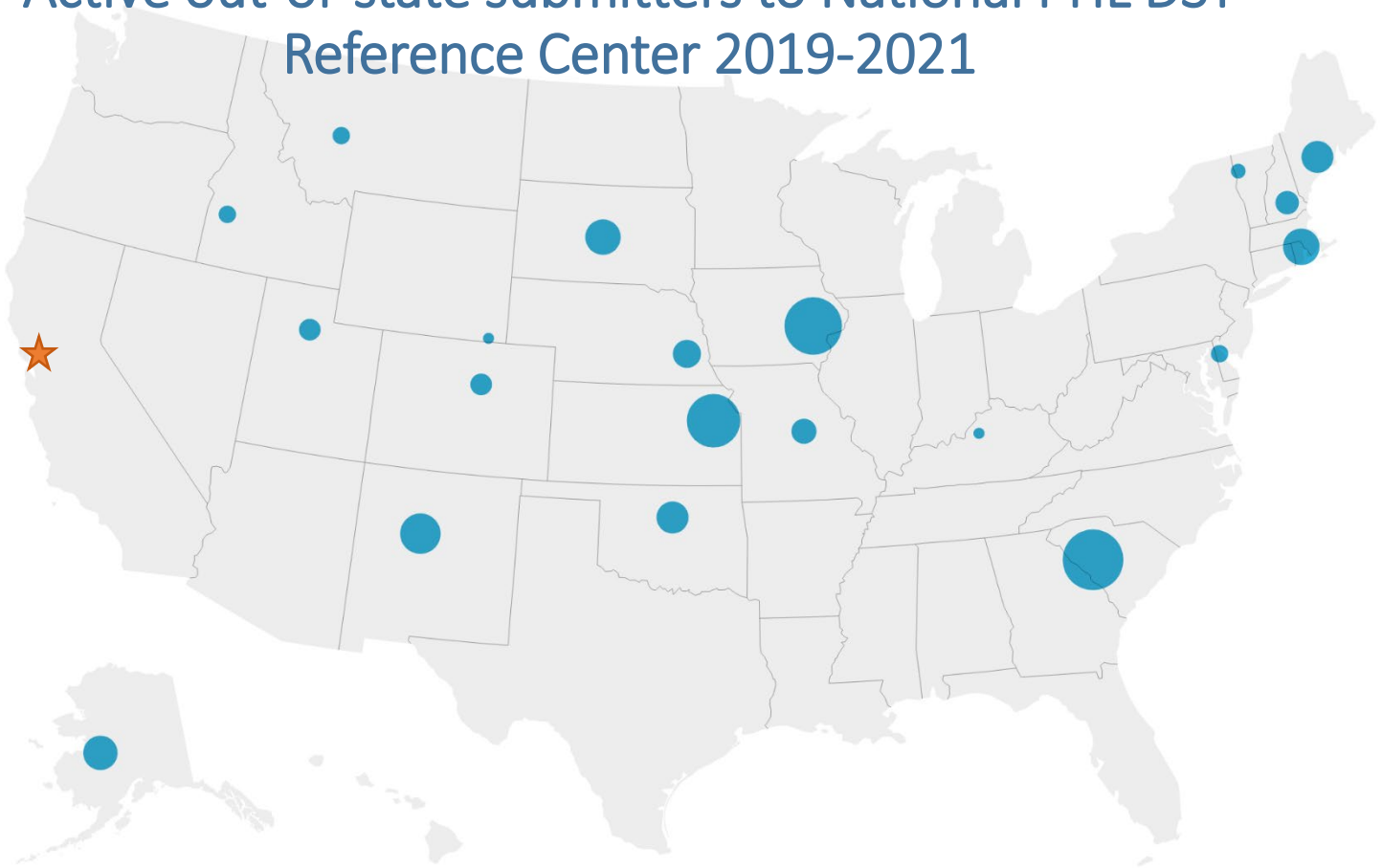
Update on Implementation of NGS-based TB Drug Resistance Detection at the California Dept. of Public Health

Varvara Kozyreva
Mycobacterial, Mycotic, and Parasitic Diseases Section
Microbial Diseases Laboratory Program
California Department of Public Health
✉ varvara.kozyreva@cdph.ca.gov
@varvarakozyreva

APHL 12th National Conference on Laboratory Aspects of Tuberculosis
July 13th, 2022

