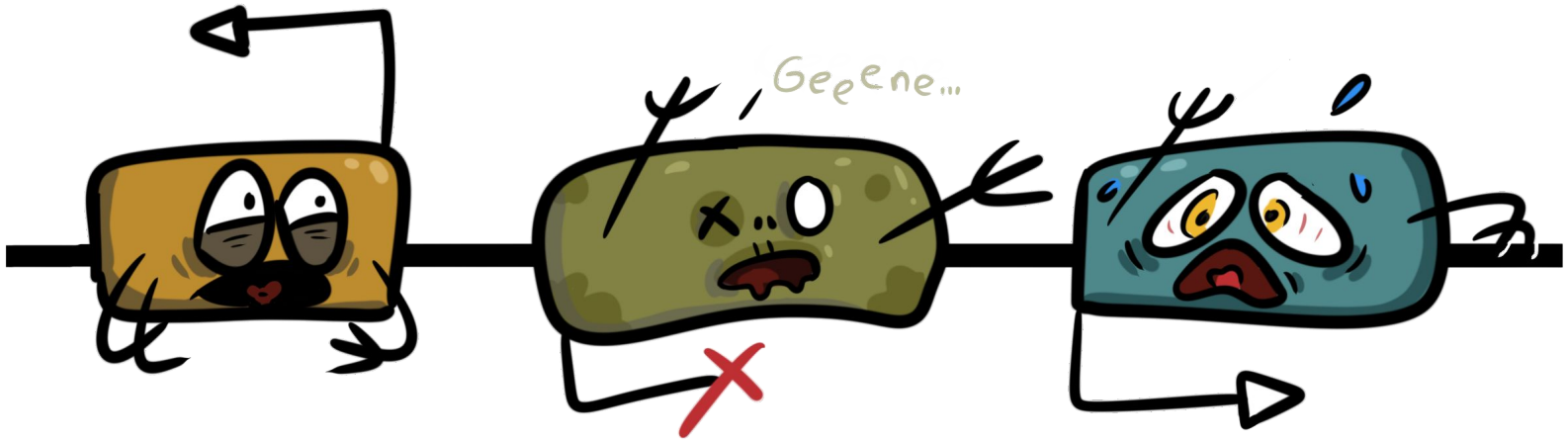
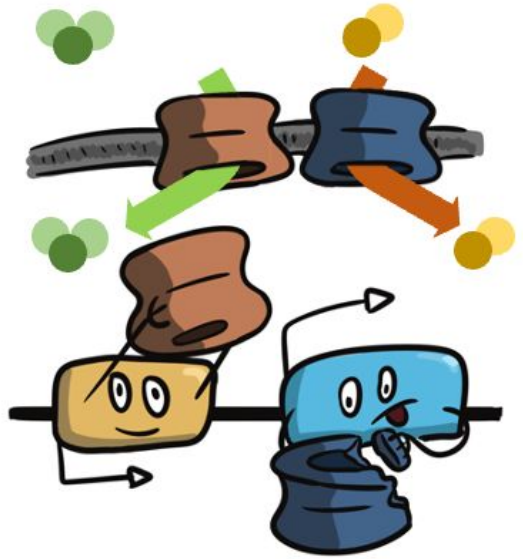


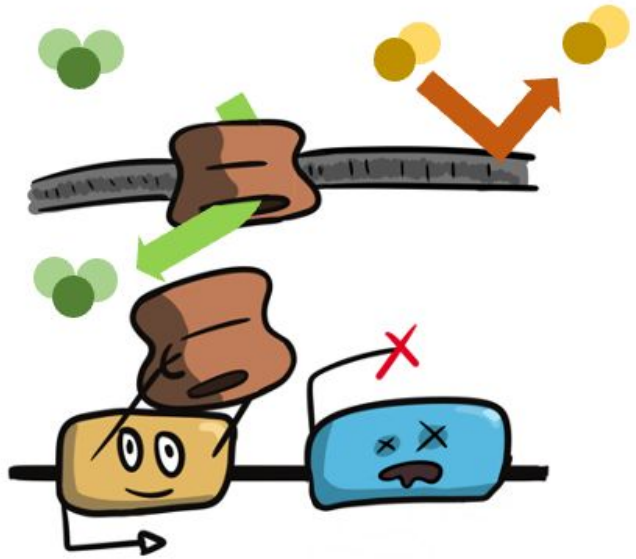
Dead Gene Walking

Team Prasad-Feldgarden:
Arjun Prasad, Team co-Lead
Michael Feldgarden, Team co-Lead
Adrien Assie, Tech Lead
Ana Ramos, Writer
EB Dickinson
Chienchi Lo
Erin Young

