.3.4. RASE ja ResFinderi töövoogude võrdlus ajakulu alusel

RASE tarkvara ajakulu varieerus sellest, kas analüüsida lahjendamata isolaadi tüvesid või lahjendatud simuleeritud metagenoomse taustaga proove. Lahjendamata proovide analüüsimisel oli ajakulu suurem, kus kõigi 25 *K. pneumoniae* proovi puhul kulus programmil pea 48 tundi. Simuleeritud metagenoomse taustaga proovide analüüsi ajakulu oli märgatavalt väiksem *K. pneumoniae* puhul, kus kõigi 175 (25 proovi × 7 lahjendust) lahjenduse analüüsi viisin läbi 13

tunniga. *E. coli* puhul täheldasin, et isolaatide analüüsimisel sai RASE veel kiiremini hakkama, kus 21 proovi töötlemiseks kulus pea kuus tundi ja 147 (21 proovi × 7 lahjendust) proovi analüüs võttis maksimaalselt neli tundi aega.

RASE tarkvara töökiirus sõltus oluliselt kasutatava andmebaasi mahust ning osaliselt ka lugemite arvust. *K. pneumoniae* andmebaas sisaldas 1511 genoomi ja *E. coli* oma 148 genoomi, mille vastu uuritavaid järjestusi analüüsiti. Lahjendamata proovide lugemipaaride keskmine oli *K. pneumoniae*'l 1,5 miljonit, aga *E. coli*'l piirdus 1,2 miljoniga. Lahjendatud proovide keskmine oli *K. pneumoniae*'l 470 000 lugemit, kui *E. coli*'l oli 425 000 lugemit. Tuleb tõdeda, et ajaline erinevus lahjendamata ja lahjendatud proovide vahel võis olla tingitud kõigi lahjendatud proovide identsest metagenoomsest taustast, mis võis vähendada arvutuskoormust. Seega mõjutasid nii andmebaasi suurus kui ka lugemite arv analüüsi ajakulu.

ResFinderi kasutamine oli ajaliselt oluliselt vähemnõudlik võrreldes RASE tarkvaraga. Analüüs toimus käsurealt, kus iga proov vajas minimaalseid sisendmuudatusi peamiselt failinimede täpsustamisel. Käskude käsitsi käivitamine oli tülikam, kuid kaheksa proovi (üks lahjendamata ja seitse lahjendatud) analüüsi ajakulu oli umbes neli minutit ja arvestades kogu andmestikku, siis *K. pneumoniae* proovide töötlemiseks kulus keskmiselt 13 tundi ja *E. coli* puhul 11 tundi. Ajakulu vähendas märgatavalt *nohup* käsu kasutamine, mis võimaldas mitme proovi paralleelset analüüsi, mida RASE puhul ei saanud rakendada.

2.4. Arutelu

Käesoleva töö raames uurisin antimikroobse resistentsuse ennustamist sekveneerimisandmetest kahe tarkvaraga: RASE ja ResFinder, et määrata uuritavate proovide resistentsust ja tundlikkust. Praktilises töös analüüsisin 25 erinevat *Klebsiella pneumoniae* ja 21 *Escherichia coli* tüve uurides, kui tundlikud programmid on. WGS-i antimikroobse resistentsuse ennustamise programmid on populaarsed, et vähendada laboratoorset ajakulu ning leida resistentseid fenotüüpe kiiremini (Břinda et al., 2020; MacFadden et al., 2014).

Käesoleva töö tulemused viitavad siiski, et andmetöötlus ning RASE tarkvara kasutamine võivad osutada ajamahukamaks, kui esmalt eeldatud. RASE tarkvara analüüsimisel osutus oluliseks andmebaasi suurus, kus *K. pneumoniae* andmebaas oli pea 10 korda suurem *E. coli* omast. Ajaliselt kulus RASE-l kõikide proovide analüüsile pea kolm päeva, siis lahjendatud proovide ajakulu piirdus ühe päevaga. Seega võib järeldada, et RASE tarkvara arvutuskiirus sõltub andmebaasi suurusest.

ResFinder sai võrreldes RASE-ga paremini hakkama *Klebsiella pneumoniae* lahjendatud proovide resistentsuse tuvastamisega meropeneemi näitel. Tuleb tõdeda, et ResFinder otsib otseselt teadaolevaid resistentsusgeene, aga RASE otsib genoomipõhise naabertüpiseerimise abiga lähedasemaid naabreid uuritavale tüvele. ResFinderi andmetel tuvastati *Klebsiella pneumoniae* tüvedest õigesti meropeneemi resistentsusgeene (joonis 6), kus lahjendatud proovides (10^0-10^6) määrati õigesti 29% ja tundlikkust 98%. RASE leidis meropeneemile resistentsust 4% ja tundlikkust 100% õigesti. Antud tulemustest on kooskõlas RASE tundlikkuse protsent (Carroll et al., 2025). Edasises töös peab kasutatavasse RASE andmebaasi lisama rohkem resistentseid meropeneemi genoome, et usaldusväärsemaid tulemusi saada.

RASE *K. pneumoniae* andmebaas oli suunatud Euroopa päritoluga tüvedele. Käesolevas töös kasutatud tüved polnud tõenäoliselt kooskõlas andmebaasis esindatud referentsidega, mis võis olla üheks põhjuseks madalamale täpsusele. Uuritavad proovid võeti juhuslikult, mis andsid LS skoori üle 0,5. Teisest küljest on kirjanduse ülevaate järgi *Klebsiella pneumoniae* resistentne β-laktaamidele (Li et al., 2023) ja see kajastub tsefepiimi tulemustes (joonis 5), kuid mitte meropeneemi tulemustes (joonis 6). Antud töö väikse valimi tõttu on võimalik, et kõik olulised genotüübid ja fenotüübid polnud andmebaasis esindatud, mis võis mõjutada ennustustulemuste vastavust kirjanduses mainitule.

ResFinderi ja RASE täpsus kombinatsioonravimite resistentsuse tuvastamisel jäi kahetiseks. *K. pneumoniae* puhul uuritud piperatsilliini ja tasobaktaami kombinatsiooni puhul leidis RASE resistentsust 13% ja tundlikkust 75% kõikidest lahjendusastmetes sh lahjendamata proovist. ResFinder tuvastas seevastu resistentsusgeene ka proovides, mis olid fenotüüpiliselt tundlikud (joonis 7). Õigesti ennustatud resistentsuse osakaal oli 31%, kuid tundlikkuse täpsus jäi madalamaks – 71%. Lahjendamata proovide seast ei tuvastanud ResFinder ühtegi tundlikku proovi. See võib olla seotud ResFinderi tööpõhimõttega, mis ei arvesta geeniekspressiooni ega mutatsioonide mõju valgu funktsioonile.

Efflukspumpade, poriinvalkude ja mutatsioonide olemasolu võivad viia olukorrani, kus proov klassifitseeritakse ekslikult resistentseks (Elías-López et al., 2024; Hussein et al., 2024). Piperatsilliini ja tasobaktaami kombinatsiooni puhul tuvastas ResFinder bla-OXA1 geen mitmest proovist: SRR5385697, SRR5385727, SRR5385805, SRR5385829, SRR5385966, SRR5386739, SRR5386795, SRR5386823, SRR5386853, SRR5387337. Kui lisada ResFinderi tulemustele PointFinderi analüüs, siis ilmnes, et eelmainitud proovidest (va SRR5385697) esines ompK36 geenis punktmutatsioone, mida on kirjanduses seostatud resistentsuse tekkega (Elías-López et al., 2024; Shropshire et al., 2021). blaOXA-1 geeni amplifikatsioon koos OmpK36 poriini mutatsioonidega võib viia β-laktaam resistentsuseni (Shropshire et al., 2021). *E. coli* trimetoprimmi ja sulfametoksasooli analüüsil leidsin, et nii RASE kui ka ResFinder suudavad tundlikkust ennustada (joonis 9), kuid *K. pneumoniae* piperatsilliini ja tasobaktaami kombinatsioonil ei suutnud (joonis 7). Antud töös ei vastanud kõik genotüüpilised leiud fenotüüpilisele resistentsusele, mistõttu on oluline edasine analüüs, kuidas ja miks sellised vastuolud tekivad.