National PHL DST Reference Center

Active out-of-state submitters to National PHL DST Reference Center 2019-2021

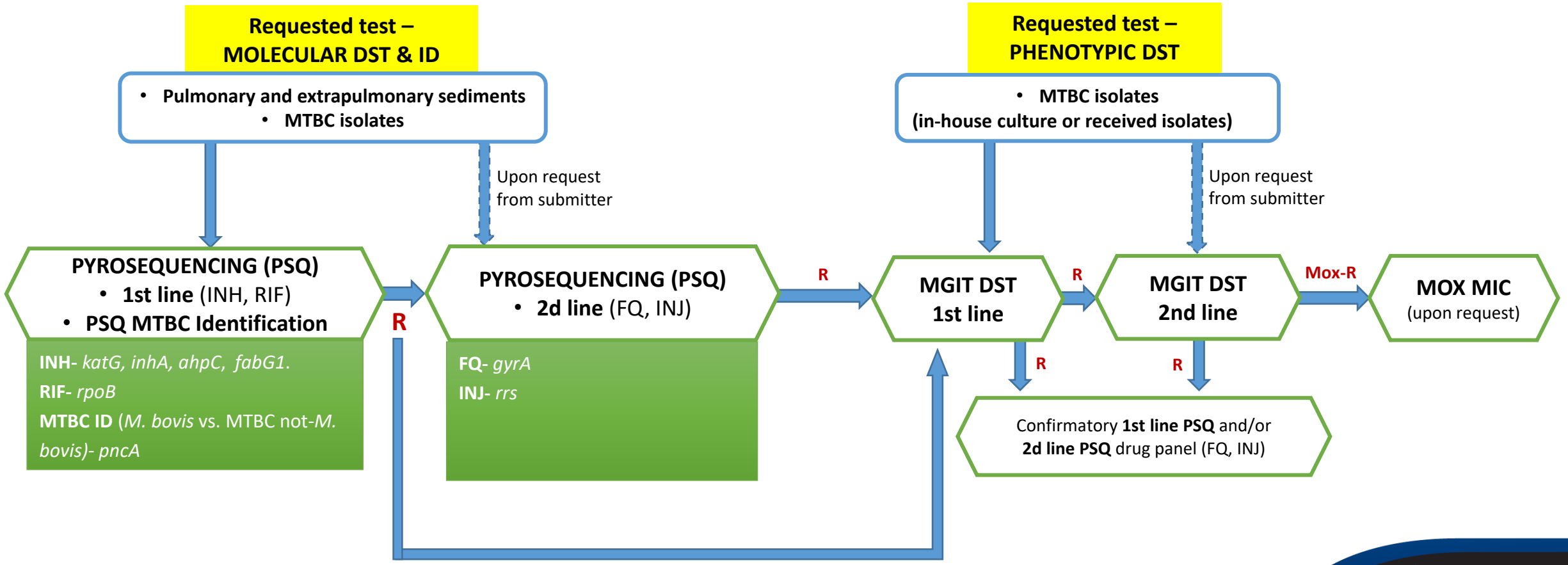


- 20 PHL submitters nation-wide + 30 California submitters

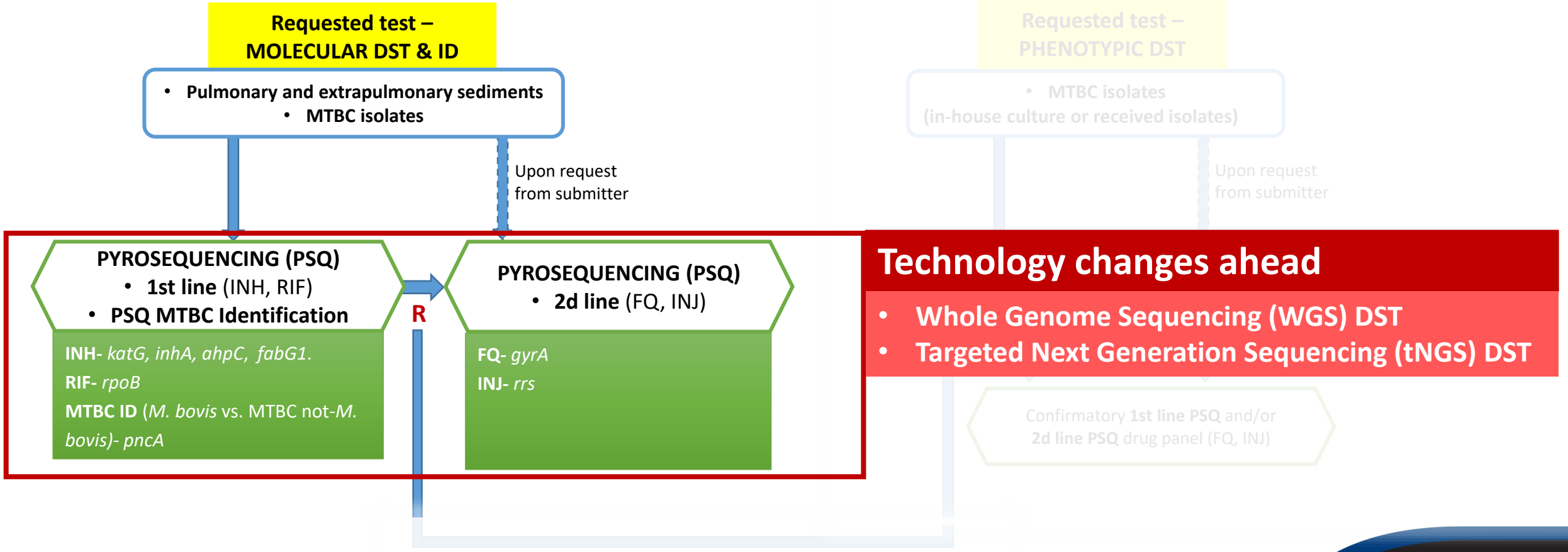
Size of circles corresponds to the average number of annual submissions (test performed) per site

https://datawrapper.dwcdn.net/s3jD6/4/

Current TB Drug Susceptibility Testing (DST) algorithm



Current TB Drug Susceptibility Testing (DST) algorithm



Overview of new & upcoming mDST tests

Targeted Next Generation Sequencing (tNGS) DST

- Purpose: expand mDST panel and replace outdated PSQ technology.

