

令和5年度 希少感染症診断技術研修会 2024年2月15日

# 結核菌のゲノム利用

御手洗 聡

結核予防会結核研究所抗酸菌部

長崎大学大学院医歯薬学総合研究科基礎抗酸菌症学



## 抗酸菌感染症における背景状況（日本国内）

- 結核

- 国内では漸減しており、2022年時点で低まん延状態にある（8.2/10万人）。
- 患者の主体は依然として高齢者であり、過去の既感染者からの発病であるが、近年では非特異的新規感染からの発病や、海外からの輸入（海外既感染者の発病）が認められ、それぞれに（あるいは相互に）新たな感染伝搬を起こしている。
- Anonymous contactによる多重・多方向感染から、Specific transmissionへの変化→伝搬経路の特定の可能性

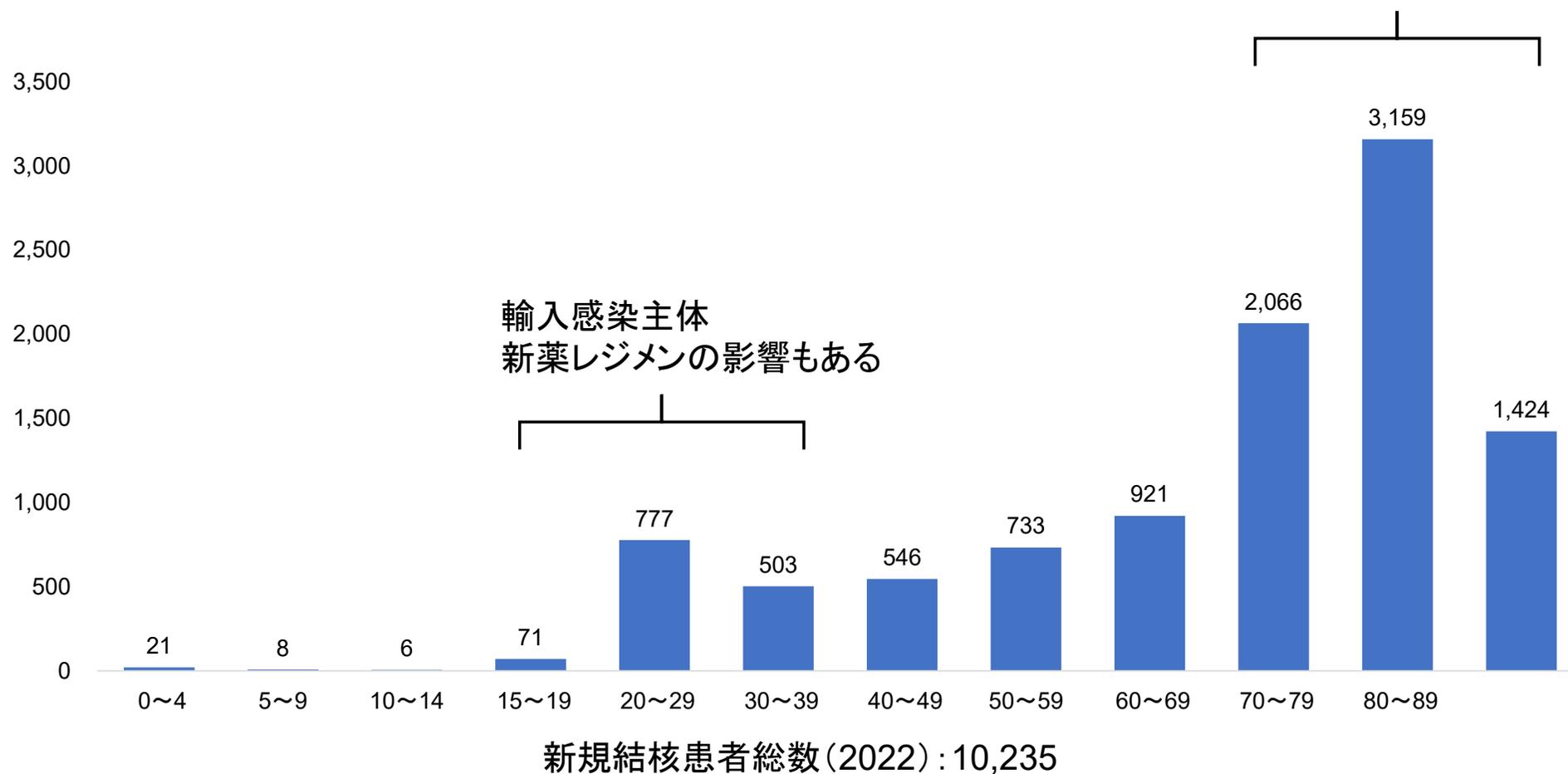
- 非結核性抗酸菌症

- 患者の急速な増加
- 基本的には環境感染だが、ヒト-ヒト感染を疑わせる事例がある。
- 環境からの再感染を起こす場合が多く、感染経路の特定と制御は必要（結核よりもむしろ前向き）。