E. coli proovide analüüsil leidsin, et tsiprofloksatsiini resistentsuse ennustustulemused olid RASE tarkvaral märgatavalt kõrgemad kui ResFinderil (joonis 8), kuid tundlikkus jäi madalaks (33%). Antud antibiootikumi suur resistentsuse osakaal uuritavates tüvedes ei korreleeru kirjanduses mainituga, kus tsiprofloksatsiin on üks tõhusamaid antibiootikume *E. coli* vastastes infektsioonide vastu võitlemises (Ericsson, 2003; Morita ja Kuroda, 2025; Qadri et al., 2005). Teisalt korreleerub tundlikkuse protsent tsiprofloksatsiini ResFinderi tundlikkuse omaga, kus alternatiivses uuringus oli antud antibiootikumi täpsus pea 98% (Verschuuren et al., 2022).

Käesoleva töö tulemused näitavad, et RASE ega ResFinder ei suutnud metagenoomsetest andmetest resistentsust usaldusväärselt tuvastada. Samas oli tundlikkuse ennustamine lahjendatud proovides märkimisväärselt täpsem, ületades 90% läve. Lahjendatud proovide põhjal oli *K.*

pneumoniae keskmine resistentsuse määramise täpsus RASE-l 5,75% ja ResFinderil 25,75%. *E. coli* puhul olid vastavad näitajad 13,25% ja 16,25%. Eelmainituga võrreldes olid lahjendamata proovide keskmised resistentsuse ennustustulemused kõrgemad: *K. pneumoniae* puhul RASE täpsus 30,25% ja ResFinderil 100%. *E. coli* puhul seevastu 34,75% ja 69,75%. ResFinder suutis resistentsusgeene tuvastada kuni lahjenduseni 10^4 (joonis 5), *E. coli* korral tuvastas RASE resistentsust kuni lahjenduseni 10^3 (joonis 8). Antud tulemused viitavad selgelt sellele, et isolaadipõhistes proovides (ilma metagenoomse taustata) on antibiootikumi resistentsuse ennustamine tunduvalt usaldusväärsem, samas kui metagenoomne lisa võib vähendada märgatavalt ennustuste täpsust. Tulemust võisid mõjutada ka tausta-DNA liigipuhtus.

RASE madalate ennustustulemuste taga võis peituda mitmed asjaolud. Tuleb tõdeda, et horisontaalne geeniülekanne võis RASE tulemusi mõjutada, mida tarkvara arendaja on välja toonud (Břinda et al., 2020). Teiseks võimalikuks selgituseks võis olla asjaolu, et RASE on algselt välja töötatud kolmanda põlvkonna sekveneerimisandmete jaoks. Siin töös kasutasin NGS andmeid, mis võivad olla seotud madalama täpsusega. Kolmas selgitus tuleneb RASE artiklis öelduga, et mitme mikroorganismi kombinatsioonil võivad k-meerid kattuda RASE andmebaasis ja tekitada eksitavaid tulemusi (Břinda et al., 2020). Seetõttu on vajalik arendada meetodeid, mis suudaksid paremini toime tulla keeruka taustaga ja madala bakterikontsentratsiooniga proovidest sõltumata kasutatavast sekveneerimisandmete tüübist. Edasistes töödes võiks kaaluda alternatiivsete AMR tööriistade kasutamist, millest Core-Gene-AMR-model põhineb konserveerunud geenidel (Nguyen et al., 2020). Antud mudeli põhjal saab resistentsust ennustada osalistest genoomidest ja ka oluliste AMR geenide osalisel puudumisel (Nguyen et al., 2020).

Proovide tundlikkuse ja resistentsuse tuvastamist ning ennustamist võis mõjutada kontaminatsioon algsetest uuritavatest tüvedest (Anjum et al., 2017). Kraken2 analüüs, mis viidi läbi pärast tulemuste analüüsi, näitas et enamik proove kuulusid vastavale liigile. *E. coli* puhul leiti, et 21/21 proovi vastasid taksonoomilistele ootustele. Erandinda ilmus, et üks *K. pneumoniae* uuritavatest (SRR3185094) polnud taksonoomiliselt puhas (lisa 9). Antud leid võis mõjutada nii genotüüpilisi kui ka fenotüüpilisi ennustustulemusi, eriti kui metagenoomne taust on keerukas. Ideaalses olukorras oleks pidanud taksonoomilise kontrolli läbi viima kohe pärast uuritavate proovide LS skoori kinnitamist, et eemaldada ebasobivad tüved.

ResFinder oli stabiilsem ja tundlikum metagenoomsete proovide resistentsuse ennustamisel, samas kui RASE saavutas kõrge tundlikkuse samas andmestikus. Tulemusi mõjutasid nii andmebaaside

koostis kui ka võimalik kontaminatsioon uuritavates tüvedes, mis viitab kvaliteedikontrolli vajadusele enne AMR analüüsi. Usaldusväärse resistentsuse ennustamine keerulisemates proovides on jätkuvalt väljakutse, mistõttu peaksid edasised arendused keskenduma nii andmebaaside täiendamisele kui ka tarkvarade täiustamisele.

KOKKUVÕTE

Käesoleva töö põhieesmärk oli uurida kahe erineva antimikroobse resistentsuse tuvastamise tarkvara tööefektiivsust kasutades teise põlvkonna sekveneerimisandmeid. Uuringus keskendusin *Klebsiella pneumoniae* ja *Escherichia coli* tüvede resistentsuse ning tundlikkuse prognoosimisele nii isolaadipõhises kui ka simuleeritud metagenoomse taustaga proovides.

ResFinder oli stabiilsem resistentsuse ennustamisel isolaadipõhistes ehk lahjendamata proovides, saavutades *Klebsiella pneumoniae* puhul 100% täpsuse võrreldes RASE 30,25%-ga. *Escherichia coli* puhul tuvastas samuti ResFinder resistentsust paremini 69,75% täpsusega RASE 34,75% ees. Tuleb tõdeda, et RASE tarkvara suutis pakkuda väga kõrget usaldusväärsust tundlikke proovide puhul, mis viitab RASE sobilikkusele tundlikkuse ennustamiseks.

Simuleeritud metagenoomse taustaga proovides vähenes resistentsuse ennustamise täpsus märkimisväärselt mõlema programmi puhul. RASE saavutas metagenoomse taustaga ehk lahjendatud proovides *Klebsiella pneumoniae* resistentsuse keskmise täpsuse kõigest 5,75% ja ResFinder 25,75%. *Escherichia coli* korral olid vastavad näitajad 13,25% ja 16,25%. Tundlikkuse määr oli endiselt kõrge, kus RASE tuvastas tundlikkust mõlema liigi lahjendatud proovide puhul 95% ja ResFinder 93% täpsusega. See viitab, et metagenoomne taust mõjutab programmide võimet resistentsust ennustada.

Lahjendusastmete analüüs näitas, et *K. pneumoniae* puhul suutis ResFinder säilitada mõningase töökindluse kuni lahjenduseni 10^4, kus resistentsusgeene veel tuvastati. *E. coli* korral suutis RASE leida resistentsust kuni 10^3 lahjenduseni. RASE puhul oli resistentsuse tuvastamine lahjendatud proovides peaaegu puudulik (<6%), samas kui tundlikkuse prognoos püsis kõrgel kogu lahjendusastmete vahemikus. Seega toimis resistentsuse usaldusväärne ennustamine peamiselt lahjendamata proovides ning mõningal juhul mõõduka lahjenduse korral.

Detection and prediction of antimicrobial resistance from Illumina sequencing data

Heleri Savolainen

Summary

Antimicrobial resistance (AMR) poses a significant threat to public health due to increasing ineffectiveness of antibiotics against pathogenic bacteria. AMR can arise from various bacterial mechanisms such as horizontal gene transfer, altered gene regulation and modifications in porins and efflux pumps. Traditional resistance detection starts with phenotypic testing using polymerase chain reaction (PCR). While PCR is widely used in today's clinical microbiology, bioinformatic methods are gaining increasing importance. Although traditional approaches aim to detect known resistance genes, high genetic variability among AMR genes still exists. Therefore, broader data integration is essential to improve the performance of bioinformatic tools.