  To be used on processed clinical *specimens* directly & for *culture*
- Approach: Amplicon-based sequencing of specific genomic areas associated with drug resistance (hot spots)

  Illumina sequencing technology
- Status: In process of development

  Expected test roll-out timeline: by the end of 2022

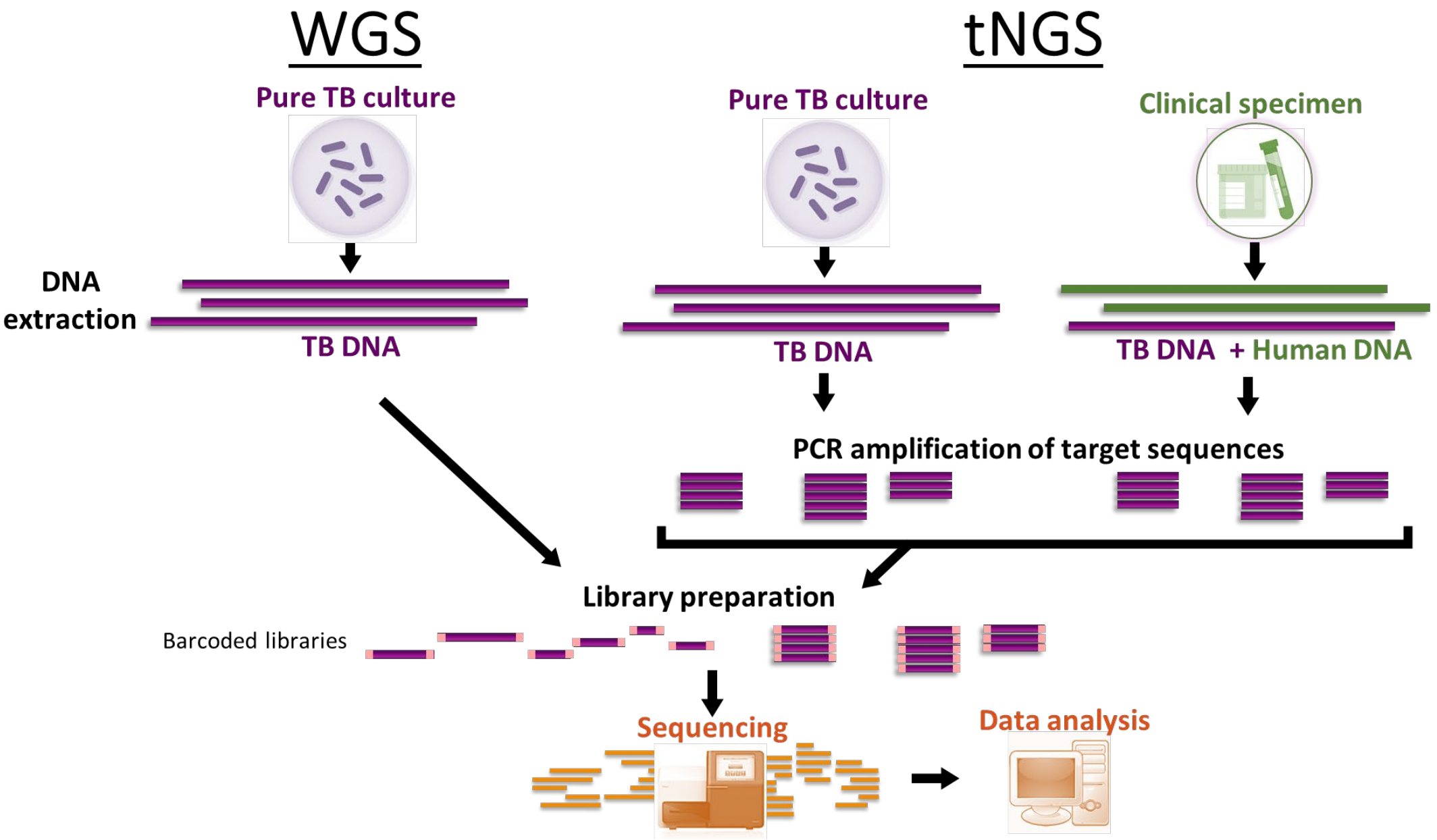
Whole Genome Sequencing (WGS) DST

- Purpose: in-depth mDST characterization from *pure culture*
- Approach : Genome-wide sequencing with consequent analysis of full-length genes of drug targets

  Illumina sequencing technology
- Status: Validation completed, report is being reviewed. LIMS assay building

  Expected test roll-out timeline: Q3 2022

WGS vs. tNGS. Principle



tNGS & WGS mDST targets

Red text- new targets in tNGS not covered by PSQ
Purple text- additional targets in WGS not covered in tNGS

	Drug	PSQ loci	Added tNGS loci	Added WGS loci
1st line	Isoniazid	*katG, fabG1/inhA, ahpC*	*= PSQ*	*= tNGS*
	Rifampicin	*rpoB*	*= PSQ*	*= tNGS*
	Pyrazinamide	*pncA* (only used for *M. bovis* ID in PSQ)	*= PSQ + pncA* (PZA interpretation)	*= tNGS*
	Ethambutol	-	*embB*	*= tNGS*
2nd line	Amikacin/Kanamycin	*rrs*	*= PSQ + eis*	*= tNGS*
	Capreomycin	*rrs*	*= PSQ*	*= tNGS + tlyA*
	Fluoroquinolones	*gyrA*	*= PSQ + gyrB*	*= tNGS*
	Ethionamide	-	-	*ethA*
Other	Bedaquiline	-	*Rv0678 (mmpR), atpE*	*= tNGS + pepQ*
	Linezolid	-	*rplC, rrl*	*= tNGS*
MTBC ID confirmation		*IS6110*	*= PSQ*	- (lineage assignment instead)