# 新規結核患者数：2022

感染主体  
70年以降のRFP導入期が反映され始める



# 抗酸菌症診断技術

- 古典的抗酸菌感染症検査技術

- 症状・理学所見等による推定 (Scoring) → 途上国・小児等では現在も
- 細菌学的方法 → 確定診断だが、感度85%程度が現状の限界
  - 抗酸菌塗抹鏡検
  - 抗酸菌培養
  - 遺伝子検査
  - 抗酸菌種同定
  - 薬剤感受性試験
- 免疫学的診断法 → 抗原・抗体の特異性に依存 感染か発病か
  - 抗原免疫反応 (ツベルクリン, IGRA)
  - 抗原検出
  - 抗体検出
- 放射線学的方法 → 高感度だが一般的に特異性の問題あり
- 血液・生化学検査 → 検体によっては高精度 (ADAなど)



# 結核菌ゲノム情報の利用

- 結核の診断
  - 結核菌検出
  - 遺伝子変異検出による薬剤感受性検査
  - 結核菌の毒性評価？
- 結核の分子疫学
  - RFLP
  - Spoligotyping
  - VNTR
  - SNPs analysis



# Whole Genome Sequence

- Use WGS data for species identification
- Average Nucleotide Identity (ANI)
  - ANI is a measure of nucleotide-level genomic similarity between the coding regions of two genomes.
  - ANI  $\geq 95\%$  means identical species.
- Genome-to-Genome Distance Calculator (GGDC)
  - In silico DDH values are determined by identifying regions of high homology between the genome sequences of control and comparison strains on a computer and determining their pairwise distances.
  - GGDC  $\geq 70\%$  means identical species.