Whole genome sequencing (WGS) allows for a more comprehensive detection of resistance determinants. However, accurate prediction also depends on the quality of available reference databases which remain as a limitation to computational AMR detection. These factors highlight the growing need for bioinformatic approaches that support real-time clinical decision making.

In this thesis, the bioinformatic tools ResFinder and RASE (Resistance-Associated Sequence Elements) were evaluated for predicting antimicrobial resistance using second-generation (NGS) sequencing data. RASE employs genomic-neighbour typing, while ResFinder detects known resistance genes. Both tools were tested on clinically relevant *Klebsiella pneumoniae* and *Escherichia coli* strains. The aim was to assess each tool's accuracy in predicting resistance and susceptibility in isolates and simulated metagenomic samples as well as across a dilution series. The experimental part involved analyzing 25 *K. pneumoniae* and 21 *E. coli* examples. Tool outputs were compared to lab phenotypes and Kraken2 to verify sample purity. Accuracy and sensitivity were measured across dilutions.

ResFinder performed better in resistance prediction for isolates achieving 100% accuracy for K. pneumoniae and 69,75% for E. coli. RASE was more accurate for susceptibility, maintaining 95% accuracy in diluted samples. Resistance detection declined in metagenomic conditions but maintained reliable for susceptibility. The results indicate that current AMR prediction tools are more robust for susceptibility than resistance, highlighting the need for improved database coverage and performance in low-DNA or complex samples.

KIRJANDUSE LOETELU

Anjum, M. F., Zankari, E., & Hasman, H. (2017). Molecular Methods for Detection of Antimicrobial Resistance. *NAT MICROBIOL*, *5*, 3–4.

Bortolaia, V., Kaas, R. S., Ruppe, E., ... Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *J ANTIMICROB CHEMOTH*, *75*, 3491–3500.

Břinda, K., Callendrello, A., Ma, K. C., ... Hanage, W. P. (2020). Rapid inference of antibiotic resistance and susceptibility by genomic neighbour typing. *NAT MICROBIOL*, *5*, 455–464.

Carroll, A. C., Mortimer, L., Ghosh, H., ... MacFadden, D. R. (2025). Rapid inference of antibiotic susceptibility phenotype of uropathogens using metagenomic sequencing with neighbor typing. *NAT MICROBIOL*, *13*, 1366–24.

Carroll, A. C., Mortimer, L., MacFadden, D., & McGeer, A. (2024b). *Rapid Inference of Antibiotic Susceptibility Phenotype of Uropathogens using Metagenomic Sequencing with Neighbour Typing—RASE Database for EuSCAPE* (Version v1) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.11121227

Carroll, A. C., Mortimer, L., MacFadden, D., & McGeer, A. (2024a). *Rapid Inference of Antibiotic Susceptibility Phenotype of Uropathogens using Metagenomic Sequencing with Neighbour Typing—RASE Databases for E. coli and Klebsiella spp* (Version v1) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.11042579

Chien, J.-Y., Chiu, W.-Y., Chien, S.-T., Chiang, C.-J., Yu, C.-J., & Hsueh, P.-R. (2016). Mutations in *gyrA* and *gyrB* among Fluoroquinolone- and Multidrug-Resistant Mycobacterium tuberculosis Isolates. *ANTIMICROB AGENTS CH*, *60*, 2090–2096.

Clausen, P. T. L. C., Aarestrup, F. M., & Lund, O. (2018). Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC BIOINFORMATICS*, *19*, 1–4.

Cloete, T. E. (2003). Resistance mechanisms of bacteria to antimicrobial compounds. *INT BIODETER BIODEGR*, *51*, 277–282.

Drlica, K., & Zhao, X. (1997). DNA Gyrase, Topoisomerase IV, and the 4-Quinolones. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS,* 377–381.

Drouin, A., Giguère, S., Déraspe, M., Marchand, M., Tyers, M., Loo, V. G., Bourgault, A.-M., Laviolette, F., & Corbeil, J. (2016a). Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. *BMC GENOMICS*, *17*, 2–5.

Drouin, A., Raymond, F., Letarte St-Pierre, G., Marchand, M., Corbeil, J., & Laviolette, F. (2016b). Large scale modeling of antimicrobal resistance with interpretable classifiers. *arXiv*, 1–4.

ECDC, 2024. (2023). *Antimicrobial consumption in the EU/EEA (ESAC-Net)—Annual epidemiological report for 2023* (pp. 9–13) [Surveillance report]. European Centre for Disease prevention and Control.

Elías-López, C., Muñoz-Rosa, M., Guzmán-Puche, J., Pérez-Nadales, E., Chicano-Galvez, E., & Martínez-Martínez, L. (2024). Porin expression in clinical isolates of Klebsiella pneumoniae. *ANN CLIN MICROB ANTI*, *23*, 6–7.

Ellington, M. J., Ekelund, O., Aarestrup, F. M., ... Woodford, N. (2017). The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: Report from the EUCAST Subcommittee. *CLIN MICROBIOL INFEC*, *23*, 2–22.

Ericsson, C. D. (2003). Travellers' diarrhoea. *INT J ANTIMICROB AG*, *21*, 116–124.

Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D., & Aarestrup, F. M. (2022). ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microbial Genomics*, *8*.

Gin, A., Karlowsky, J. A., Walkty, A., Rubinstein, E., & Zhanel, G. G. (2007). Piperacillin-tazobactam: A beta-lactam/beta-lactamase inhibitor combination. *EXPERT REV ANTI-INFE*, *5*, 365–383.

Hasman, H., Hammerum, A. M., Hansen, F., ... Skov, R. L. (2015). Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. *EUROSURVEILLANCE*, *20*.

Hendriksen, R. S., Bortolaia, V., Tate, H., Tyson, G. H., Aarestrup, F. M., & McDermott, P. F. (2019). Using Genomics to Track Global Antimicrobial Resistance. *Front. Public Health*, *7*, 1–2.

Hu, K., Meyer, F., Deng, Z.-L., Asgari, E., Kuo, T.-H., Münch, P. C., & McHardy, A. C. (2024). Assessing computational predictions of antimicrobial resistance phenotypes from microbial genomes. *BRIEF BIOINFORM*, 2.

Hussein, R. A., AL-Kubaisy, S. H., & Al-Ouqaili, M. T. S. (2024). The influence of efflux pump, outer membrane permeability and β-lactamase production on the resistance profile of multi, extensively and pandrug resistant Klebsiella pneumoniae. *J INFECT PUBLIC HEAL*, *17*, 102544.

Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: Past, present and future. *CURR OPIN MICROBIOL*, 72–80.

Hyun, J. C., Kavvas, E. S., Monk, J. M., & Palsson, B. O. (2020). Machine learning with random subspace ensembles identifies antimicrobial resistance determinants from pan-genomes of three pathogens. *PLOS COMPUT BIOL*, *16*, 1–2.

Kosek, M., Bern, C., & Guerrant, R. L. (2003). The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *B WORLD HEALTH ORGAN*, *81*.

Kowalska-Krochmal, B., & Dudek-Wicher, R. (2021). The Minimum Inhibitory Concentration of Antibiotics: Methods, Interpretation, Clinical Relevance. *Pathogens MDPI*, 1–3.

Larrañaga, P., Calvo, B., Santana, R., & Robles, V. (2006). Machine learning in bioinformatics. *BRIEF BIOINFORM*, 87–88.

Li, Y., Kumar, S., Zhang, L., Wu, H., & Wu, H. (2023). Characteristics of antibiotic resistance mechanisms and genes of *Klebsiella pneumoniae*. *OPEN MED-WARSAW*, *18*, 2–5.

Li, Y., Metcalf, B. J., Chochua, S., Beal, B. W. (2016). Penicillin-Binding Protein Transpeptidase Signatures for Tracking and Predicting β-Lactam Resistance Levels in Streptococcus pneumoniae. *MBIO*, e00756-16–18.

MacFadden, D. R., Leis, J. A., Mubareka, S., & Daneman, N. (2014). The Opening and Closing of Empiric Windows: The Impact of Rapid Microbiologic Diagnostics. *CLIN INFECT DIS*, 1199–1201.

