

第94回総会教育講演

結核分子疫学研究における全ゲノム解析の役割

1.2 瀧井 猛将

要旨：次世代シーケンシング (next generation sequencing, NGS) の技術革新により結核菌の遺伝情報解析が詳細に行えるようになってきた。全ゲノム (whole genome sequencing, WGS) 解析法では分子疫学研究で重要なクラスターの細分化や菌の関連性の解析の能力が従来法に比較して格段に高く、さらに薬剤耐性予測も可能である。本稿では結核の分子疫学における WGS 解析の役割について概説する。

キーワード：結核, 分子疫学, 全ゲノム解析

1. はじめに

結核は、多くは潜在性結核患者であり、5~15%が発症する¹⁾。致死率50%、適切な治療により95%が治癒するが、5%は再発・再燃する²⁾。結核の発症時期は数カ月から数十年後まで様々であるため、感染時期や感染の場所の特定が難しい³⁾。接触者健診で行うツベルクリン反応や interferon- γ release assays (IGRA) 検査では感染の有無の判別は可能であるが、同一の菌による感染の証明は難しい。そのため、感染源の特定、感染経路の解明など結核の伝播の実像を把握するためには、菌の型別情報を活用することは有用である。

2. 結核の分子疫学研究に用いられている遺伝子型別法

結核菌の遺伝子型別法では、図1に示した遺伝子配列を利用して株の異同を判別する⁴⁾⁵⁾。手技の詳細については他の成書等を参考にされたい。

(1) トランスポゾン IS6110 を利用した Restriction Fragment Length Polymorphism (RFLP) 法

この方法は、菌のゲノム中の任意の位置に挿入されているトランスポゾン IS6110 を解析することで、菌株の異同を判別する手法であり、長年遺伝子型別法として用いられてきた。原理は IS6110 遺伝子をプローブとしたサザンブロット法であり、株ごとに IS6110 を含む

DNA 断片数や移動度が異なることから異同が判別できる。長所としては視認性が高く、クラスター分解能がスポリゴタイピング法 (後述) より優れている。短所としては、結果の取得に1カ月程度を要すること、他の泳動との比較やデータベース化が難しいことが挙げられる。

(2) 反復配列を利用した Variable Number of Tandem Repeats (VNTR) 法

この手法は菌のゲノムのミニサテライト領域中の VNTR 領域の違いを利用して菌の異同を判別する手法である (図1)。VNTR 法では12~24カ所の領域を対象にして、株間で各領域での反復数の一致、不一致から菌株の異同を判別する。手技としては、ポリメラーゼ連鎖反応 (polymerase chain reaction, PCR) 法による VNTR 領域の増幅とアガロースゲル電気泳動法等を用いた増幅 DNA 断片長の測定からなる。1単位の反復配列長 (既知) を用いて測定長から反復数を算出する。長所としては、DNA が少量で解析できることから1週間程度で結果が得られる。結果の数値化が可能であり、過去の検査結果や他施設での結果と比較が容易なため遺伝子型別法として多用されている。一方、クラスター解析の分解能は検査領域数に依存するため、(例えば12領域など) 検査領域数が少ない場合には、疫学的な関係性がなくとも遺伝子型別結果が一致することがある。また、他の VNTR 検査と比較する際には検査対象の VNTR 領域を揃える必要がある。

¹公益財団法人結核予防会結核研究所抗酸菌部, ²名古屋市立大学大学院薬学研究科

連絡先：瀧井猛将, 公益財団法人結核予防会結核研究所, 〒204-8533 東京都清瀬市松山3-1-24

(E-mail: t-takii@jata.or.jp)

(Received 26 Aug. 2019)