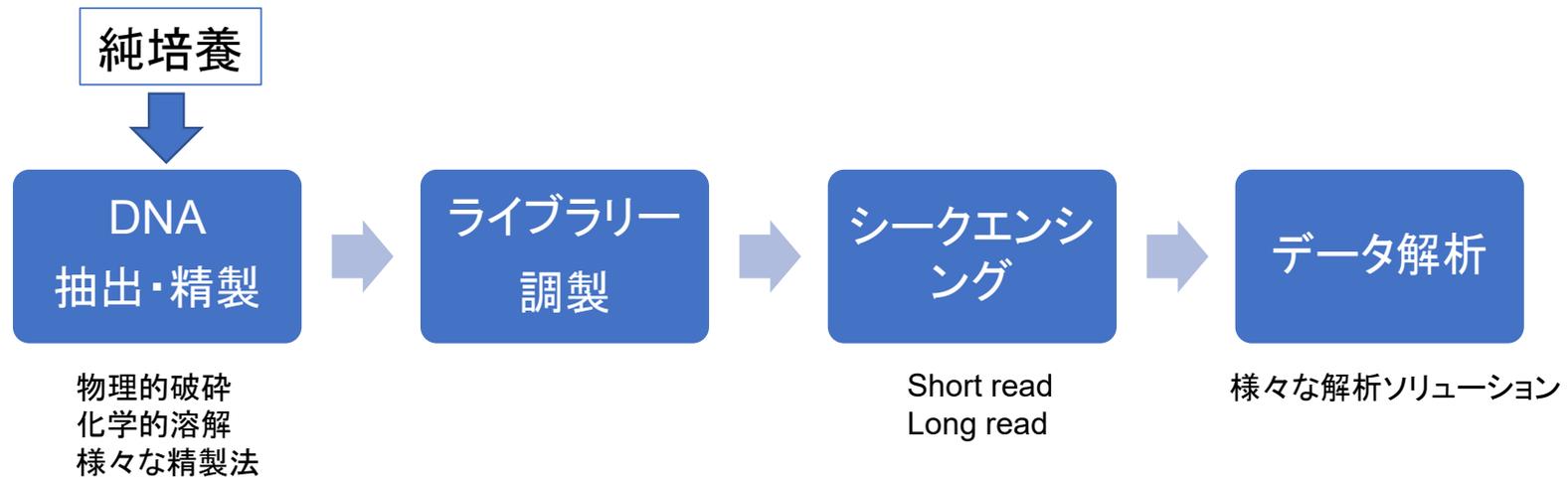


# ANI example

M23061	Species	ANI%
	<i>Mycobacterium heckeshornense</i>	98.9071
	<i>Mycobacterium xenopi</i>	90.3417
	<i>Mycobacterium noviomagense</i>	83.6582
	<i>Mycobacterium kyorinense</i>	80.8522
	<i>Mycobacterium sp. MFM001</i>	80.8462
	<i>Mycobacterium celatum</i>	80.7749
	<i>Mycobacterium fragae</i>	80.3969
	<i>Mycobacterium branderi</i>	80.3405
	<i>Mycobacterium talmoniae</i>	79.8962
	<i>Mycobacterium avium</i> subsp. <i>hominissuis</i>	79.8651



# ゲノムシーケンスのWorkflow



# 結核菌ゲノム解析

- 結核菌ゲノム解析手順書の公開
  - [https://jata.or.jp/dl/pdf/data/TB\\_genom\\_v1.2.pdf](https://jata.or.jp/dl/pdf/data/TB_genom_v1.2.pdf)

Institution, laboratory name 結核研究所抗酸菌部 Location 東京都清瀬市松山3-1-24 Head/Responsible person 村瀬 良朗	SOP (Standard Operating Procedure)  <i>Mycobacterium tuberculosis</i> のゲノム解読に関する 標準作業手順書	Code: Version: no.1.1 Date of release: 2023/06/xx Page: 1 of 7
---	---	--

- 結核菌ゲノム解析実習
  - 結核研究所: 年1回
- 結核菌ゲノム解析外部精度評価
  - 2022年度から年1回希望施設に対して実施



# 原理

## 1.2 ゲノムDNAの調製

- 患者より分離培養された*M. tuberculosis* 株を固形培地で増菌する。クロロホルム存在下でガラスビーズと菌体を激しく攪拌する事により、*M. tuberculosis* のもつ脂質に富んだ厚い細胞壁を効率的に溶解・破碎し、ゲノムDNAを抽出する。RNase A処理、フェノール・クロロホルム処理によりRNA、タンパク質、脂質、糖類等の不純物を除去し、イソプロピルアルコール沈殿によりゲノムDNAを回収し、バッファーに溶解する。
- ※市販のゲノムDNA調製カラムキットなどを用いることもできる。この場合、結核菌の不活化を確実に実施すること。