Breaking a gene can lead to resistance

- **Genes that are the target of antibiotics**
  - *K. pneumoniae* CirA - siderophore receptor used to import iron and the antibiotic cefiderocol
- **Porins that let antibiotics into the cell**
  - *K. pneumoniae* OmpK35/K36
  - Loss-of-function mutations confer resistance to carbapenems and other beta-lactams
- **Regulatory genes**
  - *P. aeruginosa* NalD - efflux system regulator
  - *P. aeruginosa* AmpD - beta-lactamase gene regulator
  - confer resistance to carbapenems and other beta-lactams
  - *A. baumannii* AdeS - efflux pump regulator

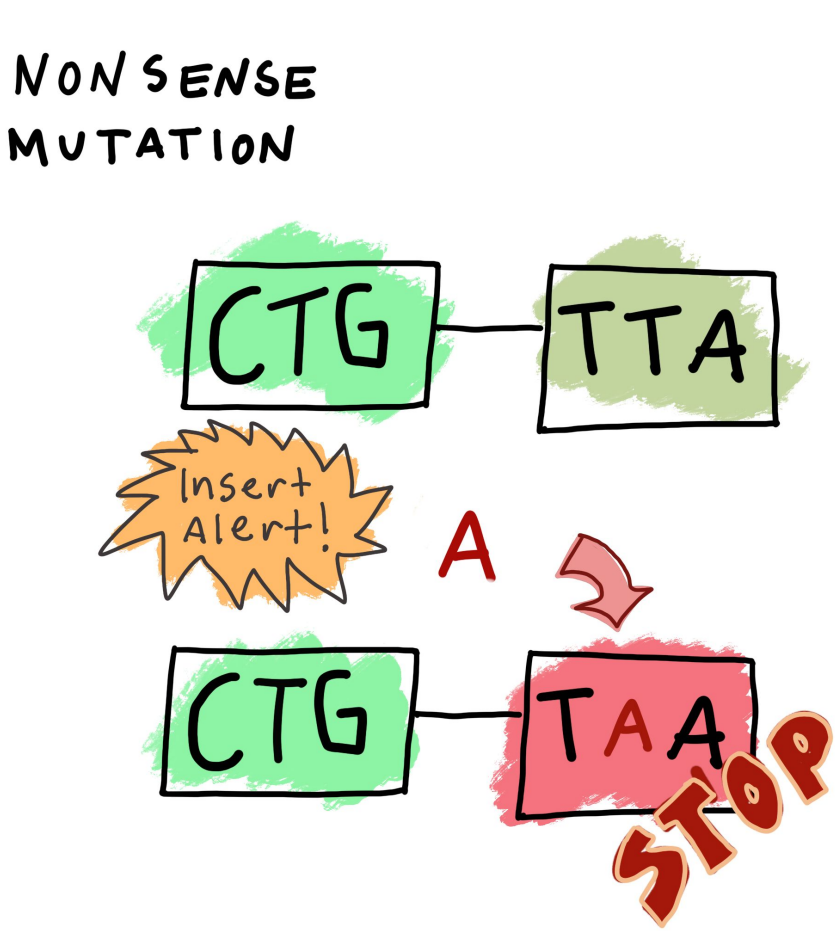