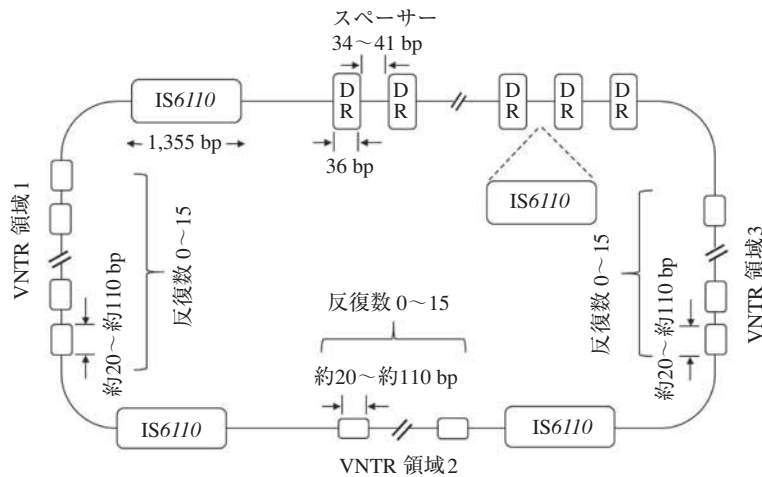


図1 結核菌のゲノム中の特徴のある遺伝子

IS6110: トランスポゾン, DR: Direct Repeat, VNTR: Variable Numbers of Tandem Repeats

(3) Direct Repeat (DR) 配列の有無を利用したスポリゴタイピング法

この手法では43カ所のDR配列のスペーサー領域を標的としたドットプロット法でDR領域有無を検出する(図1)。長所としては、菌数がきわめて少量でも検査可能であり、検査対象領域が国際的に共通であるため互換性も高く、菌の分離地域とスポリゴタイピングの情報を収載したデータベース(SITVIT WEB, http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/)がネット環境上で利用可能である。スポリゴタイピング法は結核菌の地理的分布や系統分類に優れた解析能力を示すが、同系統の菌を対象としたクラスター解析では一致例が多く、RFLP法、VNTR法と比較して劣る。

(4) 次世代シーケンシング(next generation sequencing, NGS)を利用した全ゲノム(whole genome sequencing, WGS)法

WGS法は、結核菌株の系統解析、クラスター解析、株間の関連性解析に加えて薬剤耐性予測等も同時に行える。解析は、検体からのNGSによる結核菌全ゲノム遺伝子配列の取得とバイオインフォマティクスソフトを用いたデータ解析の2つのステップで構成されている(図2)。

例として、われわれはIllumina社のshort-read型のNGS器機であるMiSeqを用いている。MiSeq ver. 3の試薬を使用した場合、1リード(取得した遺伝子配列)長が300 bp×2で約4400万リードを取得できる。計算上では約70株の結核菌を平均100カバレッジ(遺伝子当たりのリードの数)で情報を取得できるが、約700株の結核菌を解析した研究では、1回のNGSに24株を用いて平均カバレッジ85であった⁹⁾。取得したゲノム配列はfastqファイルとして共用が可能である。続のステップでは取得した

遺伝子配列情報を参照配列である結核菌 *Mycobacterium tuberculosis* H₃₇Rv株⁷⁾に当てはめることから始める。H₃₇Rv株の配列情報はNCBI(GenBank: AL123456.3)からダウンロード可能であり、バイオインフォマティクスソフトで解析する。われわれはTGS-TB⁸⁾(<https://gph.niid.go.jp/tgs-tb/>)を用いているが、他にもweb上で利用可能なソフトや⁹⁾、より詳細に設定ができる有償のソフトもある。いずれのソフトにおいてもゲノム中の一塩基多型(single nucleotide polymorphisms, SNPs)、挿入・欠失(Short Insertion/Deletion, InDel)を利用して解析を行う。WGS解析は、結核菌の系統解析、薬剤耐性遺伝子変異の検出、SNPs変異箇所の比較による関連性の解析の他、*in silico*でのスポリゴタイピング、IS6110の検出、一部のVNTRの検出も行える⁸⁾。WGS解析では、VNTR法で同一のクラスターである株間でも数十から数百のSNPsの違いを検出できることから¹⁰⁾、クラスターをさらに細分化できる¹¹⁾。また、SNPsの違いから菌の関連性を詳細に解析できることから、南アフリカで行われた治験の再発事例では、再燃、再感染、複合感染の解析に用いられている¹²⁾。一方、WGSによる薬剤感受性の予測については、薬剤感受性試験の結果と比較して感度が低い薬剤もあり⁹⁾、開発の余地が残されている。今後、まだ見つかっていない遺伝子変異による耐性化の情報が蓄積されることで、包括的で精度の高い方法としての利用が期待できる。

