TOTAL RESISTOME PREDICTION

BBS SSP 2022

Friday, May 27th, 2022

Karyn Mukiri

ABOUT ME

- First-year MSc student
  - Undergrad: Biotechnology (B. Tech at Mac)
    - Mandatory programming courses (C++, Java, Python)
- Dabbled in both academia and industry
  - Academic experience was largely wet lab (Biohybrids Lab, Chem Eng. & fourth-year capstone)
  - Industry experience at two Coca-Cola production facilities
    - First introduction to real "ontologies"

ANTIBIOTIC RESISTANCE ONTOLOGY
KNOWLEDGE SPACE

Antibiotic molecule

Resistance-modifying agents

Mechanism of antibiotic resistance

Determinant of antibiotic resistance

Antibiotic target

Antibiotic biosynthesis

Component of AMR phenotype terminology

MODEL ONTOLOGY
ALGORITHM SPACE

Protein homolog model

Protein variant model

Protein overexpression model

rRNA gene variant model

Toxin system meta model

Protein knockout model

Nonfunctional insertion model

Efflux pump system meta-model

Protein domain meta-model

CARD CURATED MODEL TYPES AND DETECTION PARAMETERS

| Model Type / Model Parameter	Sum of Model Parameter Count
efflux pump system meta-model	33
efflux pump components	33
gene cluster meta-model	12
gene order	12
protein domain meta-model	1
domain order	1
protein homolog model	4577
BLASTP bit-score	4577
protein knockout model	11
BLASTP bit-score	11
protein overexpression model	32
single resistance variant	13
BLASTP bit-score	13
multiple resistance variants	1
deletion mutation from peptide sequence	1
nonsense mutation	1
deletion mutation from nucleotide sequence	1
insertion mutation from nucleotide sequence	1
frameshift mutation	1
protein variant model	462
BLASTP bit-score	163
single resistance variant	163
no association with resistance TB	17
multiple resistance variants	17
deletion mutation from nucleotide sequence	17
high confidence TB	14
insertion mutation from nucleotide sequence	14
nonsense mutation	13
minimal confidence TB	9
moderate confidence TB	8
indeterminate confidence TB	7
insertion mutation from peptide sequence	6
deletion mutation from peptide sequence	6
snp in promoter region	3
co-dependent single resistance variant	2
frameshift mutation	2
disruptive mutation in regulatory element	1
rRNA gene variant model	180
BLASTN bit-score	84
single resistance variant	83
high confidence TB	4
no association with resistance TB	3
moderate confidence TB	3
disruptive mutation in regulatory element	1
multiple resistance variants	1
minimal confidence TB	1
nonfunctional insertion model	0
N/A	0
Grand Total	**5308**

Supported by RGI

4

PERFECT

KNOWN KNOWNS

Matches reference
sequence
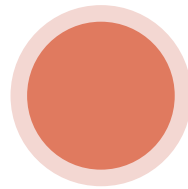
STRICT

KNOWN UNKNOWNS

Similarity within model
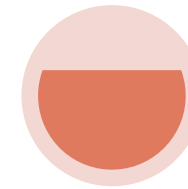
LOOSE/DISCOVERY

UNKNOWN UNKNOWNS

Similarity outside of
model
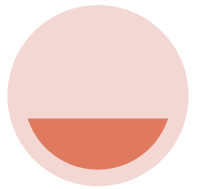
B I T S C O R E

Measures sequence similarity independent
of database size

PROJECT TL;DR

1. Fine-tune bitscore cut-offs to reduce false positive and false negative hits

2. Implement additional statistics to further reduce false positive hits

3. Develop additional code to address current gaps in RGI resistome prediction

WRIGHT CLINICAL COLLECTION

(691 sequenced MDR genomes)

PROTEIN HOMOLOG
86.48%

PROTEIN VARIANT
7.33%

PROTEIN OVEREXPRESSION
6.19%

rRNA GENE VARIANT
0%

1) Presence/absence of homologous proteins
2) Similarity to CARD curated reference sequences based on **BLAST bitscore cut-offs**

1) Presence/absence of homologous proteins
2) Secondary search for **curated antibiotic resistance-conferring mutations**

Detects overexpression based on the presence or absence of mutations

Detects rRNA genes containing AMR mutations, differentiating them from their wild type counterparts

NDM-1
(metallo-beta-lactamase)

SNP in *Staphylococcus aureus* gyrB
(confers resistance to aminocoumarin antibiotics)

Mutations in MexR (repressor of multi-drug efflux operon) of *Pseudomonas aeruginosa*

Point mutation in 16S rRNA of *Mycobacteroides abscessus* that confers aminoglycoside resistance