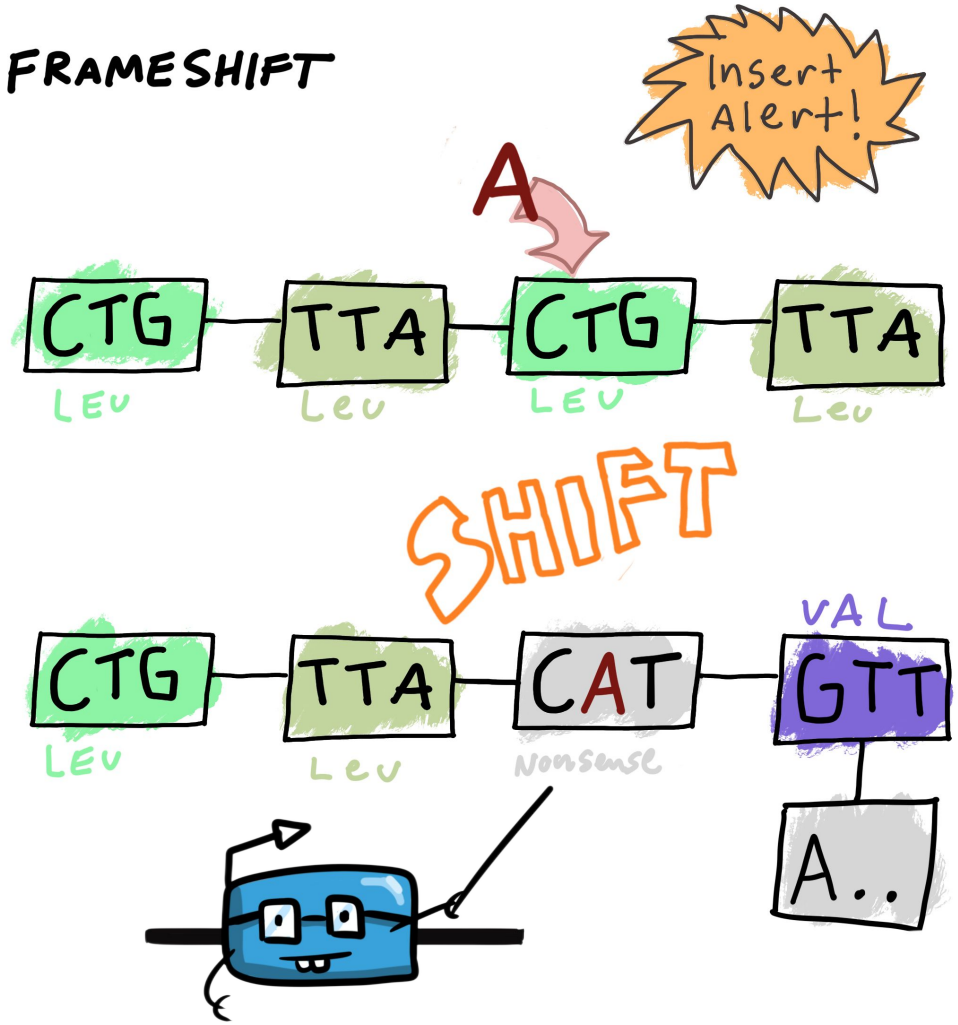


Project Overview

- Existing AMR gene detection software screens for known mutations
- Develop a tool and database to identify loss-of-function mutations in genes known to confer resistance when disrupted
  - Identify lesions that should ‘break’ the gene:
    - Frameshifts
    - Internal stops (nonsense mutations)
- Run the tool(s) on assemblies in NDARO to characterize the presence of novel loss-of-function mutations in those genes for three taxa
  - *Acinetobacter baumannii*
  - *Klebsiella pneumoniae*
  - *Pseudomonas aeruginosa*

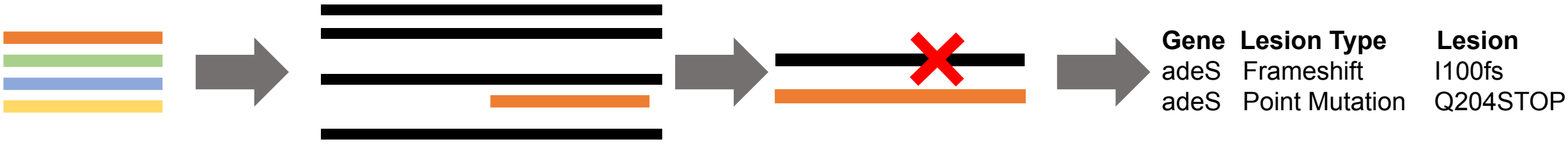
What are Nonsense Mutations and Frameshifts?