McManus, M. C. (1997). Mechanisms of bacterial resistance to antimicrobial agents. *Primer*, *54*, 1420–1422.

Moeckel, C., Mareboina, M., Konnaris, M. A., Chan, C. S. Y., & Mouratidis, I. (2024). A survey of k-mer methods and applications in bioinformatics. *Computational and Structural Biotechnology Journal*, 2289–2290.

Morita, D., & Kuroda, T. (2025). Recent Antimicrobial Resistance Situation and Mechanisms of Resistance to Key Antimicrobials in Enterotoxigenic *Escherichia coli*. *BIOL PHARM BULL*, *48*, 222–229.

Nguyen, M., Olson, R., Shukla, M., VanOeffelen, M., & Davis, J. J. (2020). Predicting antimicrobial resistance using conserved genes. *PLOS COMPUT BIOL*, 1–3.

Pesesky, M. W., Hussain, T., Wallace, M., Patel, S., Andleeb, S., Burnham, C.-A. D., & Dantas, G. (2016). Evaluation of Machine Learning and Rules-Based Approaches for Predicting Antimicrobial Resistance Profiles in Gram-negative Bacilli from Whole Genome Sequence Data. *FRONT MICROBIOL*, 2–3.

Pidcock, K. A., & Piotrowski, N. A. (2023). Antibiotics. *Magill's Medical Guide (Online Edition)*, 1–3.

Pretzel, E. (2015). About Antibiograms. *Minnesota Department of Health Infectious Disease Epidemiology, Prevention, and Control Division*.

Qadri, F., Svennerholm, A.-M., Faruque, A. S. G., & Sack, R. B. (2005). Enterotoxigenic *Escherichia coli* in Developing Countries: Epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. *CLIN MICROBIOL REV*, *18*, 465–483.

Seoane, A., Bou, G., & Complejo Hospitalario Universitario A Coruña. (2020). Bioinformatics approaches to the study of antimicrobial resistance. *REV ESP QUIM*, *34*, 15–17.

Shropshire, W. C., Aitken, S. L., Pifer, R., & Shelburne, S. A. (2021). IS *26* -mediated amplification of *bla* OXA-1 and *bla* CTX-M-15 with concurrent outer membrane porin disruption associated with *de novo* carbapenem resistance in a recurrent bacteraemia cohort. *J ANTIMICROB CHEMOTH*, *76*, 385–395.

Sikic, K., & Carugo, O. (2010). Protein sequence redundancy reduction: Comparison of various methods. *Bioinformation*, *5*(6), 234–239.

Swedan, S. F., & Aldakhily, D. B. (2024). Antimicrobial resistance, biofilm formation, and molecular detection of efflux pump and biofilm genes among Klebsiella pneumoniae clinical isolates from Northern Jordan. *NAT MICROBIOL*, *10*, 2–3.

Tenover, F. C. (2006). Mechanisms of antimicrobial resistance in bacteria. *AM J INFECT CONTROL*, *5*, S3–S5.

Verschuuren, T., Bosch, T., Mascaro, V., Willems, R., & Kluytmans, J. (2022). External validation of WGS-based antimicrobial susceptibility prediction tools, KOVER-AMR and ResFinder 4.1, for Escherichia coli clinical isolates. *CLIN MICROBIOL INFEC*, *28*, 1465–1470.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *GENOME BIOL*, *15*, 1–12.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *J ANTIMICROB CHEMOTH*, *11*, 2640–2644.

KASUTATUD VEEBIAADRESSID

Bioinformaatika töögrupp (2025). Kasutatud 22.05.2025. https://github.com/bioinfo-ut/random

BV-BRC (*Bacterial and Viral Bioinformatics Resource Center*) (2025). Kasutatud 11.05.2025. https://www.bv-brc.org/

Coursera (2025). Kasutatud 14.01.2025. https://coursera.org/share/ddc3059c29f8c822354020cbfed78dca

Core-Gene-AMR-models (2025). Kasutatud 25.05.2025. https://github.com/jimdavis1/Core-Gene-AMR-Models

DrugBank (2025). Kasutatud 19.05.2025. https://go.drugbank.com/

ENA (*European Nucleotide Archive*) (2025). Kasutatud 11.05.2025. https://www.ebi.ac.uk/ena/browser/home

Illumina (2024). Kasutatud 27.10.24. https://www.illumina.com/techniques/sequencing/dna-sequencing/whole-genome-sequencing.html

Kover–AMR Platform (2017) (*PATRIC database*). Kasutatud 27.01.25. https://aldro61.github.io/kover-amr-platform/index.html

PAHO (*Pan American Health Organization*). Kasutatud 20.05.2025. https://www3.paho.org/data/index.php/en/mnu-topics/antimicrobial-resistance/578-amr-eva-en.html

RASE (2020). Kasutatud 16.01.2025. https://github.com/c2-d2/rase-db-skeleton

ResFinder (2025). Kasutatud 13.05.2025. http://genepi.food.dtu.dk/resfinder

ResFinder bitbucket (2020). Kasutatud 5.03.2025. https://bitbucket.org/genomicepidemiology/resfinder/src/master/

TechTarget (2024). Autor Gavin Wright. Kasutatud 2.11.24. https://www.techtarget.com/whatis/definition/empirical-analysis

LISAD

Lisa 1.

***Klebsiella pneumoniae* kasutatavad tüved RASE ja ResFinderi andmeanalüüsis**. Tabelis on välja toodud tundlikkus ja resistentsus erinevatele antibiootikumidele, mis on BV-BRC kinnitatud. Sekveneerimislugemid pärinevad ENA andmebaasist, sellest tulenevalt ka tabelis olev proovi nimi.

Proovi nimi	Resistentne	Tundlik
ERR3164722	Klooramfenikool, tsiprofloksatsiin, ertapeneem, gentamütsiin	Amikatsiin, tseftasidiim/avibaktaam, kolistiin, fosfomütsiin, imipeneem, tigetsükliin, trimetoprim/sulfametoksasool
ERR1023711	Amikatsiin, amoksitsilliin/klavulaanhape, ampitsilliin, tsefasoliin, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, gentamütsiin, nitrofurantoiin, norfloksatsiin, piperatsilliin/tasobaktaam, tikartsilliin/klavulaanhape, tobramütsiin, trimetoprim/sulfametoksasool	Tsefepiim, tsefoksitiin, meropeneem
ERR1218573	Tsefuroksiim, amoksitsilliin/klavulaanhape, ampitsilliin, asteronaam, tsefalotiin, tsefepiim, tsefotaksiim, tseftasidiim, tsiprofloksatsiin, gentamütsiin, tigetsükliin, trimetoprim	Amikatsiin, tsefoksitiin, ertapeneem, meropeneem
SRR3185094	Amikatsiin, amoksitsilliin/klavulaanhape, ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefotaksiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, ertapeneem, gentamütsiin, imipeneem, levofloksatsiin, meropeneem, moksiofloksantiin, nitrofurantoiin, piperatsilliin/tasobaktaam, tigetsükliin, tobramütsiin, trimetoprim/sulfametoksasool	Ertapeneem
SRR5386491	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tsefoksitiin, tseftasidiim, tseftriaksoon, tsefuroksiim, imipeneem, meropeneem, piperatsilliin/tasobaktaam, tetratsükliin	Amikatsiin, tsiprofloksatsiin, gentamütsiin, levofloksatsiin, tobramütsiin, trimetoprim/sulfametoksasool

SRR5387337	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, gentamütsiin, tetratsükliin, tobramütsiin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, imipeneem, levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam
SRR8607464	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tsefotaksiim, klooramfenikool, tsiprofloksatsiin, kolistiin, fosfomütsiin, imipeneem, meropeneem, piperatsilliin/tasobaktaam, polümüksiin B, tobramütsiin, trimetoprim/sulfametoksasool	Gentamütsiin, minotsükliin, tetratsükliin, tigetsükliin, fosfomütsiin
SRR5385697	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, levofloksatsiin, nitrofurantoiin, tetratsükliin, trimetoprim/sulfametoksasool	Amikatsiin, gentamütsiin, imipeneem, meropeneem, piperatsilliin/tasobaktaam
SRR5386595	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, gentamütsiin, tetratsükliin, tobramütsiin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, imipeneem, levofloksatsiin, meropeneem
SRR5385727	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, levofloksatsiin, tetratsükliin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, gentamütsiin, imipeneem, meropeneem, piperatsilliin/tasobaktaam
SRR5385904	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, gentamütsiin, levofloksatsiin	Tetratsükliin, trimetoprim/sulfametoksasool
SRR5385721	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, gentamütsiin, levofloksatsiin, piperatsilliin/tasobaktaam, tobramütsiin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, imipeneem, meropeneem, tetratsükliin