WGS法での短所としては、他の手法に較べて解析にかかる費用が高額であることが挙げられるが、NGS器機や試薬の価格も廉価になってきたことから、今後、全ゲノム情報を用いた分子疫学研究が主流になる。実際、2019年に報告によると2016年時点で、26のヨーロッパ諸国のうち20カ国が遺伝子型別法を導入しており、10カ国のヨーロッパ諸国が遺伝子型別法にWGSを取り入れて

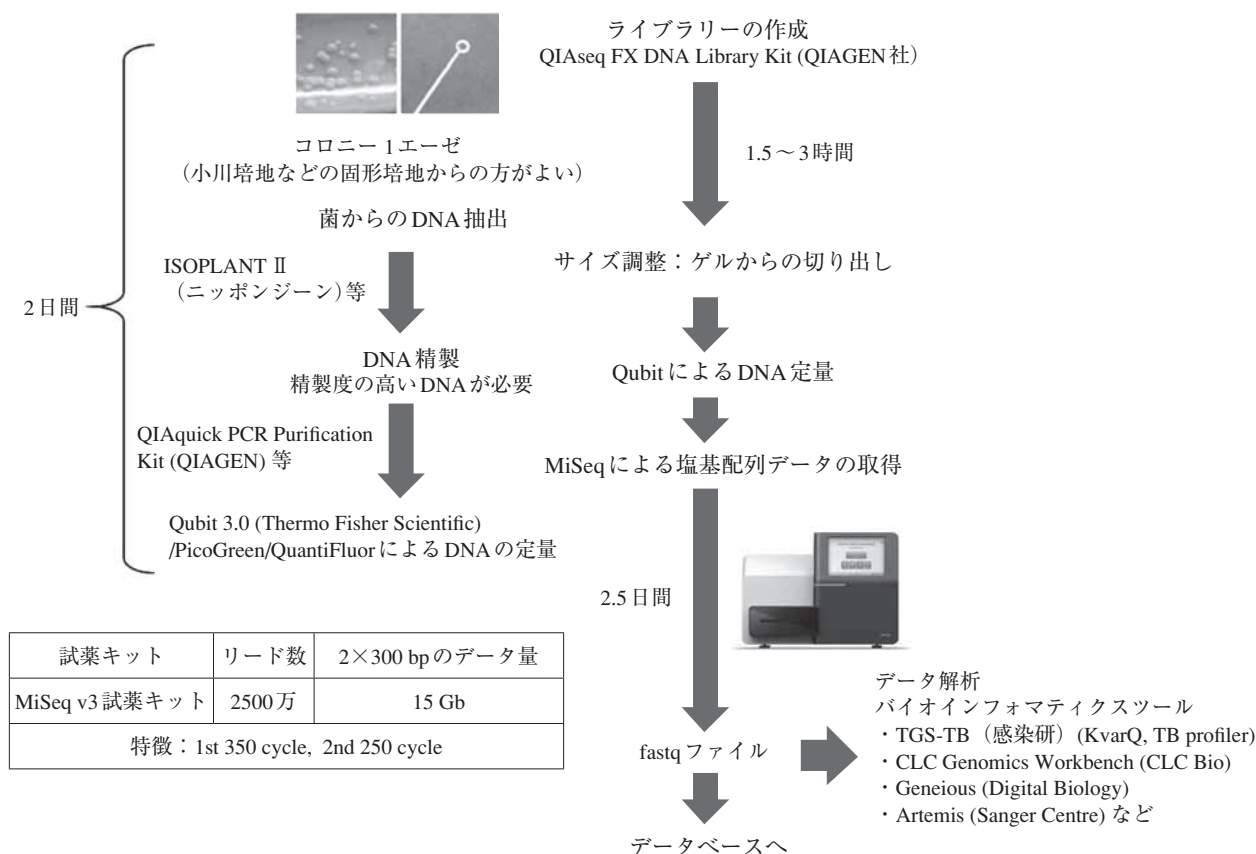


図2 WGS解析のフロー (Illumina MiSeqの使用例)⁶⁾

いる¹³⁾。

表1に結核菌の遺伝子型別法の手法の長所と短所をまとめた。分子疫学研究で解析の目的とする情報とそれを得るためにかかる費用・時間等とを勘案して手法を選択する必要がある。

3. 遺伝子型別の結果の解釈

結果の解釈の際に、疫学的な関連性の有無と菌株間の遺伝子型別の異同について4つ場合が想定される(表2)。疫学的な関連性と遺伝子型別の結果が合致している場合には同じ初発患者由来の感染、不一致の場合には別の菌株の感染の同時期の発症であると考えられる。一方、疫学的な関連性と遺伝子型別の結果に齟齬がある場合には次のことが考えられる。

ケース1：疫学的な関連性があるが、遺伝子型別の結果が異なる場合

遺伝子型別のわずかな違いは、増殖中に生じた変異による可能性がある。菌が増殖している間にも自然変異が1年あたり0.5 SNPs程度生じる¹⁴⁾ため、WGS法では同一の菌株でも分離時期が異なるとSNPsが見られる。一方、VNTR法で数領域が異なる場合やRFLP法でバンドの位置・数に大きな違いが見られる場合は別の菌による感染

である。

ケース2：遺伝子型別の結果は同一であるが、疫学的な関連性がない場合

クラスター分解能が低い方法で解析している場合では過去に地域で流行を起こした株の影響(結果)が反映されていることがある。例えば、日本ではM株¹⁵⁾、カナダではQuebec株¹⁶⁾が報告されている。一方、クラスター分解能が高い方法で解析している場合には、laboratory cross-contaminationが起きた可能性¹⁷⁾やcasual contactで予期しない感染伝播が起きた¹⁸⁾ことが考えられる。