# クロロホルムビーズ破碎

クロロホルムビーズ破碎法

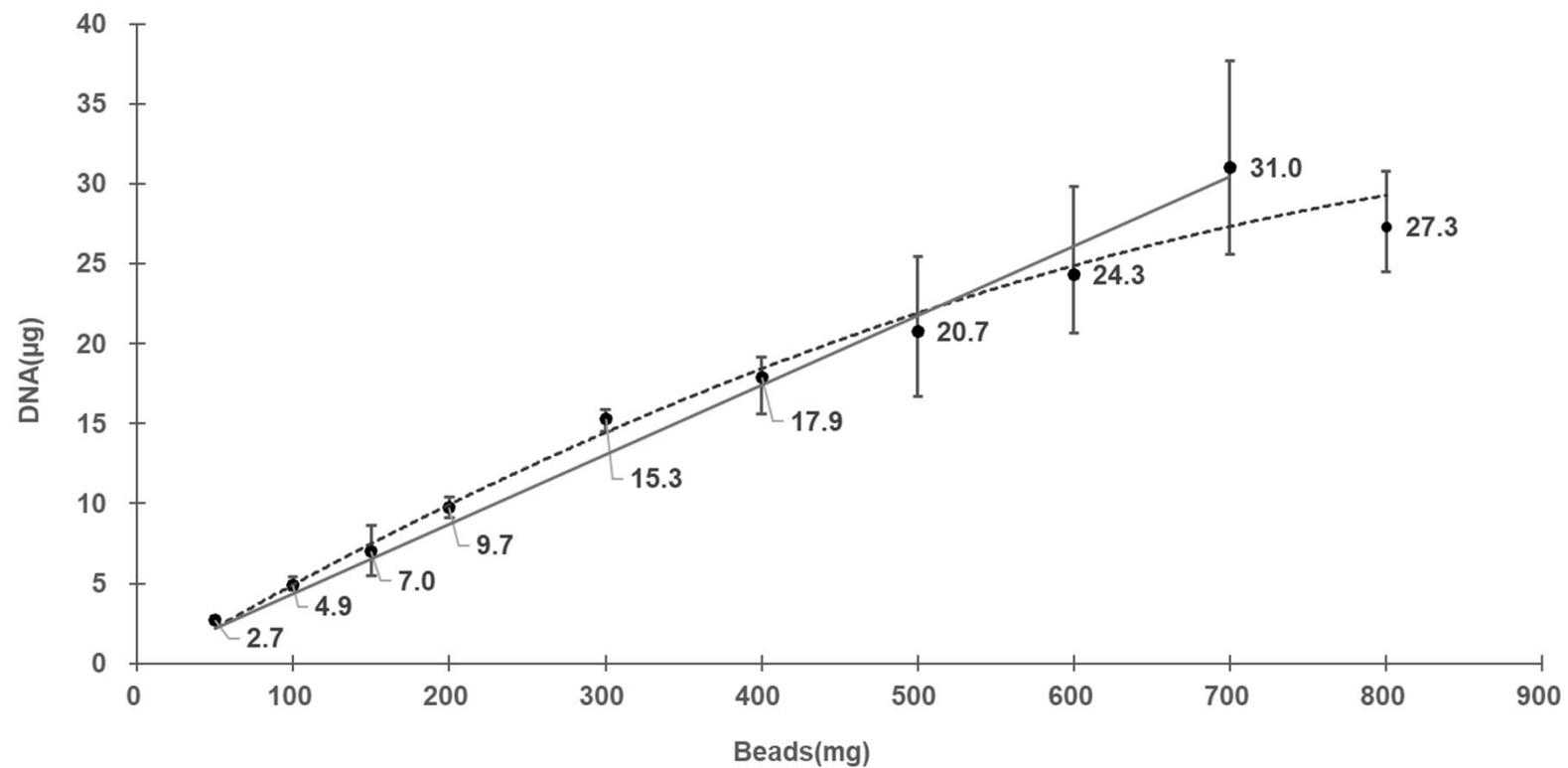
項目	所要時間 min/株	備考	作業領域
NaCl/TE/クロロホルム+菌体	5		BSL3
粉碎 Voltex	5	Voltex 2-7分	BSL3
遠心	10		BSL2以下
RNaseA処理	22	インキュベート20分	BSL2以下
PCIAA抽出	15	PCIAA 遠心5分x2	BSL2以下
クロロホルム処理	12	クロロホルム遠心10分	BSL2以下
エタノール沈殿	20	遠心15分	BSL2以下
<b>合計</b>	<b>89</b>		