SRR5386303	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, levofloksatsiin, piperatsilliin/tasobaktaam, trimetoprim/sulfametoksasool	Amikatsiin, gentamütsiin, imipeneem, meropeneem, tobramütsiin
SRR5386412	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, levofloksatsiin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, gentamütsiin, imipeneem, meropeneem, piperatsilliin/tasobaktaam, tetratsükliin, tobramütsiin
SRR5386739	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftriaksoon, tsiprofloksatsiin, gentamütsiin, tetratsükliin, tobramütsiin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, imipeneem, levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam
SRR5387223	Ampitsilliin, tsefasoliin, tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, trimetoprim/sulfametoksasool	Amikatsiin, ampitsilliin/sulbaktaam, asteronaam, tsefepiim, tsefoksitiin, tseftasidiim, gentamütsiin, imipeneem, meropeneem piperatsilliin/tasobaktaam, tetratsükliin, tobramütsiin
SRR5385767	Ampitsilliin, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsefuroksiim, tetratsükliin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, tsiprofloksatsiin, gentamütsiin, imipeneem, levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam, tobramütsiin
SRR538789	Ampitsilliin, ampitsilliin/sulbaktaam, tsefasoliin, tsefoksitiin, tseftasidiim, tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, gentamütsiin, imipeneem, levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam, tetratsükliin, trimetoprim/sulfametoksasool	Amikatsiin, asteronaam, tsefepiim, tobramütsiin
SRR5385805	Ampitsilliin, ampitsilliin/sulbaktaam, tsefasoliin, tsefepiim, tseftasidiim,	Amikatsiin, tsefoksitiin, gentamütsiin, imipeneem,

	tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, tetratsükliin, trimetoprim/sulfametoksasool	levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam
SRR5385966	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, levofloksatsiin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, gentamütsiin, imipeneem, levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam, tetratsükliin
SRR5386664	Ampitsilliin, ampitsilliin/sulbaktaam, tsefasoliin, tseftriaksoon, tsefuroksiim, gentamütsiin, tetratsükliin, trimetoprim/sulfametoksasool	Amikatsiin, asteronaam, tsefoksitiin, tseftasidiim, imipeneem, levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam
SRR5386853	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, levofloksatsiin, nitrofurantoiin, tetratsükliin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, gentamütsiin, imipeneem, levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam
SRR5386823	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, gentamütsiin, levofloksatsiin, tetratsükliin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, imipeneem, meropeneem, piperatsilliin/tasobaktaam
SRR5386795	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tseftasidiim, tseftriaksoon, tsefuroksiim, gentamütsiin, tetratsükliin, tobramütsiin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, imipeneem, levofloksatsiin, meropeneem
SRR5385829	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsefuroksiim, tsiprofloksatsiin, gentamütsiin, tetratsükliin, tobramütsiin, trimetoprim/sulfametoksasool	Amikatsiin, tsefoksitiin, imipeneem, levofloksatsiin, meropeneem, piperatsilliin/tasobaktaam

Lisa 2.

***Escherichia coli* kasutatavad tüved RASE ja ResFinderi andmeanalüüsis**. Tabelis on välja toodud tundlikkus ja resistentsus erinevatele antibiootikumidele, mis on BV-BRC kinnitatud. Sekveneerimislugemid pärinevad ENA andmebaasist, sellest tulenevalt ka tabelis olev proovi nimi.

Proovi number	Resistentne	Tundlik
ERR2146823	-	Amikatsiin, amoksitsilliin/klavulaanhape, tsefotaksiim, tseftasidiim, tsiprofloksatsiin, gentamütsiin, meropeneem, piperatsilliin/tasobaktaam
ERR2146822	Tsiprofloksatsiin	Amikatsiin, amoksitsilliin/klavulaanhape, tsefotaksiim, tseftasidiim, gentamütsiin, meropeneem, piperatsilliin/tasobaktaam
ERR3062081	Tsefotaksiim, tseftasidiim, gentamütsiin, tobramütsiin, trimetoprim	Ampitsilliin, tsefuroksiim, tsefalotiin, tsiprofloksatsiin
ERR3062030	Trimetoprim	Ampitsilliin, tsefotaksiim, tseftasidiim, tsefuroksiim, tsefalotiin, tsiprofloksatsiin, gentamütsiin, tobramütsiin
ERR2205971	Ampitsilliin, tsefotaksiim, tsefoksitiin, tseftasidiim	Kolistiin, imipeneem, meropeneem
SRR4017845	Amoksitsilliin/klavulaanhape, ampitsilliin, asteronaam, tsefasoliin, tsefotaksiim, tsefoksitiin, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, ertapeneem, gentamütsiin, levofloksatsiin, piperatsilliin/tasobaktaam, trimetoprim/sulfametoksasool	Amikatsiin, tsefepiim, doripeneem, imipeneem, meropeneem, tetratsükliin, tigetsükliin
SRR5420067	Ampitsilliin, tsefasoliin, tseftriaksoon, tsiprofloksatsiin, levofloksatsiin, tetratsükliin	Amikatsiin, amoksitsilliin/klavulaanhape, tsefepiim, tseftasidiim, ertapeneem, gentamütsiin, meropeneem, nitrofurantoiin, piperatsilliin/tasobaktaam, tobramütsiin, trimetoprim/sulfametoksasool

SRR8844665	Ampitsilliin, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, trimetoprim/sulfametoksasool	Amikatsiin, doksütsükliin, gentamütsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin
SRR8844668	Ampitsilliin, tsiprofloksatsiin	Amikatsiin, ampitsilliin/sulbaktaam, asteronaam, tsefepiim, tseftasidiim, tseftriaksoon, doksütsükliin, gentamütsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin, trimetoprim/sulfametoksasool
SRR8844686	Ampitsilliin, tsiprofloksatsiin, doksütsükliin, gentamütsiin, trimetoprim/sulfametoksasool	Amikatsiin, ampitsilliin/sulbaktaam, asteronaam, tsefepiim, tseftasidiim, tseftriaksoon, imipeneem, meropeneem, minotsükliin, nitrofurantoiin
SRR8844688	Ampitsilliin, asteronaam, tsefasoliin, tsefepiim, tseftriaksoon, tsiprofloksatsiin, trimetoprim/sulfametoksasool	Amikatsiin, doksütsükliin, gentamütsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin
SRR8844689	Ampitsilliin, tsiprofloksatsiin, doksütsükliin, gentamütsiin, trimetoprim/sulfametoksasool	Amikatsiin, ampitsilliin/sulbaktaam, asteronaam, tsefepiim, tseftasidiim, tseftriaksoon, imipeneem, meropeneem, minotsükliin, nitrofurantoiin
SRR8844699	Ampitsilliin, tsefasoliin, tseftriaksoon, tsiprofloksatsiin, nitrofurantoiin	Amikatsiin, ampitsilliin/sulbaktaam, asteronaam, tsefepiim, tseftasidiim, gentamütsiin, imipeneem, meropeneem, minotsükliin, trimetoprim/sulfametoksasool
SRR8844702	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, doksütsükliin, trimetoprim/sulfametoksasool	Gentamütsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin
SRR8844741	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, doksütsükliin, trimetoprim/sulfametoksasool	Amikatsiin, gentamütsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin

SRR8844747	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tseftriaksoon, tsiprofloksatsiin, gentamütsiin	Amikatsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin, trimetoprim/sulfametoksasool
SRR8844754	Ampitsilliin, tsefasoliin, tseftriaksoon, tsiprofloksatsiin, doksütsükliin	Amikatsiin, ampitsilliin/sulbaktaam, asteronaam, tseftasidiim, gentamütsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin, trimetoprim/sulfametoksasool
SRR8844755	Ampitsilliin, tsefasoliin, tseftriaksoon, tsiprofloksatsiin, doksütsükliin	Amikatsiin, ampitsilliin/sulbaktaam, gentamütsiin, imipeneem, meropeneem, minotsükliin, trimetoprim/sulfametoksasool
SRR8844764	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, gentamütsiin	Imipeneem, meropeneem, minotsükliin, nitrofurantoiin, trimetoprim/sulfametoksasool
SRR8844765	Ampitsilliin, ampitsilliin/sulbaktaam, asteronaam, tsefasoliin, tsefepiim, tseftasidiim, tseftriaksoon, tsiprofloksatsiin, doksütsükliin, trimetoprim/sulfametoksasool	Amikatsiin, gentamütsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin
SRR8844770	Ampitsilliin, ampitsilliin/sulbaktaam, tsefasoliin, tseftriaksoon, tsiprofloksatsiin	Amikatsiin, doksütsükliin, gentamütsiin, imipeneem, meropeneem, minotsükliin, nitrofurantoiin, trimetoprim/sulfametoksasool

Lisa 3.