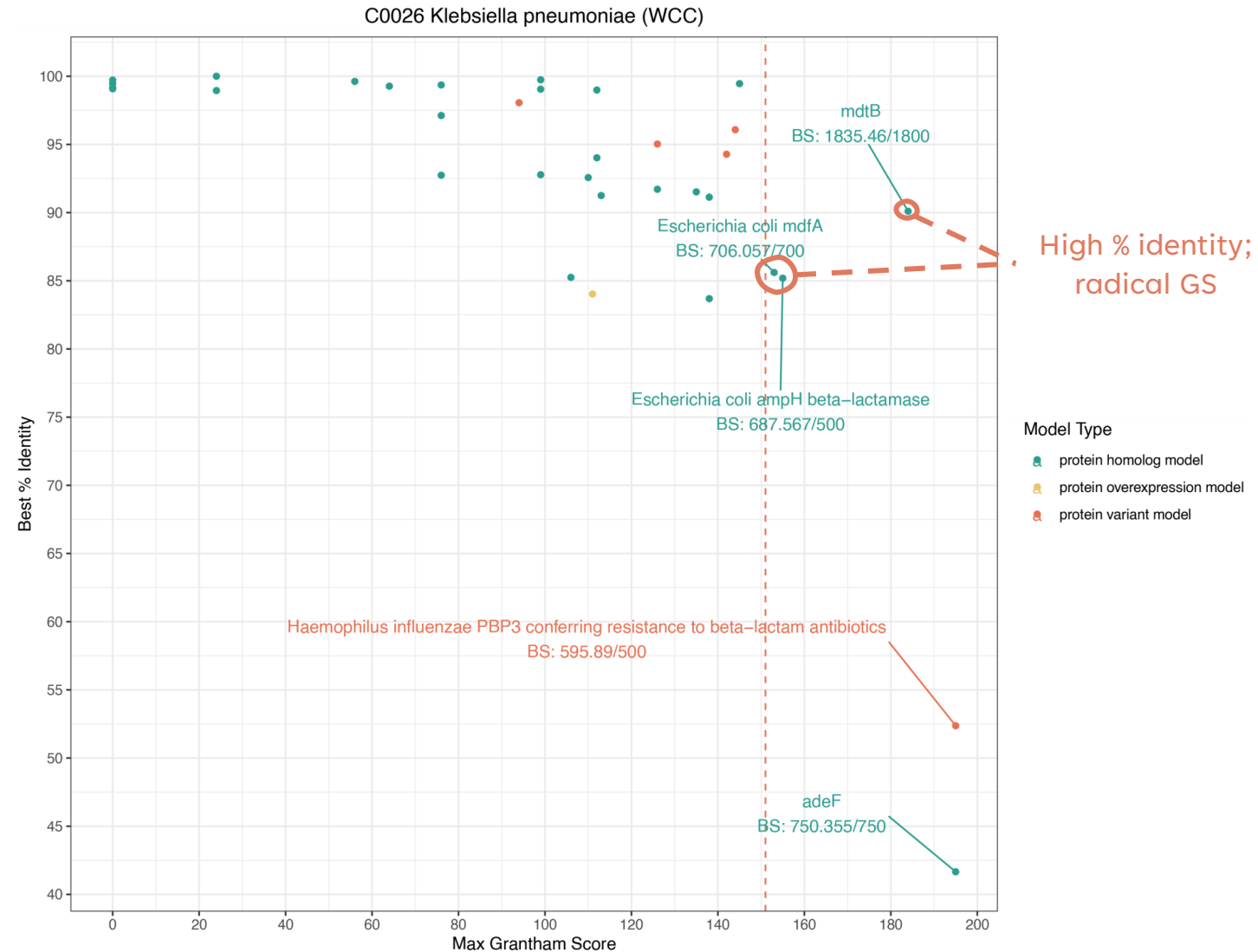BLAST: Model Comparison via Positives (30 files; 3 