(従来法) ISOPLANT抽出+磁気ビーズ精製

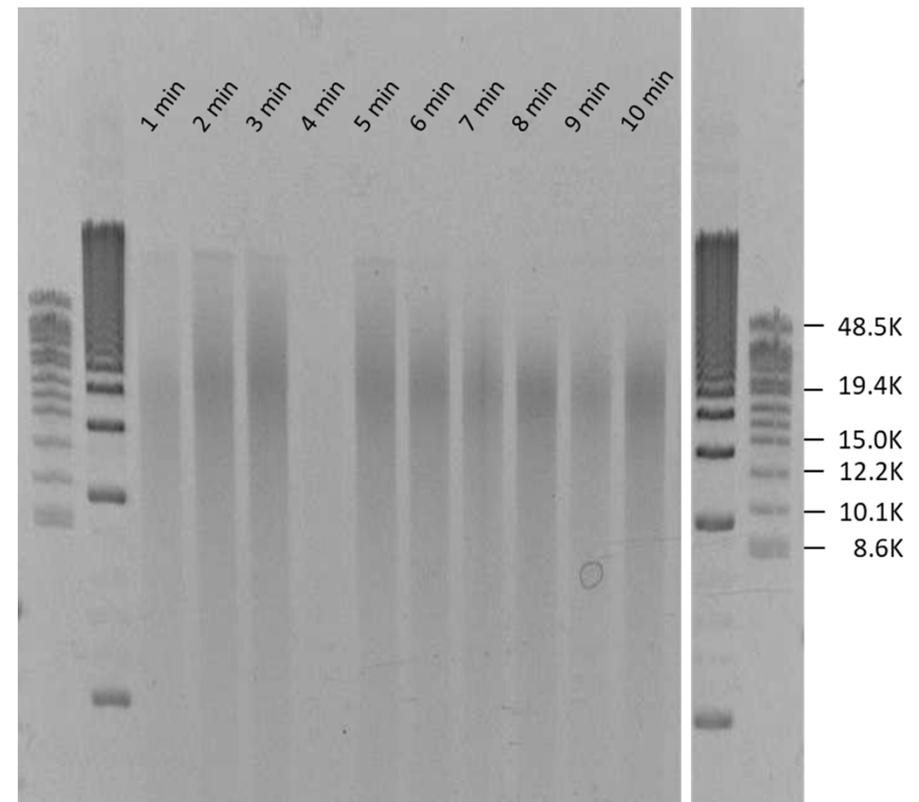
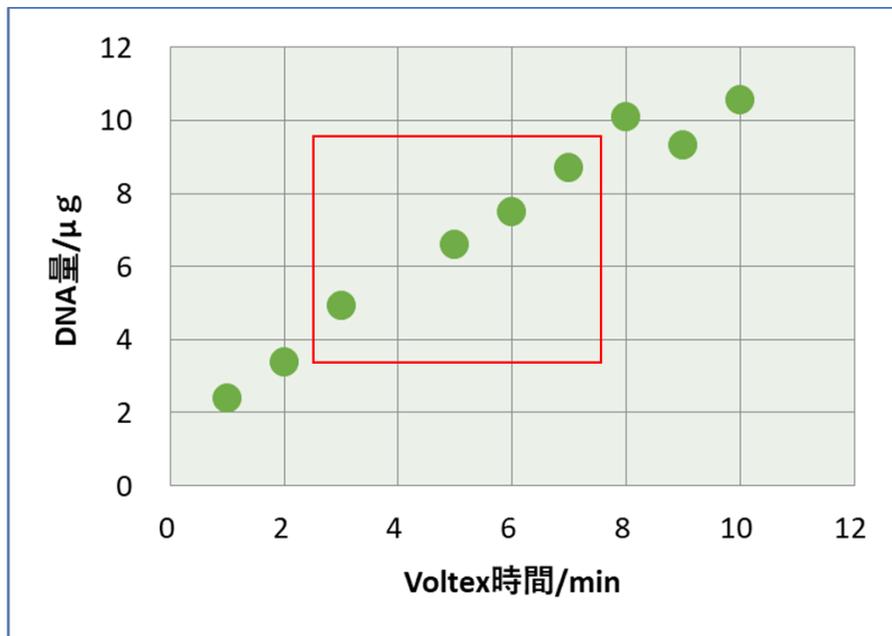
項目	所要時間 min/株	作業時間 min/株	備考	作業領域
I液作業(SDS処理)	5	5		BSL3
1~7日 37°C保管	1440		24Hとして計算	BSL3
80°C インキュベート	20	20	80°Cインキュベート 20分	BSL3
II液作業(塩化ベンジル)	5	5		BSL3
Voltex+55°Cインキュベート	40	40	Voltex 2分x2、インキュベート15分x2、冷却1分	BSL2以下
III液作業(酢酸アンモニウム)	12	12	添加後 氷上インキュベート 10分	BSL2以下
遠心	15	15		BSL2以下
エタノール沈殿	15	15	遠心10分	BSL2以下
磁気ビーズ精製	30	30	Sera-pure	BSL2以下
<b>合計</b>	<b>1582</b>	<b>142</b>		



# ビーズ量とDNA収量の相関関係



# Voltex 時間と回収DNAの分子量



# ライブラリー調製

- QIAseq FX DNA Library Kitを用い、抽出したゲノムDNAから、iSeq、MiniSeq、MiSeq、NextSeqなどのイルミナシーケンサーで解析可能な分子構造を持つDNA断片 (ライブラリー)を調製する。
- キットに含まれるアダプターを用い、最大96株のゲノムDNAを株ごとに異なるDNA配列 (インデックス)で標識したライブラリーを同時に調製する。
- シーケンスの品質を高めるため、アガロースゲル抽出により、適切なサイズのライブラリーを精製する。



# 機材

- 恒温器(インキュベーター): M. tuberculosis培養用
- VORTEX-GENIE 2 (Electro Scientific Industries)
- TurboMix™(Electro Scientific Industries): VORTEX-GENIE 2マイクロチューブ用アタッチメント
- 小型遠心機: 各種
- 小型インキュベーター: MULTI-SHAKER OVEN HB (TAITEC) 等、37°C , 65°Cの設定が可能なもの
- 微量分光光度計 (Nano Drop One [Thermo Fisher Scientific]等)
- Qubit® 4 Fluorometer (Thermo Fisher Scientific) あるいは同等品
- 8連チューブ用あるいは96wellプレート用アルミブロック
- AMPure XP精製用磁石スタンド(Fast Gene Magna Stand等)
- サーマルサイクラー(ヒートリッドの温度制御[切、67°C、105°C]が可能なもの)
- 電気泳動装置: Mupid®-exU (株式会社ミューピッド) あるいは同等品
- 振とう機: アガロースゲルの染色に使用可能なもの
- LEDトランスイルミネーター: FAS-V (日本ジェネティクス) あるいは同等品
- イルミナシーケンサー: iSeq、MiniSeq、MiSeqのいずれか (illumina)