Tausta-DNA faili koostamine. Simuleeritud taust koosnes ühest inimese ja kuuest sagedasti inimese veres leiduva bakteri DNA-st. Inimese DNAst proovi (SRA kood on ERR191146) on võetud juhuslikult miljon juhuslikku lugemit seqtk programmiga kasutades käsku:

```
seqtk sample -s88 ERR194146_all.fastq 1000000 > sample_ERR194146.fastq.1000000
```

Bakterliigid oli *Bacillus cereus, Ralstonia solanacearum, Clostridium botulinum, Staphylococcus epidermidis, Pseudomonas fluorescens, Streptomyces venezuelae*. Varasemast tööst olid kõigist bakteritest olemas proovid 15 000 juhusliku lugemiga. Miljon inimese DNA lugemit ja 6*15 000 bakteri lugemit pandi kokku käsuga:

```
cat sample_BC.fastq.15000 sample_CB.fastq.15000 sample_PF.fastq.15000
sample_RP.fastq.15000 sample_SE_2.fastq.15000 sample_SV.fastq.15000
sample_ERR194146.fastq.1000000 > segamata_lahjendus.fastq
```

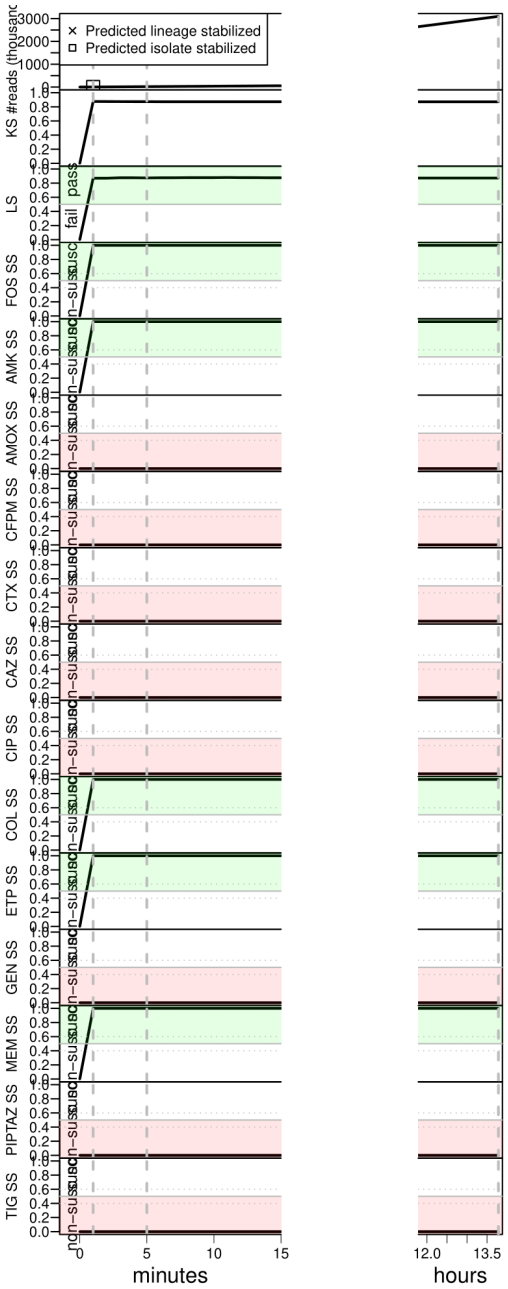
Edasi segati simuleeritud prooviks:

```
#!/bin/bash

cat segamata_lahjendus.fastq |\

awk '{printf("%s",$0);n++;if(n%4==0){printf("\n");}else{printf("\t\t");}}' |\

shuf |\

sed 's/\t\t/\n/g' > human_background_DNA.fastq
```

Tegu on töögrupis tehtud tausta-DNA failiga, mille koostas töögrupis olev kolleeg.

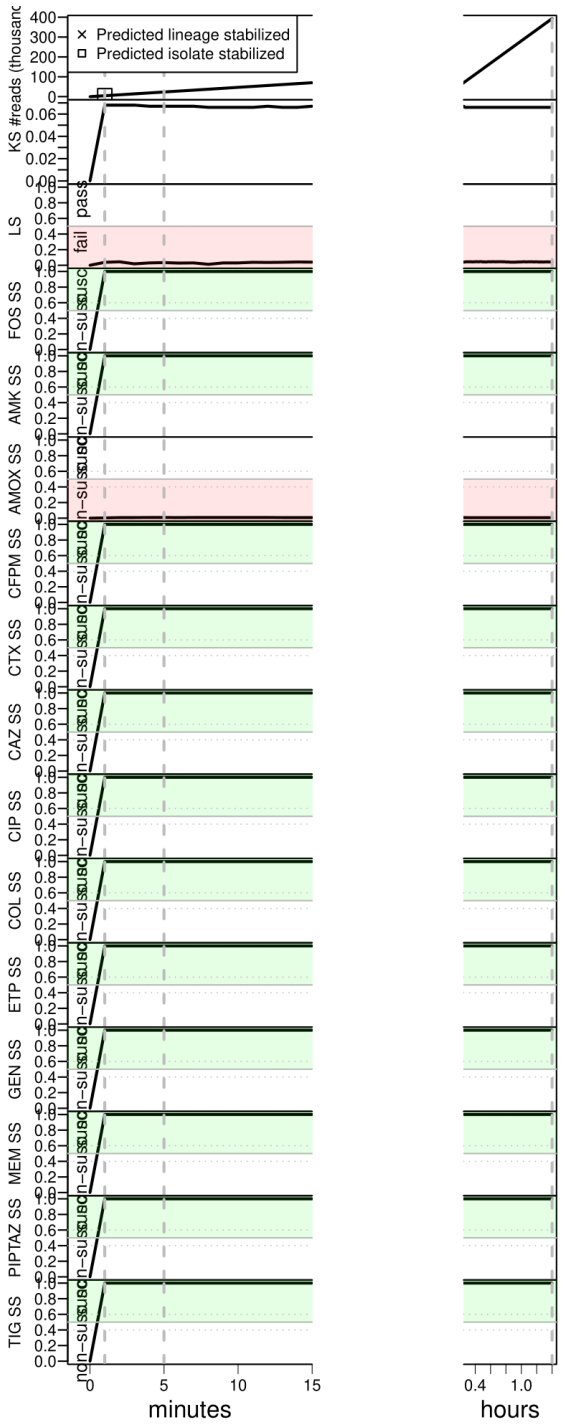
Lisa 4.