WGS DST Validation

Performed by Molecular Characterization Unit MDL:
Dr. Matt Sylvester
Dr. Rituparna Mukhopadhyay
John Crandall
John Bell
Emmanuel Flores

Method:

- DNA extraction: InstaGene DNA extraction from pure culture (evaluated against CTAB & Zymo Quick-DNA HMW MagBead Kit).
- Sequencing: Nextera DNA prep library prep, MiSeq platform, 2 x 500 cycle sequencing runs with V3-600c Illumina sequencing chemistry

Data analysis

- varpipe_wgs v1.0.1 (CDC)
- Both major variants (mutations supported by >50% of reads) and minor variants (<50% reads support) are interpreted
- Only high-confidence variants (WHO) are interpreted at initial assessment (but lower confidence variants with strong association to R phenotype detected in this validation study may be added to the assay later)

Validation Panel- total 194 MTBC cultures (clinical routine isolates, MPEP strains)

- Including 22 *M. bovis* isolates.
- 42 Pan-susceptible isolates.
- 152 Resistant (to one or more drugs) isolates.

Reference methods:

- Phenotypic DST: MGIT in-house testing, Agar Proportion for MPEP tested cultures
- Molecular DST: PSQ in-house testing, Sanger for MPEP & MDDR tested cultures
- *M. bovis* ID: PSQ in-house testing

WGS DST Validation results

➢ Performance of WGS DST for prediction of susceptibility phenotype:

- The WGS DST assay in general showed high concordance with phenotypic DST, though PZA, ETA, and EMB can be problematic.

	INH	RIF	RFB	MOX	AMK	CAP	KAN	PZA	ETA	EMB	Overall
TP	82	17	5	16	2	2	2	7	29	5	167
TN	91	166	93	153	73	72	56	139	100	131	1074
FP	0	0	2	0	0	0	0	0	1	2	5
FN	9	0	0	2	0	0	0	16	8	0	35
Sensitivity	90%	100%	100%	89%	100%	100%	100%	30% (!)	78%	100%	83%
Specificity	100%	100%	98%	100%	100%	100%	100%	100%	99%	98%	100%
Accuracy	95%	100%	98%	99%	100%	100%	100%	90%	93%	99%	97%
PPV	100%	100%	71%	100%	100%	100%	100%	100%	97%	71%	97%
NPV	91%	100%	100%	99%	100%	100%	100%	90%	93%	100%	96%

- Reasons for pDST vs. WGS discrepant results:
  - Low reproducibility of PZA and EMB susceptibility results obtained with MGIT
  - Mutations lacking sufficient data (per WHO) to support their correlation with S/R phenotype (low confidence) [ETA, RFB] - need to be further validated & included into assay.
  - Heteroresistance to MOX due to low frequency *gyrA* mutation missed by WGS

WGS DST Validation results

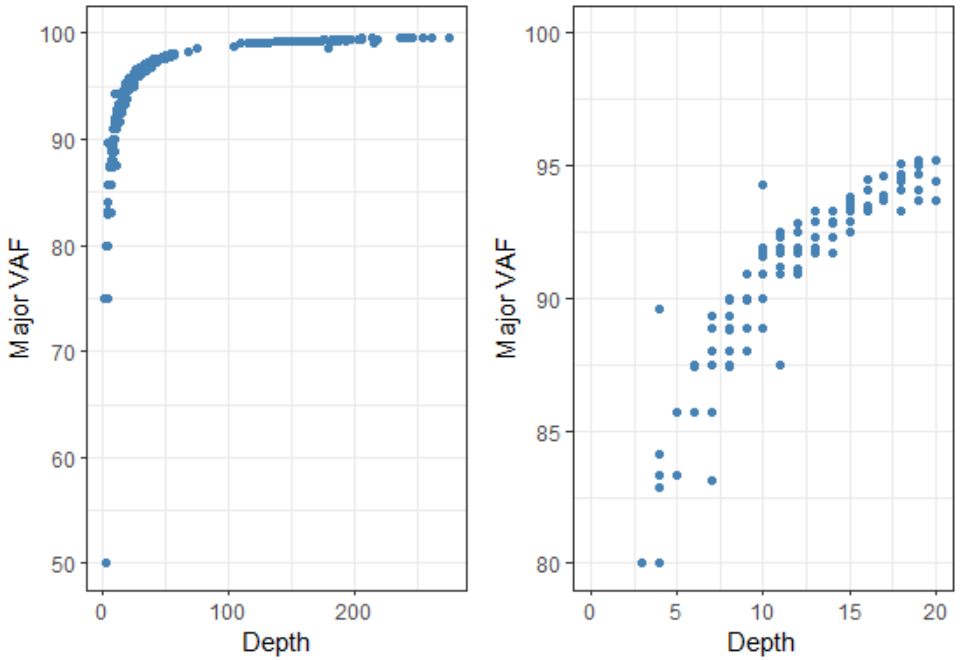
- Concordance of WGS DST with Pyrosequencing:
  - 100% concordance for *major* variants at loci targeted by PSQ.
- Reproducibility of variant detection – 97.8%
  - Discrepancy in *gyrA* allele call in a single sample with low frequency (8.3%) variant present
- Repeatability of variant detection – 100%

WGS DST Validation results

➢Limit of Detection for Major Variants

- **As Biological LOD**: Not applicable since input material for the rxn is standardized
- **As Bioinformatic LOD** (*in silico* downsampling): determined variant allele frequencies (VAF) at different coverage depth values:

Major VAF as a function of loci-specific depth of coverage



- No discrepancies in major allele calls with overall depth of coverage as low as 5x
- Single FP minor frequency allele call with very low read support (~ 3 x) ⇒ Set min allele support threshold to 10 reads (10x)

WGS DST Validation results

➢Heteroresistance (Detection Limit of Minor Variants)

- **As Biological LOD:** extracted gDNA mixed at different ratios to model heterogeneous population
  - Sensitivity of minor variant detection was reduced for the variant supporting coverage < 7x
    ⇒ Supports setting the min allele coverage threshold to 10 reads (10x)
- **As Bioinformatic LOD** (*in silico* downsampling):
  - Different allele frequencies at 10x variant coverage were modeled to predict the required min overall genome coverage:
    - Would require > 100x overall coverage to detect variants with at least 10% frequency
    - With average coverage routinely achieved for samples, the WGS DST assay is not likely to consistently detect variants with frequency < 20%

WGS DST Validation results

➢Cross-reactivity (Contamination)

- **Biological contamination (*in vitro* spike-in):** extracted TB gDNA mixed with potential contaminating organisms' DNA at various ratios
  - NTM: no effect of up to 30% contamination.
  - *Pseudomonas aeruginosa*: tolerance of up to 10% contamination.
- **Bioinformatic mocked contamination** (*in silico* read mix): non-MTBC sequences added to TB WGS reads at different ratios
  - No issues detecting major variants up to 30% contamination.
- Higher level contamination with both Biological and Bioinformatic mocked contamination manifested as pipeline QC failure.

Current progress on tNGS DST – Development stage.

Initial tNGS approach:

➢ Method:

- DNA extraction: Heat-kill DNA extraction from culture or sediment followed by column purification
- Targeted PCR amplification of R-conferring loci - classical PCR design:
  - Primer pairs generating amplicons of 400-800 bp, e.g.:

  - 16 targets (15 drug markers + IS6110 for ID) covered by 22 amplicons
- Sequencing: Nextera XT library prep, iSeq platform, 2 x 150 cycle sequencing runs with V2 Illumina sequencing chemistry

Former Supervisor of DST Unit- Dr. **Linlin Li**, PhD, PHM

➢ Data analysis

- varpipe_amplicons (CDC)

➢ Achieved results: Successful proof of concept for using tNGS as a molecular DST method for cultures & sediments

- Concordance of mutations detected with tNGS vs. PSQ & Sanger
- Great sequencing coverage for isolates
- Good sequencing coverage for 3+/4+ sediments
- Ability to detect heteroresistance

➢ Limitations:

- Worked well as singleplex PCR, but some loci dropped out with duplex PCR
- Suboptimal amplification of certain loci in 1+/2+ sediments
- Attempted Nested PCR design: preliminary success in increasing sensitivity but increased TAT and risk of contamination

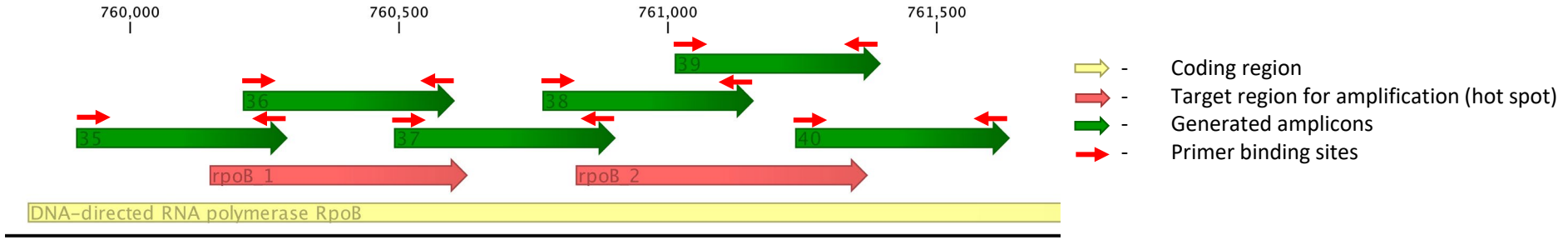
Current progress on tNGS DST – Development stage.

New tNGS approach:

Dr. **Matt Sylvester**, PhD, PHM
DST Unit Supervisor

➢ Method:

- DNA extraction, Library prep, Sequencing, and Data Analysis- the same
- Tiled[1,2] PCR design:
  - Primers designed to produce overlapping amplicons (using PrimalScheme[3] tool), e.g.:



  - Two designs are being tested with amplicon sizes ~400 bp and ~1200 bp
  - Each primer set is tested as 2 pools of primers and 4 pools of primers (to avoid cross-reaction between overlapping primer sets)
  - 16 targets (15 drug markers + IS6110 for ID) covered by either 52 amplicons (400 bp amplicons) or 34 amplicons (1200 bp amplicons)

1. Quick, J., Grubaugh, N., Pullan, S. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* **12,** 1261–1276 (2017). https://doi.org/10.1038/nprot.2017.066
2. Freed NE, Vlková M, Faisal MB, Silander OK. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. Biol Methods Protoc. 2020 Jul 18;5(1):bpaa014. doi: 10.1093/biomethods/bpaa014. PMID: 33029559; PMCID: PMC7454405.
3. https://primalscheme.com/

Current progress on tNGS DST – Development stage.