Pathogens; WCC)
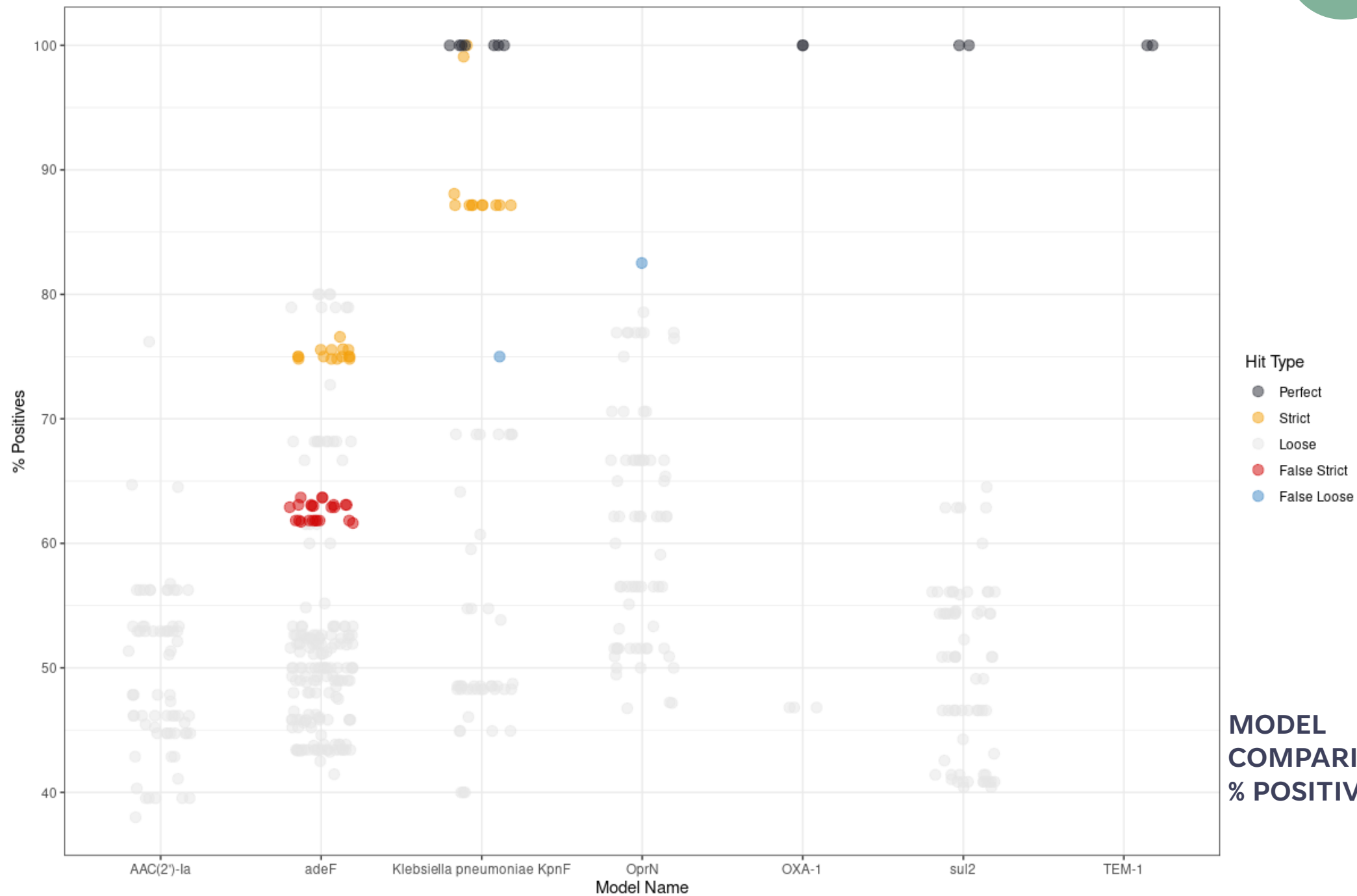
MODEL COMPARISON USING % POSITIVES