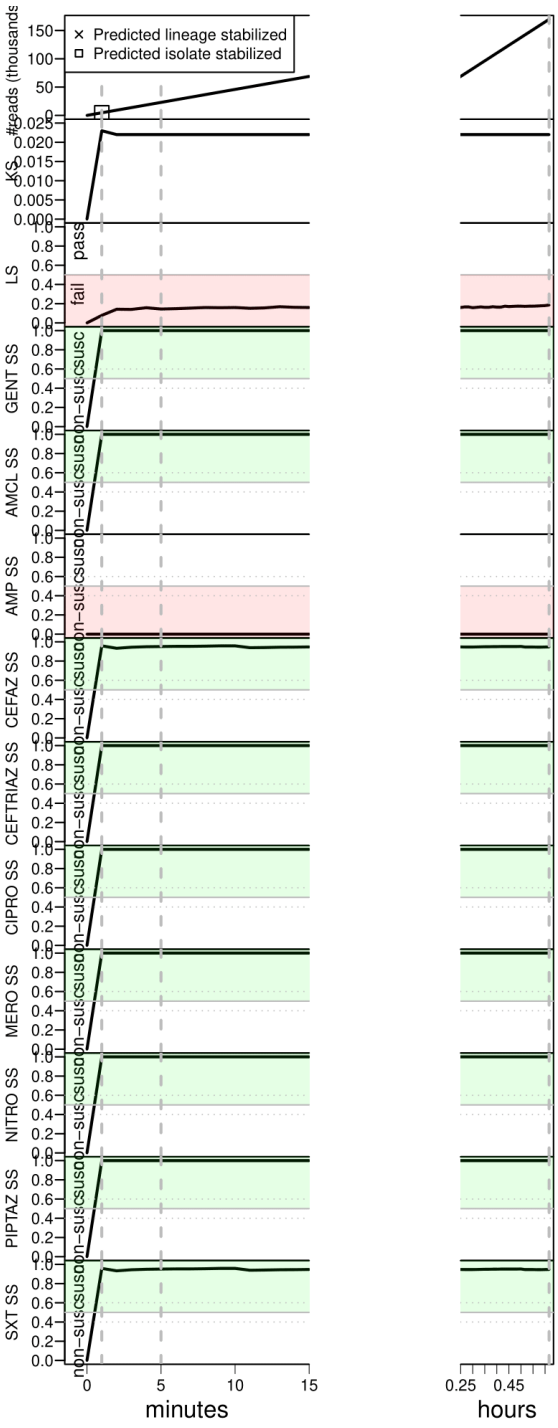
RASE tarkvara väljundfail R-graafikuna. Näites on toodud välja *Klebsiella pneumoniae* proovi SRR5385805 graafik, mis sisaldab nii antibiootikumide resistentsust kui ka sugulusskoori (LS) ja tundlikkuse skoori (SS). Rohelisega määratud alad viitavad tundlikkusele ning punasega märgitud alad resistentsusele. X-teljel on määratud aeg, mis on möödunud reaalajas sekveneerimise algusest. Antud tüved pole reaalajas sekveneeritud, seega tarkvara määrab ise aja sekveneerimise algusest ka NGS FASTQ faili puhul. Antud näite puhul on tundlikkust ennustatud järgmiste antibiootikumide puhul: fosfomütsiin (FOS), amikatsiin (AMK), kolistiini (COL), ertapeneem (ETP), meropeneem (MEM). Resistentsust on ennustatud amoksitsilliini (AMOX), tsefepiimi (CFPM), tseftriaksooni (CTX), tseftasidiimi (CAZ), tsiprofloksatsiini (CIP), gentamütsiini (GEN), piperatsilliin/tasobaktaami (PIPTAZ) ja tigetsükliini (TIG) puhul.



Lisa 5.

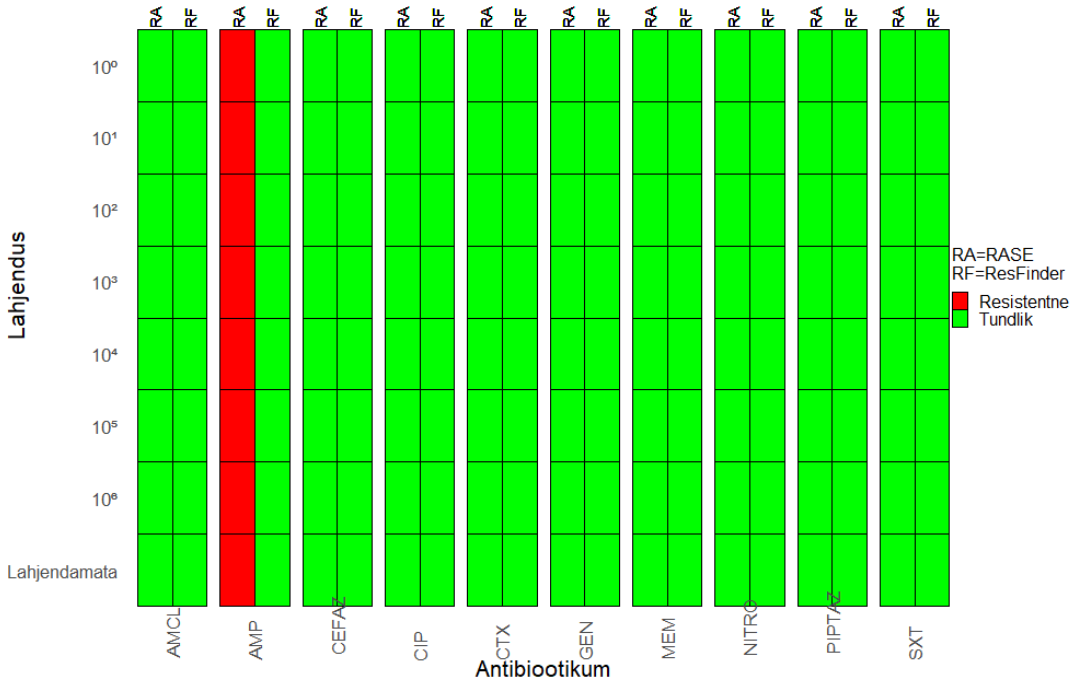
Tausta-DNA põhjal koostatud RASE R-graafikud *Klebsiella pneumoniae* ja *Escherichia coli* andmebaaside vastu. Vasakpoolsel R-i graafikul on välja toodud RASE andmebaasi *K. pneumoniae* tausta-DNA proovi ennustused, mille LS skoor on 0,042. Resistentsust tuvastati amoksitsilliini (AMOX) vastu, mida on näha ka tundlikkuse skoori (SS) läbikukkumisest. Parempoolsel graafikul on RASE andmebaasi *E. coli* tasuta-DNA proovi ennustused erinevate antibiootikumide vastu. Resistentsust tuvastati ampitsilliini (AMP) vastu.





Lisa 6.

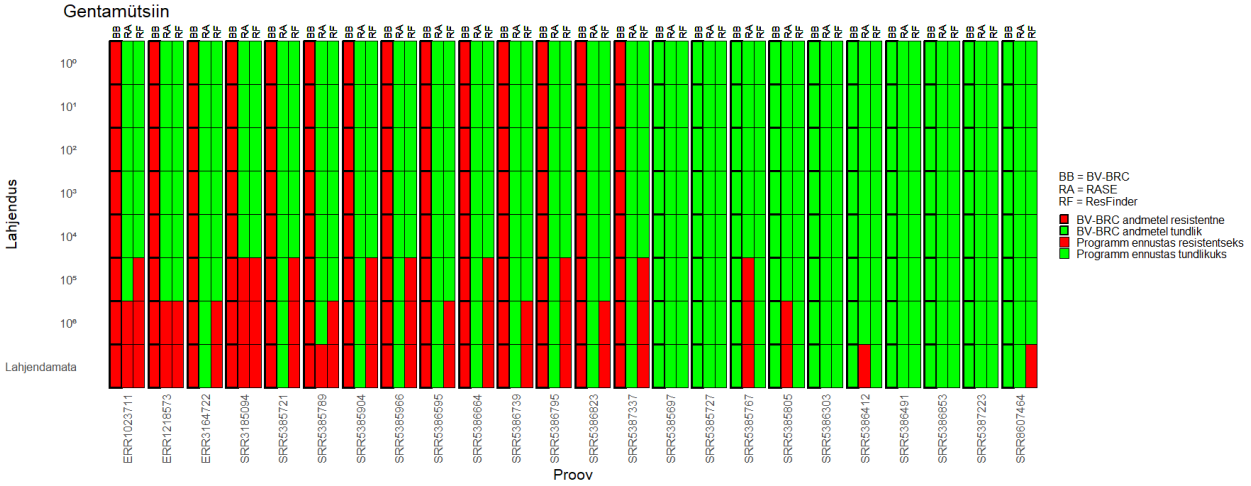
Simuleeritud inimese vere DNA proovi AMR-i ennustustulemused *Escherichia coli* RASE andmebaasi ja ResFinderi andmebaaside põhjal. Graafiku x-teljel on erinevad antibiootikumid, mida leidus RASE andmebaasis. Y-teljel on lahjendamata proov, mis on ainult tausta-DNA ning edasi 10^0-10^6 on lahjendatud tausta-DNA proovi fenotüübid RASE andmebaasis leidus 10 antibiootikumi, millest tausta-DNA proov oli resistentne ainult ampitsilliinile (AMP). ResFinderi andmebaasides ei leitud ühtegi resistentsusgeeni antud antibiootikumide suhtes.