New tNGS approach. Preliminary results:

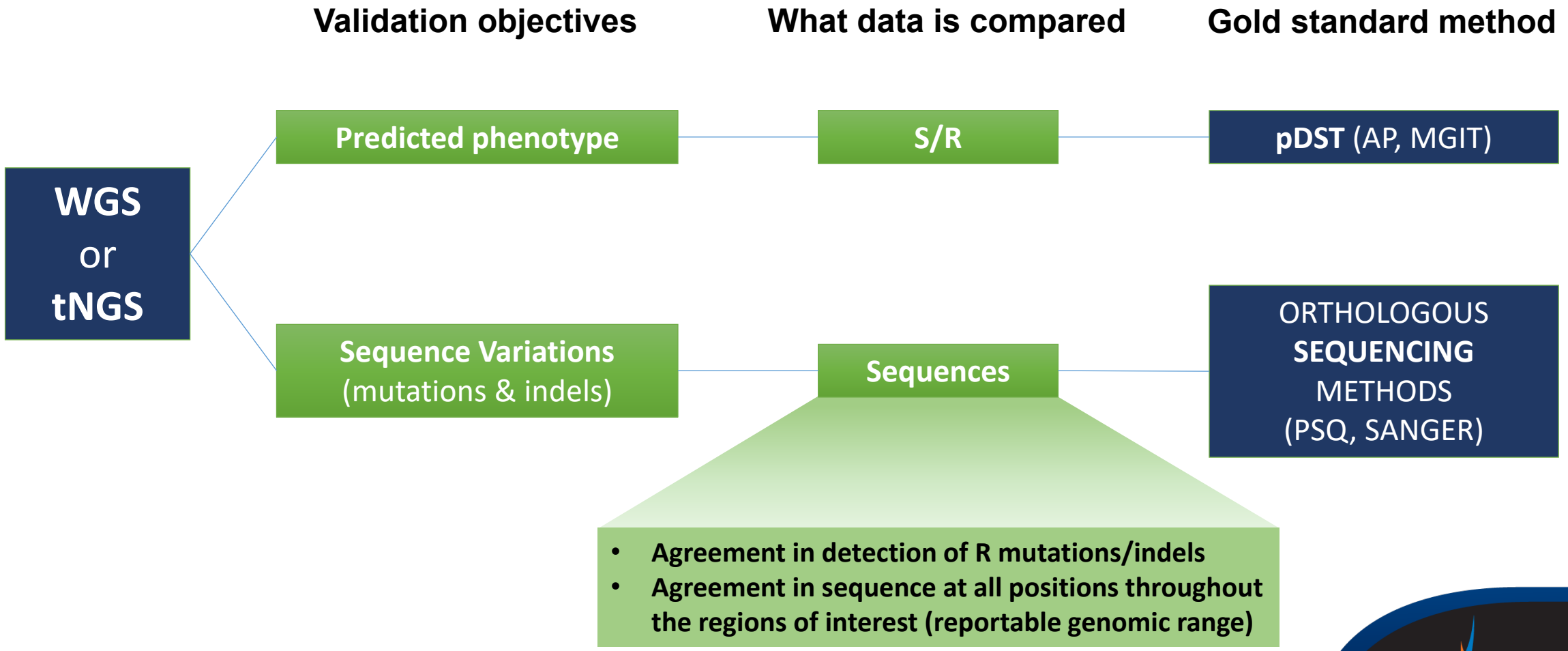
- Successful amplification and sequencing from pure culture and 3+/4+ sediments
  - 100% concordance between mutations detected by PSQ and this tNGS assay for tested isolates and sediments
  - Were able to generate sequences in all targeted areas with range of coverage 11x-6056x (8 samples per 2x150 iSeq run)
  - Both 400 and 1200 bp amplicons in either 2 primer pools or 4 primer pools worked for all samples so far

Current progress on tNGS DST – Development stage.

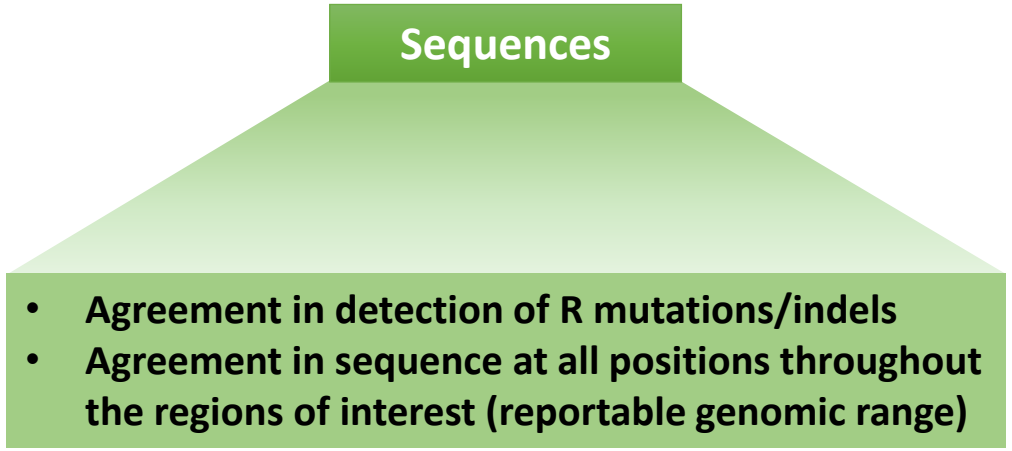
New tNGS approach. Next Steps:

- Testing 1+/2+ sediments
- Determining minimal allele frequency detectable for heteroresistant strains
- Potentially adding more loci outside of currently defined “hot spots” to include mutations with high PPV based on in-house WGS DST validation and WHO Catalogue of mutations
- Formal CLIA validation