Dead Gene Scanner.R

- Use a curated database of reference genes
- Blastn reference gene in whole genomes
  - Aim to find genes that missed annotation prediction
  - Find stop mutation in existing genes
- Checks and reports for STOP point mutation or frameshift in results
  - Check 1000 genomes/min



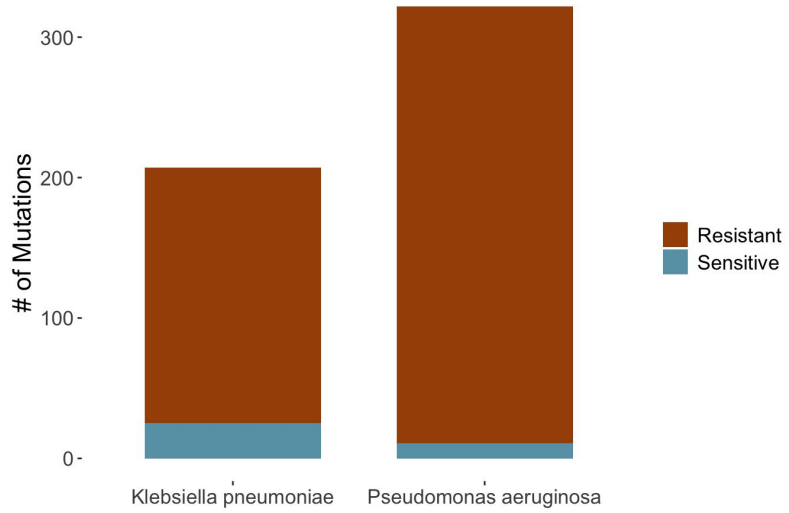
Technical approaches

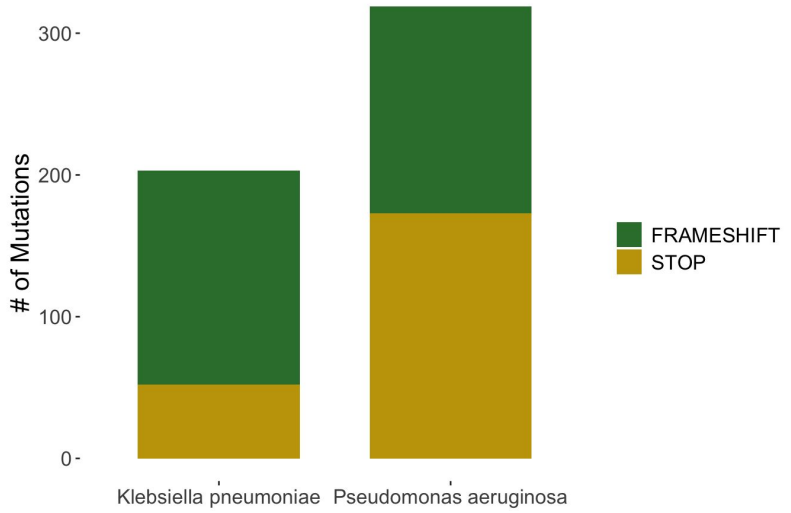
- **Assembly sequence to identify broken genes**
- Translated alignment with BLASTX
  - Stop codon detection
    - A python script to run BLASTX and check the * position in the alignment.
  - Frameshift detection
    - "Diamond" blastx has option --frameshift to allow frameshift gap aligned and show / or \ in the alignment result
    - In-frame mutation is noted with - in the alignment. However, it is common and may not cause genes broken. (default detection is disabled.)

```
usage: DGW_blast.py [-h] -i INPUT_FASTA -d DATABASE [-o OUTPUT] [--cov COV] [--id ID] [--indels] [--diamond]
                    [--version]

  Detect nonsensus/frameshift mutations by running blastx on contig fasta against target protein database

options:
  -h, --help            show this help message and exit
  -i INPUT_FASTA, --input_fasta INPUT_FASTA
                        contig fasta file
  -d DATABASE, --database DATABASE
                        protein (diamond) blast DB
  -o OUTPUT, --output OUTPUT
                        output tsv file
  --cov COV             target coverage
  --id ID               hit identity
  --indels              check in-frame insertion and deletion
  --diamond             use diamond blastx
  --threads THREADS     threads
  --verbose             Show more information in log
  --debug               Show more information in log
  --version             show program's version number and exit
```