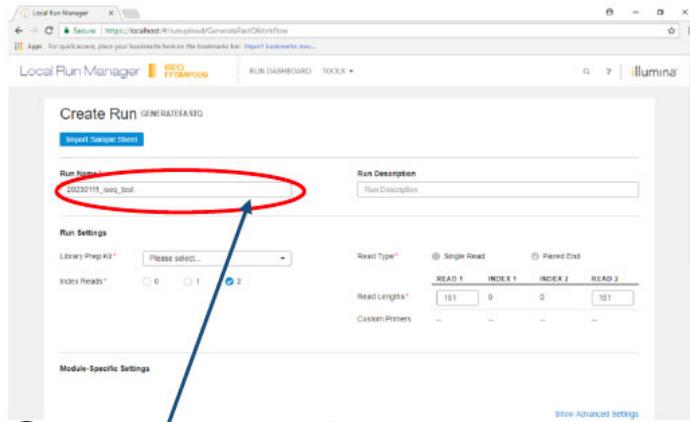
## シーケンサーと主な試薬キット毎の抗酸菌全ゲノム解析可能株数

解析可能株数の目安→	解析株数 上限の目安	データ出力量	出力データに占めるQ30以上の割合	定価(2022年6月)
iSeq 300 Cycle	4	≥1.2 Gb	≥80%	¥110,400
MiniSeq 300 Cycle Mid	10	≥2.4 Gb	≥80%	¥110,300
MiniSeq 300 Cycle High	30	≥7.5 Gb	≥80%	¥309,400
MiSeq v2 300 Cycle Nano	1	≥300 Mb	≥80%	¥58,300
MiSeq v2 300 Cycle Micro	5	≥1.2 Gb	≥80%	¥87,900
MiSeq v2 300 Cycle	18	≥4.5 Gb	≥80%	¥208,900
MiSeq v3 600 Cycle	36	≥15 Gb	≥75%	¥305,800

結核菌ゲノム解読に関する標準作業手順書  
5.5 illuminaシーケンサーのラン

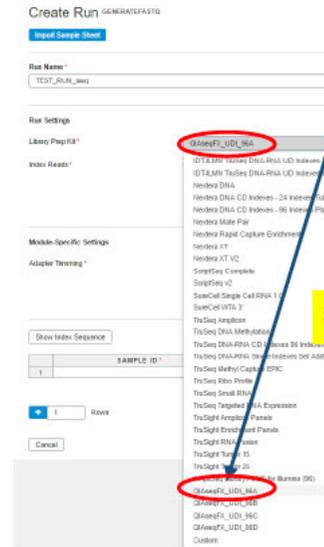
**(追補版) ライブラリー調製とシーケンサー操作ガイド\_iSeq版v1.0**





⑤ Run Nameの入力

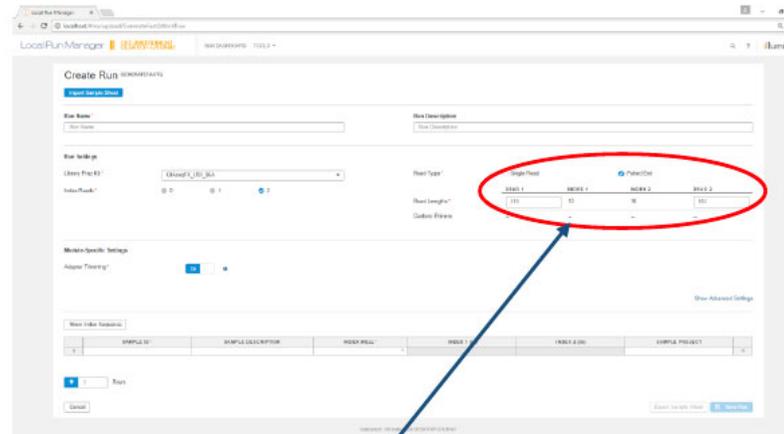
「RUN日+ライブラリー名」など一定のルールでRUNネームを付与するとよい



⑥ Library Prep Kit の選択

QIaseq FX DNA Library UDI Kit(24)を使用した場合は、QIaseqFX\_UDI\_96Aを選択

\* あらかじめQIaseq FXテンプレートをインポートしておく必要がある (p.16-17参照)