WGS法はRFLP法やVNTR法に比べて高いクラスター分解能をもつが、SNPsの相違が何個までを同一株とするかは定まっていない。集団感染事例に対してWGSを用いた研究論文のメタ解析¹⁹⁾では、5 SNPs以内を同一株としている例²⁰⁾、12 SNPs以内でも関連性があるとの報告もある²¹⁾。これは、SNPsは自然変異で起こる場合があり¹⁴⁾、完全一致でなくても同一菌株であることが否定できないためである。さらにメタ解析の研究では、WGS法で関連性が疑われ、新たな感染経路の解明に有効であった例²⁰⁾と無効であった例²¹⁾を並記している¹⁹⁾。同一株とするSNPsの許容範囲の設定は難しい課題であり、事例の蓄積や疫学情報との関連性を個々に検討する必要がある

表1 分子疫学で使用されている手法の比較

方法	長所	短所
IS6110-RFLP	高いクラスター分解能 再現性と安定性	5コピー数以下ではクラスター分解能が低下 2~3 μ g程度のDNAが必要 操作が煩雑 画像データを収載する大容量メモリーが必要
VNTR	操作が少ない 再現性と安定性 少量のDNAで解析可能 領域ごとに反復数(コピー数)をデジタル表記 できる 費用対効果が良い 結核対策への情報の還元が早い	システムによって領域のクラスター分解能が異なる クラスター分解能を上げるためには多くの領域 の解析が必要
スポリゴタイピング	簡便で安価 少量のDNAで解析可能 多検体に対応 結果をデジタル表記できる On-lineサイトで他の施設での結果を比較できる (SITVIT WEB) IS6110が低コピー数の菌株でも解析できる	IS6110-RFLPより低いクラスター分解能 北京型ではパターンが同じ
WGS	高いクラスター分解能 詳細な遺伝子情報が取得可能 (近縁性解析, 系統解析, 薬剤耐性予測等)	使用機器, 解析費用が高額 データの標準化が難しい 公共利用できるデータベースが未整備