8

C0026 Klebsiella pneumoniae (WCC)

ASSESSING CURATION ACCURACY OF BITSCCORE CUT-OFFS VIA PERCENT IDENTITY

mdtB
BS: 1835.46/1800

Escherichia coli mdfA
BS: 706.057/700

Escherichia coli ampH beta–lactamase
BS: 687.567/500

Haemophilus influenzae PBP3 conferring resistance to beta–lactam antibiotics
BS: 595.89/500

adeF
BS: 750.355/750

Low % identity;
passing bitscores

Model Type
- protein homolog model
- protein overexpression model
- protein variant model

Best % Identity

Max Grantham Score

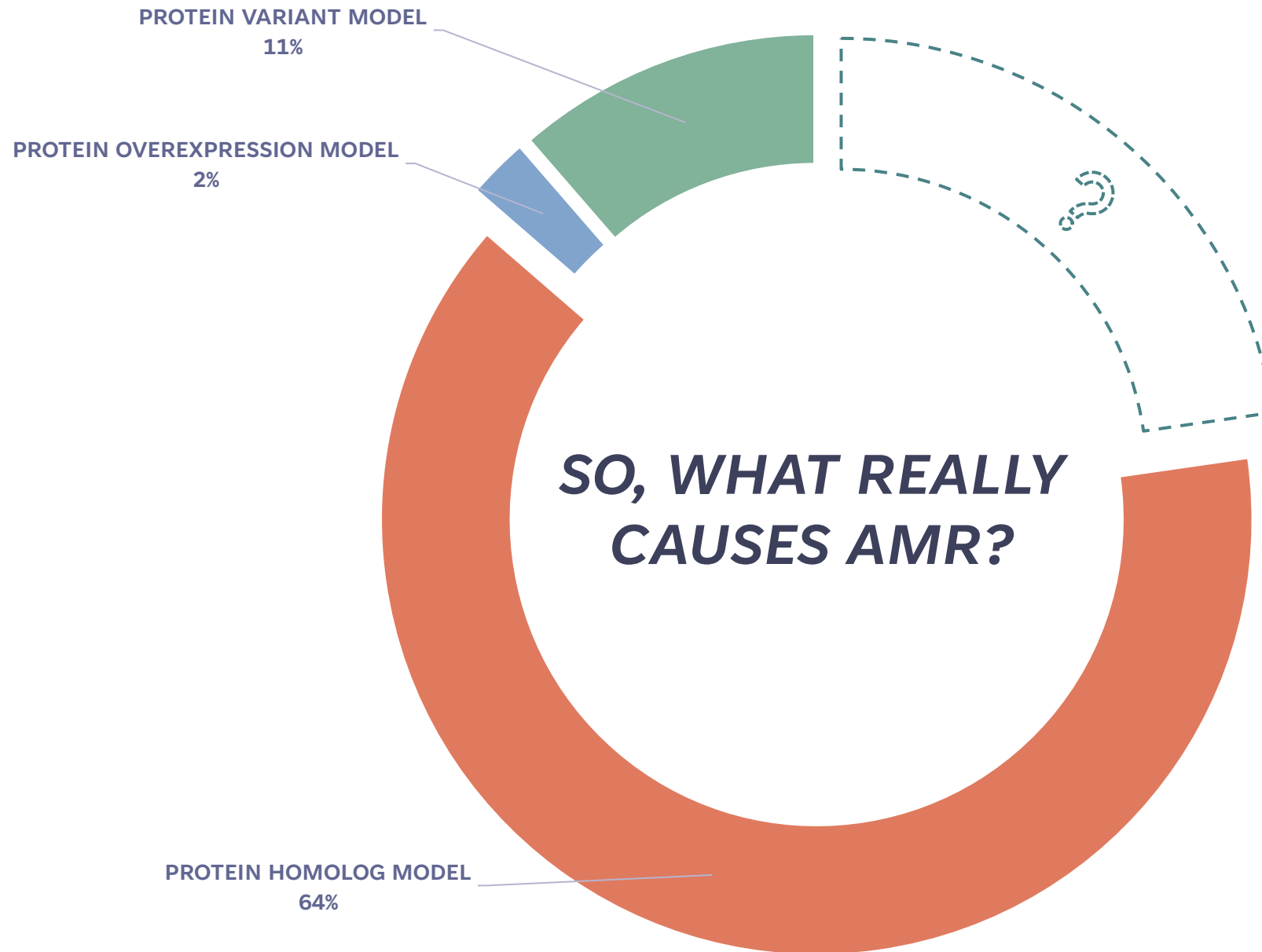# ADDITION OF GRANTHAM SCORES (GS) FOR RGI RESULT VALIDATION

- Categorizes amino acid substitutions and their level of dissimilarity

- Effect of substitutions is further categorized into four classes:

  - Conservative (0-50)

  - Moderately conservative (51-100)

  - Moderately radical (101-150)

  - Radical (≥151)

C0026 Klebsiella pneumoniae (WCC)

WHAT ABOUT THE CODE?

datacamp

PROTEIN VARIANT MODEL
11%

PROTEIN OVEREXPRESSION MODEL
2%

SO, WHAT REALLY CAUSES AMR?

PROTEIN HOMOLOG MODEL
64%

ACKNOWLEDGEMENTS

MCARTHUR LAB MEMBERS (2021-2022)

Amos Raphenya   Jalees Nasir   Emily Panousis
Brian Alcock   Arman Edalatmand   Mateusz Wlodarski
Keaton Smith   Mugdha Dave   Gene Liang
                        Sheridan Baker   Hamna Imtiaz
                        Madeline McCarthy   Sarah Yaqoob

COMMITTEE MEMBERS
Dr. Andrew McArthur
Dr. Lori Burrows
Dr. Brian Coombes

PAST LAB MEMBERS
Dr. Kara Tsang
Romeo Chalil