⑦ Run 条件の入力

Read Type 「Paired End」

Read1&2 「151」 Index 1&2 「10」



# ゲノム取り扱い上の注意点

- Illumina
  - Short readを基礎としており、depthを深く取ることによってシーケンス精度は向上するが、繰り返し配列のシーケンスに弱い。
  - 殆どIlluminaでデータ解析しているが、本当に正確か？
- Nanopore
- PacBio
  - Long readを基礎としており、PE/PPEなどの繰り返し配列領域の読み取りにも強い。



# Portable sequencing of *Mycobacterium tuberculosis* for clinical and epidemiological applications

Paula J. Gómez-González , Susana Campino , Jody E. Phelan<sup>†</sup> and Taane G. Clark <sup>†</sup>

Corresponding author: Taane Clark, Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine (LSHTM), Keppel Street, London, UK. E-mail: [taane.clark@lshtm.ac.uk](mailto:taane.clark@lshtm.ac.uk)

<sup>†</sup>Jody E. Phelan and Taane G. Clark contributed equally to this work.

## Abstract

With >1 million associated deaths in 2020, human tuberculosis (TB) caused by the bacteria *Mycobacterium tuberculosis* remains one of the deadliest infectious diseases. A plethora of genomic tools and bioinformatics pipelines have become available in recent years to assist the whole genome sequencing of *M. tuberculosis*. The Oxford Nanopore Technologies (ONT) portable sequencer is a promising platform for cost-effective application in clinics, including personalizing treatment through detection of drug resistance-associated mutations, or in the field, to assist epidemiological and transmission investigations. In this study, we performed a comparison of 10 clinical isolates with DNA sequenced on both long-read ONT and (gold standard) short-read Illumina HiSeq platforms. Our analysis demonstrates the robustness of the ONT variant calling for single nucleotide polymorphisms, despite the high error rate. Moreover, because of improved coverage in repetitive regions where short sequencing reads fail to align accurately, ONT data analysis can incorporate additional regions of the genome usually excluded (e.g. *pe/ppc* genes). The resulting extra resolution can improve the characterization of transmission clusters and dynamics based on inferring closely related isolates. High concordance in variants in loci associated with drug resistance supports its use for the rapid detection of resistant mutations. Overall, ONT sequencing is a promising tool for TB genomic investigations, particularly to inform clinical and surveillance decision-making to reduce the disease burden.