表2 疫学調査と遺伝子型別の結果の解釈

		疫学的関連性	
		あり	なし
遺伝子型別	一致	同一の初発患者由来の感染	齟齬のあるケース2
	不一致	齟齬のあるケース1	別の由来による同時期発症

ある。

4. WGSを活用した今後の研究

今まで述べてきたように全ゲノム情報には多くの情報が含まれているが、未だ十分に活用されていない。今回はNGS器機の手法の違いによる研究の詳細は述べられなかったが、short-read型のNGS器機では*Mycobacterium tuberculosis* H37Rv (lineage 4)を参照株としているため、例えば、アジアで流行している lineage 2/Beijing 系統の結核菌の全ての遺伝子の解析はできない。そのため、系統特異的な形質が見逃されている可能性がある。また、PE/PPEファミリー〔N末にPro-Glu (PE)やPro-Pro-Glu (PPE)の繰り返し構造をもつ〕遺伝子については、参照配列に照合しにくいことからあらかじめ除外しているため、病原性等に関係する重要な遺伝子等を見逃している可能性もある。このようなshort-read型の欠点を補うNGS器機として*de novo* genome assemblyを用いた器機があり²²⁾、PE/PPEファミリー-遺伝子内の病原性関連遺伝子の研究に活用されている²³⁾²⁴⁾。

ゲノム情報の取得時間を短縮して、一晩で結核菌の全

ゲノムの解析ができるNGS器機 (iSeq 100, Illumina)²⁵⁾ や手のひらサイズの小型のNGS器機で取得している報告もある²⁶⁾²⁷⁾。これらの研究では器機 (MinION sequencer, Oxford Nanopore Technology) で取得した遺伝子情報 (fastq ファイル) をクラウドに挙げて、解析は別の施設で集約して行っている。このような手法の導入により結核の疫学研究の新しい展開が期待できる。また、痰から直接WGS解析が可能であり²⁸⁾、今後、短期間で網羅的な感受性試験成績の入手が可能になることが期待できるなど、診断、治療分野への活用が試みられている (表3)。

5. まとめ

WGS解析法の導入により従前の遺伝子型別法で個別に得ていた菌の系統や相互の関連性の情報を1つの手法で得ることができるようになった。さらにWGS法では感染伝播の順番、再発、再燃、複合感染、薬剤耐性遺伝子変異の検出など多くの情報を詳細に得られることから、疫学研究だけでなく診断や治療への応用も期待できる。

一方、WGS法の限界としては排菌している患者が対

表3 WGS法を用いた結核の分子疫学研究の展開例

展開例	参考文献
1) 検体からの効率のよい試料調製試薬, 解析ソフトの開発による試料の質の向上による痰から直接WGS解析	28)
2) 測定器機・解析ソフトやシステム関係のイノベーションによる研究の発展によるPE/PPEファミリー遺伝子の解析	23) 24)
<i>de novo</i> ゲノムアセンブリーを用いた器機での全ゲノム解析	22)
NGS器機の小型化, 省力化	26) 27)
ゲノム情報のクラウドでの管理と解析専門施設での解析の集約化	26) 27)

象であり, 潜在性結核患者の解析はできないため疫学研究上のすべての疑問に対して回答できない。また, 従前の手法と較べて情報が格段に多いために解析に時間を要することから, 目的の情報を選び出して解析する専用の解析ソフトの開発も必要である。

結核は世界的な感染症であり, 地域流行や世界流行株の把握などのためには, ゲノム情報とメタデータのデータベース化と情報の共有が重要である。また, 蓄積されたゲノム情報は結核の感染性, 病原性の解明や新たな診断, 治療薬の開発に有用である。