Test Analyses - AST testset

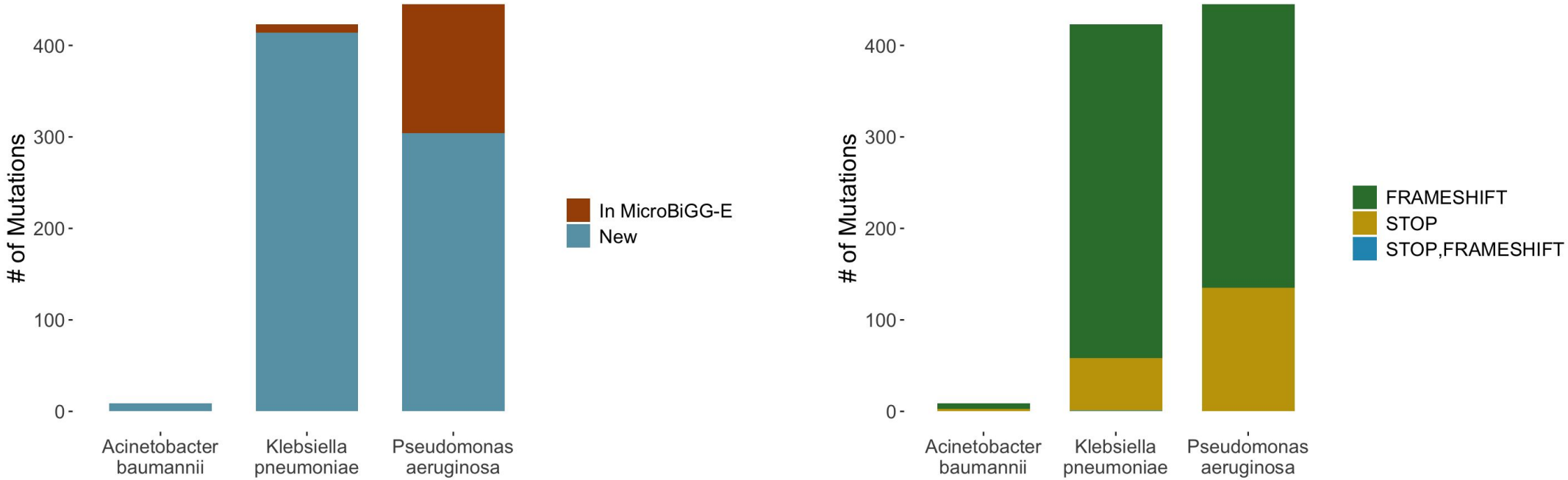
- 928 isolates
  - 589 with resistant phenotype
    - 493 mutations found
      - 130 mutations were present in MicroBIGG-E too
      - 363 new mutations in 353 isolates
  - 339 with sensitive phenotype





Test Analyses - Random testset

- 3000 isolates for *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*
  - DGW identified 877 mutations
    - 150 mutations were present in MicroBIGG-E too
    - 727 new mutations in 353 isolates



Conclusions and Future Directions

- We identified a novel OmpK35 frameshift mutation in isolate CP021955.1/SAMN04014948 which is highly carbapenem resistant and has no carbapenemases
- Compare results from the two tools on a large set of assemblies
- Run the tools on entire NDARO dataset to characterize frequency of these mutations
  - 106,000 *Klebsiella pneumonieae*
  - 39,000 *Pseudomonas aeruginosa*
  - 36,000 *Acinetobacter baumanii*