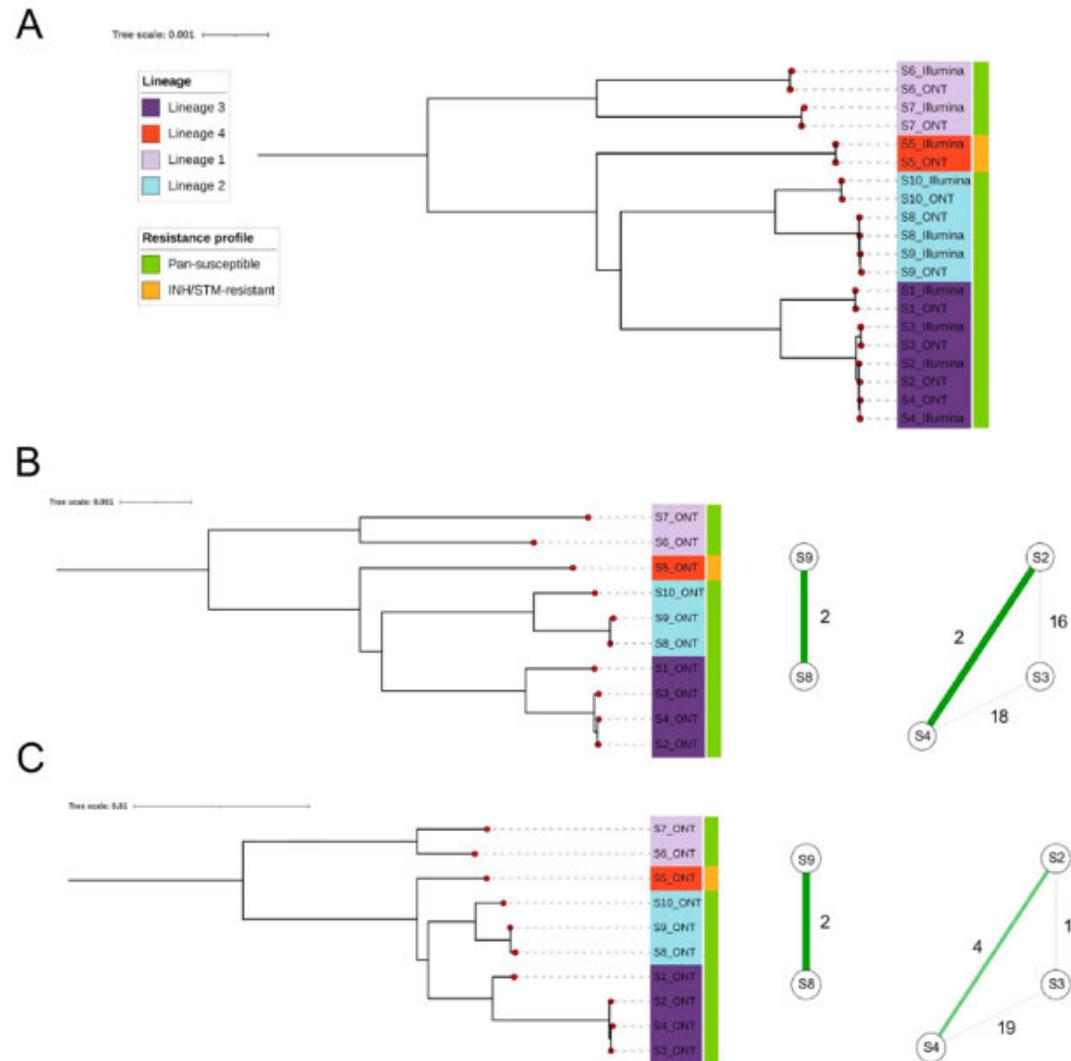
**Keywords:** *Mycobacterium tuberculosis*, tuberculosis, sequencing, genomics, mutations



**Table 2.** Concordance of variants found using Illumina and Oxford Nanopore Technology (ONT) platforms

Sample Pair	SNPs			Small Indels			Large Structural Variants <sup>a</sup>		
	ONT only	Illumina only	Both	ONT only	Illumina only	Both	ONT only	Illumina only	Both
S1	0	2	1144	3	4	94	58	9	20
S2	0	3	1154	3	8	88	64	6	17
S3	0	2	1158	5	7	84	62	6	14
S4	0	2	1156	4	7	84	66	4	14
S5	0	0	259	2	2	28	14	0	4
S6	0	5	1741	0	9	115	67	5	20
S7	0	9	1763	3	5	108	61	6	19
S8	0	4	1154	3	6	97	68	8	14
S9	0	4	1154	2	7	95	67	9	16
S10	1	5	1114	2	5	97	72	12	18

<sup>a</sup>Includes indels > 15 bp.



**Figure 2.** Phylogenetic trees and transmission networks. Maximum likelihood phylogenetic trees and transmission networks for the samples studied. Isolates are coloured by lineage. Drug-resistance profile obtained by phenotypic drug susceptibility testing is shown in the strip labels on the trees. **(A)** Phylogenetic tree reveals high degree of concordance and clustering of replicates sequenced using ONT and Illumina platforms, reconstructed with 3955 SNPs excluding genomic repetitive regions. **(B)** Phylogenetic tree of ONT sequenced samples using the 3955 SNPs, as well as transmission networks for lineage L2 (S8 and S9) and L3 (S2, S3 and S4) clusters showing SNP distances. **(C)** Phylogenetic tree of ONT sequenced samples using the 3955 SNPs in addition to 568 more polymorphic sites located in *pe/ppe* genes, as well as transmission networks for lineage L2 (S8 and S9) and L3 (S2, S3 and S4) clusters with SNP distances.

# ゲノム分子疫学の問題点

- ゲノムの一致・不一致のカットオフをいかにするか？
  - SNV  $\leq 5$  or  $\leq 12$  ?
- 得られたデータを如何にして還元するか
  - VNTRのような簡単なプロファイルではない
  - Fastqファイルのデータは膨大で、専用ソリューションとバイオインフォマティクスの知識が無いと解析できない
- 疫学情報とゲノム情報を統合するソリューションが必要（視覚的にわかりやすいことが重要）
  - 感染研ゲノムセンターが可視化ソフト”Pomegraph”



Thank you.