謝 辞

教育講演の機会を与えて頂きました第94回日本結核病学会総会会長 門田淳一先生に深く感謝申し上げます。WGS解析に用いた結核菌をご提供頂きました結核療法研究協議会に深く感謝申し上げます。WGS解析およびGReATデータベース構築に際し, 多大なご指導, ご助言を頂戴いたしました国立感染症研究所病原体ゲノム解析研究センターの黒田誠先生, 関塚剛史先生, 山下明史先生, 加藤健吾先生に深く感謝申し上げます。研究助成を頂きました日本医療研究開発機構(課題番号JP18fk0108063h), 日本学術振興会(科学研究費基盤研究C16K08346)に深くお礼申し上げます。ゲノム解析の研究にご協力頂きました研究協力者の方々に深くお礼を申し上げます。

著者のCOI (conflicts of interest) 開示: 本論文発表内容に関して特になし。

文 献

- 1) WHO. Tuberculosis. <https://www.who.int/en/news-room/fact-sheets/detail/tuberculosis>. (17 Oct. 2019)
- 2) Kumar A, Farhana A, Guidry L, et al.: Redox homeostasis in mycobacteria: the key to tuberculosis control? *Expert Rev Mol Med*. 2011; 13: e39.
- 3) 東京都福祉健康局: 東京都結核予防計画. http://www.fukushihoken.metro.tokyo.jp/iryu/kansen/toushinfiles/k_keikaku.pdf. (25 Oct. 2019)
- 4) 加藤誠也, 瀧井猛将, 大角晃弘, 他: 結核分子疫学調査の手引き, 第1版. [https://www.jata.or.jp/dl/pdf/law/2017/](https://www.jata.or.jp/dl/pdf/law/2017/09_1pdf)

09_1pdf. (25 Oct. 2019)

- 5) 加藤誠也, 御手洗聡, 瀧井猛将, 他: 結核病原体サーベイランスの実践(総説), 第1版. https://www.jata.or.jp/dl/pdf/law/2017/07_1pdf. (25 Oct. 2019)
- 6) Takii T, Wakabayashi Y, Morishige Y, et al.: Whole-genome sequencing-based epidemiological analysis of anti-tuberculosis drug resistance genes in Japan in 2007: Application of the Genome Research for Asian Tuberculosis (GReAT) database. *Sci Rep*. 2019; 9: 12823.
- 7) Cole ST, Brosch R, Parkhill J, et al.: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998; 393: 537-544.
- 8) Sekizuka T, Yamashita A, Murase Y, et al.: TGS-TB: Total Genotyping Solution for *Mycobacterium tuberculosis* Using Short-Read Whole-Genome Sequencing. *PLoS One*. 2015; 10: e0142951.
- 9) Schleusener V, Koser CU, Beckert P, et al.: *Mycobacterium tuberculosis* resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools. *Sci Rep*. 2017; 7: 46327.
- 10) Wyllie DH, Davidson JA, Grace Smith E, et al.: A Quantitative Evaluation of MIRU-VNTR Typing Against Whole-Genome Sequencing for Identifying *Mycobacterium tuberculosis* Transmission: A Prospective Observational Cohort Study. *EBioMedicine*. 2018; 34: 122-130.
- 11) Gardy JL, Johnston JC, Ho Sui SJ, et al.: Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New Engl J Med*. 2011; 364: 730-739.
- 12) Witney AA, Bateson AL, Jindani A, et al.: Use of whole-genome sequencing to distinguish relapse from reinfection in a completed tuberculosis clinical trial. *BMC Med*. 2017; 15: 71.
- 13) Andres M, van der Werf MJ, Kodmon C, et al.: Molecular and genomic typing for tuberculosis surveillance: A survey study in 26 European countries. *PLoS One*. 2019; 14: e0210080.
- 14) Kay GL, Sergeant MJ, Zhou Z, et al.: Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat Commun*. 2015; 6: 6717.
- 15) Ohkado A, Murase Y, Mori M, et al.: Transmission of specific genotype streptomycin resistant strains of *Mycobacterium tuberculosis* in the Tokyo Metropolitan Area in