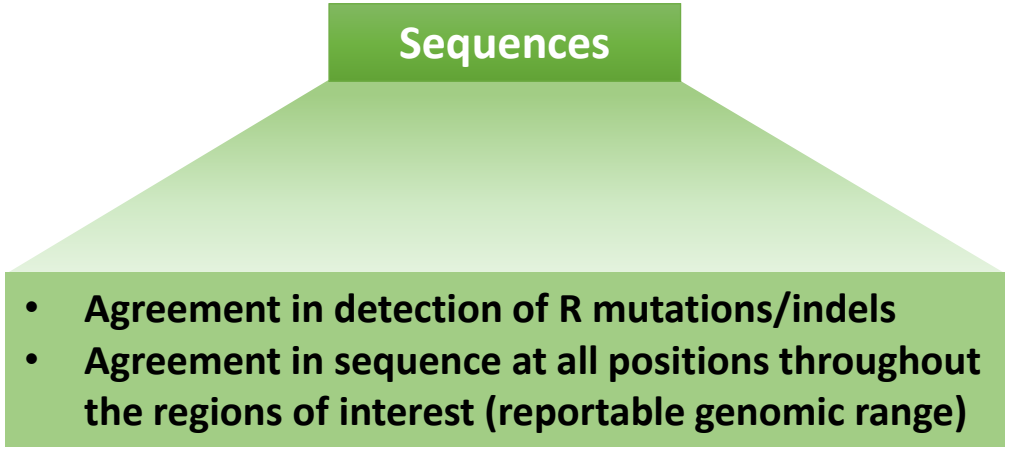
CLIA validation approach for NGS-based DST (WGS & tNGS)



CLIA validation approach for NGS-based DST (WGS & tNGS)



CLIA validation approach for NGS-based DST (WGS & tNGS)



Reasons for including variant detection throughout the targeted genomic regions into the validation:

1. To address the **limited availability of the samples** covering some rare variations that are possible in MTBC genome and can be detected by this assay while not available in current validation set[1].
2. Reporting of **novel genomic variations** associated with resistance discovered in the future[2].
3. Reporting of **silent mutations** (when clinically relevant, e.g., to explain discrepancies with other molecular methods).

Footnotes:

1.*In silico* modelled variations in select loci has been additionally included in validation, but it is not feasible to model variation in every single position where variation may arise in future.

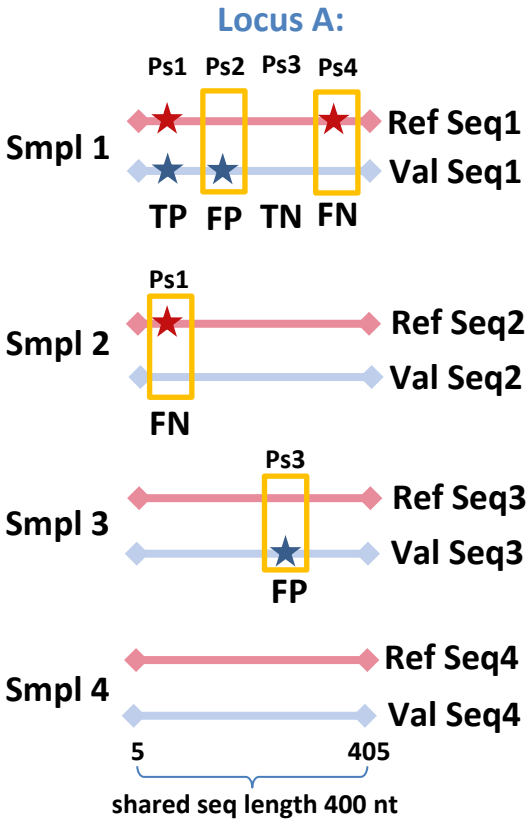
2.It is acceptable to report only those novel mutations that were detected within the genetic regions that have been demonstrated to have high accuracy of variations detection during the validation. Novel sequence variations found in genetic targets responsible for drug resistance in MTBC may be reported after the subject matter expert review, however, uncertainty of clinical effect of such mutations should be clearly conveyed to the submitter in the report. Interpretation of an effect of novel mutations on the phenotype will not be attempted when NGS DST results are reported routinely (additional validation studies will be conducted at the time to compare with pDST results)

CLIA validation approach for NGS-based DST (WGS & tNGS)

Sequences

- **Agreement in detection of R mutations/indels**
- **Agreement in sequence at all positions throughout the regions of interest (reportable genomic range)**

Example of determining accuracy of variant detection across the reportable genomic range:

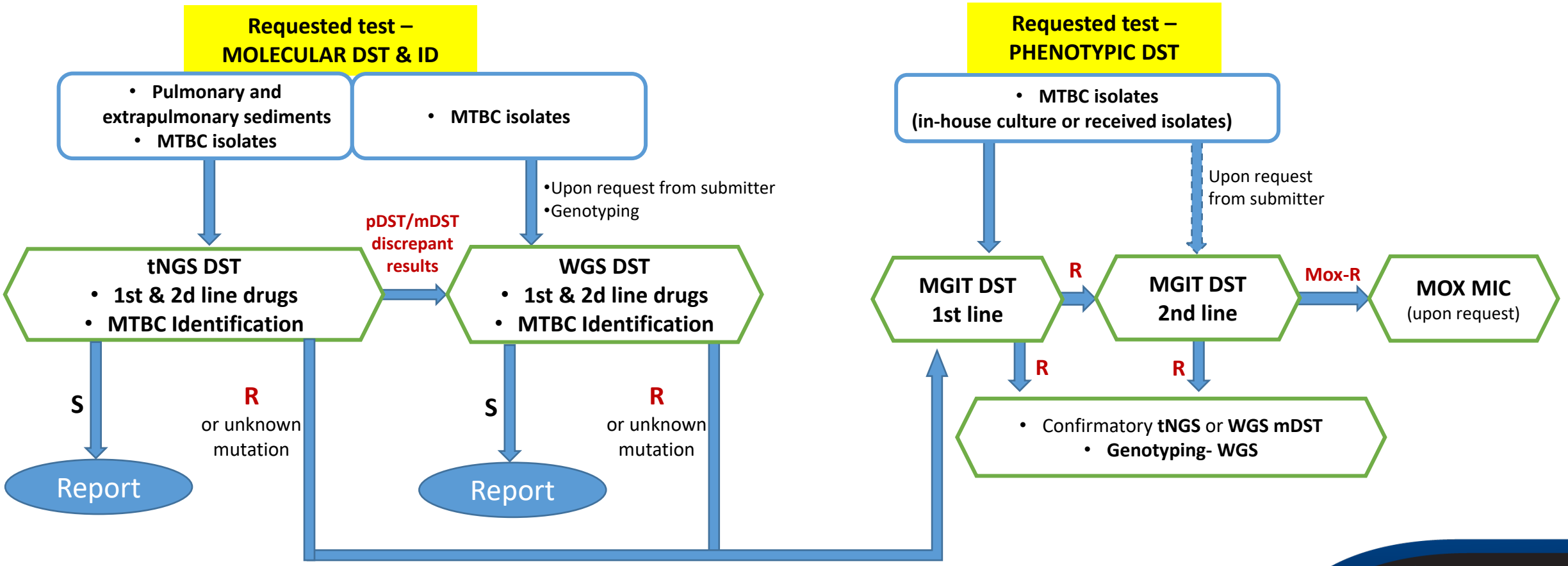


	LcA-Ps1	LcA-Ps2	LcA-Ps3	LcA-Ps4	Rest of LcA
Smpl1	TP	FP	TN	FN	TN
Smpl2	FN	TN	TN	TN	TN
Smpl3	TN	TN	FP	TN	TN
Smpl4	TN	TN	TN	TN	TN
Accuracy per position	1TP + 2TN / 4 samples = 75%	3TN / 4 samples = 75%	3TN/ 4 samples = 75%	3TN / 4 samples = 75%	4TN / 4 samples = 100%

Locus A:
accuracy 100% within reportable range 5-405 nt (exception: positions Ps1, Ps2, Ps3, Ps4 had accuracy 75% which is below the desirable threshold of 95% and will be excluded from reporting)

Conclusion: any genomic variations detected within mentioned above regions can be reported (with corresponding disclaimers regarding known association with resistance or lack of such knowledge).

Proposed TB Drug Susceptibility Testing (DST) algorithm



TURN-AROUND TIME

MON-afternoon | TUE | WED | THR | FRI

tNGS DST

TAT 3.5-4 days

Culture receipt. DNA extraction
- Sample receipt
- DNA extraction

PCR. Library prep D1
- DNA spin-column purification
- Multiplex PCR
- Amplicon Pooling
- DNA QC [optional]
- Library Prep: fragmentation & amplification

Library prep D2. Sequencer run
- Library prep: clean up
- Library QC
- Library pooling

Sequencer run. Analysis. Reporting
- Data analysis
- Reporting (afternoon)

- **Sequencing run o**n iSeq instrument (in 2 x 150 cycles run, 19 hrs run): if loaded on Wed afternoon (overnight run)- done by Thr morning

WGS DST

TAT 4-5 days

Culture receipt. DNA extraction
- Sample receipt
- DNA extraction

Library prep D1
- DNA QC
- Library Prep: fragmentation & amplification
- Library clean up

Library prep D2. Sequencer run
- Library QC
- Library pooling

Sequencer run

Sequencer run. Analysis. Reporting
- Data analysis
- Reporting (morning)

- **Sequencing run** on MiSeq instrument (in 2 x 500 cycles run, 39 hrs run): if loaded on Wed afternoon will be done by morning of Fri

ACKNOWLEDGEMENTS

Microbial Diseases Laboratory CDPH

MDL Laboratory Director
Dr. Zenda Berrada

Mycobacterial, Mycotic and Parasitic Diseases Section, DST Unit
Dr. Linlin Li (former)
Dr. Matthew Sylvester
Katelyn Chen
Tanya Grechukhina
Terry Weber
Wendy Cheung
Steven Yu
Lucy Pham
Grace Lin (former)
Dr. Ed Desmond (former)

Molecular Characterization Unit (former Core lab)
Dr. Rituparna Mukhopadhyay
John Crandall
Angela Lovett
John Bell
Emmanuel Flores

CDPH-TBCB

Dr. Pennan Barry
Dr. Tambi Shaw
& TBCB Team

CDC

Dr. Jamie Posey
Dr. Lauren Cowan
Dr. Angela Starks
Dr. Tracy Dalton
Dr. Stephanie Johnston
MDDR team
MPEP team

APHL
Dr. Anne Gaynor

Local California PHLs

Los Angeles County PHL
San Francisco PHL

University of California Santa Cruz

Balaji Sundararaman
Dr. Russ Corbett-Detig
Dr. Ed Green