Lisa 7.

Gentamütsiini resistentsusennustuste võrdlus BV-BRC (BB) andmetega *K. pneumoniae* näitel. Graafiku X-teljel on esitatud erinevad proovid *K. pneumoniae* koos nende nimetusega ENA andmebaasist. Y-teljel on lahjendusastmed: esimene tase tähistab lahjendamata proovi, sellele järgnevad lahjendused vahemikus 10^0-10^6 lahjendused, mis koosnevad proovi ja tausta-DNA segust. Iga proovi esimene tulp kajastab laboratoorselt kinnitatud resistentsust või tundlikkust (BB). Teine ja kolmas tulp näitavad vastavalt RASE (RA) ja ResFinderi (RF) ennustusi. Punase värviga on tähistatud resistentsus ja rohelisega tundlikkus.

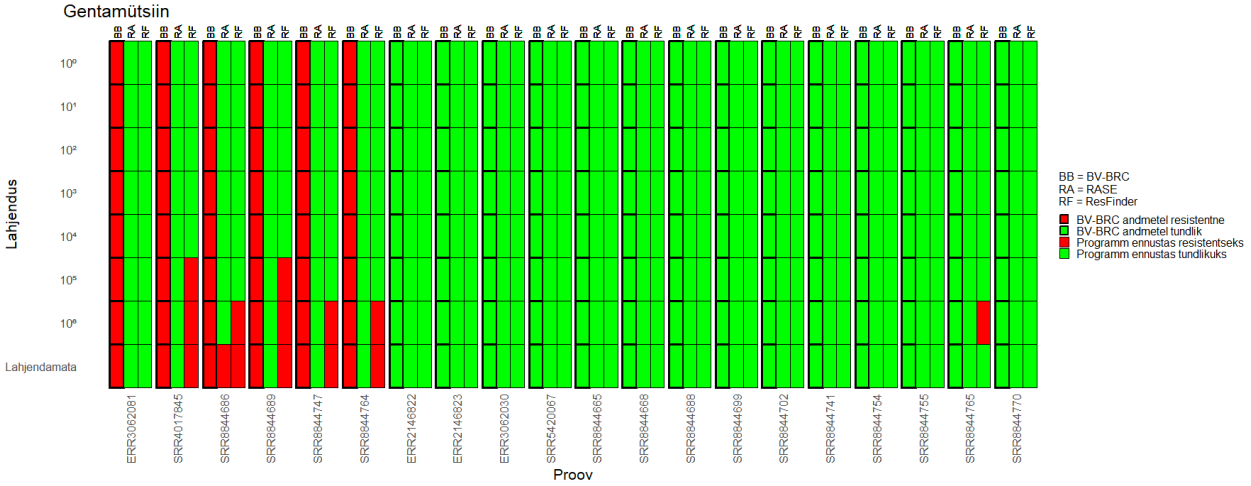
Lahjendamata proovide osas tuvastasid RASE ja ResFinder resistentsust täpsusega 29% (4/14) ja 100% (14/14). Tundlikkust tuvastati õigesti vastavalt 60% (6/10) ja 90% (9/10). Lahjendamata kuni 10^4 lahjendusastme piires oli RASE puhul resistentsuse täpsus 13% (8/64) ja ResFinderil 56% (36/64). Tundlikkuse täpsus oli vastavalt 85% (34/40) ja 98% (39/40). Kui uurida ainult lahjendusastmeid, langes resistentsuse määramine RASE-l 4% (4/98) ja ResFinderil 22% (22/98). Tundlikkuse määramine oli kõrgem 96% (67/70) ja 99% (19/70). Kogu andmestiku (sh) kõik lahjendusastmed põhjal tuvastas RASE õigesti 7% (8/112) ja ResFinder 32% (36/112) resistentsustulemustest. Tundlikkuse täpsus oli vastavalt 53% (74/140) ja 56% (79/140).



Lisa 8.

Gentamütsiini resistentsusennustuste võrdlus BV-BRC (BB) andmetega *E. coli* näitel. Graafiku X-teljel on esitatud erinevad proovid *E. coli* koos nende nimetusega ENA andmebaasist. Y-teljel on lahjendusastmed: esimene tase tähistab lahjendamata proovi, sellele järgnevad lahjendused vahemikus 10^0-10^6 lahjendused, mis koosnevad proovi ja tausta-DNA segust. Iga proovi esimene tulp kajastab laboratoorselt kinnitatud resistentsust või tundlikkust (BB). Teine ja kolmas tulp näitavad vastavalt RASE (RA) ja ResFinderi (RF) ennustusi. Punase värviga on tähistatud resistentsus ja rohelisega tundlikkus.

Lahjendamata proovide osas tuvastasid RASE ja ResFinder resistentsust täpsusega 17% (1/6) ja 83% (5/6). Tundlikkust tuvastati õigesti vastavalt 100% (14/14) ja 100% (14/14). Lahjendamata kuni 10^3 lahjendusastme piires oli RASE puhul resistentsuse täpsus 3% (1/30) ja ResFinderil 40% (12/30). Tundlikkuse täpsus oli vastavalt 100% (70/70) ja 98% (69/70). Kui uurida ainult lahjendusastmeid, langes resistentsuse määramine RASE-l 0% (0/42) ja ResFinderil 17% (7/42). Tundlikkuse määramine oli veidi kõrgem 100% (98/98) ja 99% (97/98). Kogu andmestiku (sh) kõik lahjendusastmed põhjal tuvastas RASE õigesti 2% (1/48) ja ResFinder 25% (12/48) resistentsustulemustest. Tundlikkuse täpsus oli vastavalt 100% (112/112) ja 99% (111/112).



Lisa 9.

Kraken2 analüüs uuritavate tüvede proovidega. Analüüs viidi läbi bioinformaatika õppetoolis kolleegide poolt. Kraken2 ja Bracken programmide abil analüüsiti töös kasutatud 25 *Klebsiella pneumoniae* ja 21 *Escherichia coli* sekveneerimislugemeid. Mõlema programmi puhul kasutati parameetrite vaikeväärtuseid, kuid Kraken2 juures määrati parameetri --confidence väärtuseks 0.5. *Confidence* skoor näitab, kui suurel määral peavad järjestuse k-meerid vastama andmebaasis olevatele taksonoomilistele üksustele, et lugem saaks liigituse. Kõrgem skoori väärtus suurendab klassifitseerimise täpsust, kuid võib vähendada tundlikkust, jättes rohkem lugemeid liigitamata.

Analüüs näitas, et 25 *K. pneumoniae* tüvest üks proov (SRR3185094) sisaldas vaid 80% *K. pneumoniae*'le omaseid järjestusi ja 14% lugemitest klassifitseeriti *E. coli*'ks. Sellest tulenevalt oleks antud proov tulnud analüüsist eemaldada, kuna see ei vastanud taksonoomilise ühtsuse kriteeriumile. Samas vastasid kõik 21 *E. coli* proovi kvaliteedinõuetele, kusjuures üle 90% lugemitest klassifitseeriti õigesti *E. coli*'ks.

LIHTLITSENTS

Mina, Heleri Savolainen ,
(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

   Antimikroobse resistentsuse tuvastamine ja ennustamine Illumina sekveneerimisandmetest ,
   (lõputöö pealkiri)

   mille juhendaja on Reidar Andreson ,
   (juhendaja nimi)

   reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;

2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni;

3. olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile;

4. kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Heleri Savolainen

26.05.2025