- Japan. BMC Infect Dis. 2009 ; 9 : 138.
- 16) Nguyen D, Brassard P, Menzies D, et al.: Genomic characterization of an endemic *Mycobacterium tuberculosis* strain: evolutionary and epidemiologic implications. J Clin Microbiol. 2004 ; 42 : 2573–2580.
 - 17) Wyllie DH, Robinson E, Peto T, et al.: Identifying Mixed *Mycobacterium tuberculosis* Infection and Laboratory Cross-Contamination during Mycobacterial Sequencing Programs. J Clin Microbiol. 2018 ; 56.
 - 18) Auld SC, Shah NS, Mathema B, et al.: Extensively drug-resistant tuberculosis in South Africa: genomic evidence supporting transmission in communities. The Eur Respir J. 2018 ; 52.
 - 19) van der Werf MJ, Kodmon C: Whole-Genome Sequencing as Tool for Investigating International Tuberculosis Outbreaks: A Systematic Review. Front Public Health. 2019 ; 7 : 87.
 - 20) Walker TM, Merker M, Knoblauch AM, et al.: A cluster of multidrug-resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. Lancet Infect Dis. 2018 ; 18 : 431–440.
 - 21) Fiebig L, Kohl TA, Popovici O, et al.: A joint cross-border investigation of a cluster of multidrug-resistant tuberculosis in Austria, Romania and Germany in 2014 using classic, genotyping and whole genome sequencing methods: lessons learnt. Euro Surveill. 2017 ; 22 : 30439.
 - 22) Benjak A, Sala C, Hartkoorn RC: Whole-genome sequencing for comparative genomics and de novo genome assembly. Methods Mol Biol. 2015 ; 1285 : 1–16.
 - 23) Dixit A, Freschi L, Vargas R, et al.: Whole genome sequencing identifies bacterial factors affecting transmission of multidrug-resistant tuberculosis in a high-prevalence setting. Sci Rep. 2019 ; 9 : 5602.
 - 24) Bainomugisa A, Duarte T, Lavu E, et al.: A complete high-quality MinION nanopore assembly of an extensively drug-resistant *Mycobacterium tuberculosis* Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions. Microb Genom. 2018 ; 4.
 - 25) Colman RE, Mace A, Seifert M, et al.: Whole-genome and targeted sequencing of drug-resistant *Mycobacterium tuberculosis* on the iSeq100 and MiSeq: A performance, ease-of-use, and cost evaluation. PLoS Med. 2019 ; 16 : e1002794.
 - 26) Phelan JE, O'Sullivan DM, Machado D, et al.: Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. Genome Med. 2019 ; 11 : 41.
 - 27) Pullen MF, Boulware DR, Sreevatsan S, et al.: Tuberculosis at the animal-human interface in the Ugandan cattle corridor using a third-generation sequencing platform: a cross-sectional analysis study. BMJ Open. 2019 ; 9 : e024221.
 - 28) Doyle RM, Burgess C, Williams R, et al.: Direct Whole-Genome Sequencing of Sputum Accurately Identifies Drug-Resistant *Mycobacterium tuberculosis* Faster than MGIT Culture Sequencing. J Clin Microbiol. 2018 ; 56 : e00666–18.

Review Article

WHOLE GENOME SEQUENCING AS A PROMISING TOOL FOR EPIDEMIOLOGICAL INVESTIGATIONS OF TUBERCULOSIS

^{1,2}Takemasa TAKII

Abstract The epidemiologists of tuberculosis are able to obtain huge amount of genotypic information at one time due to the innovation of new sequencing technology. The new technology, whole genome sequencing (WGS), possesses higher ability of analyzing the cluster and linkage associations among the clinical isolates of tuberculosis (TB) than traditional methods like genotyping, RFLP, VNTR and spoligotyping. Furthermore, the WGS analysis is also able to predict drug susceptibility of the isolate against anti-TB drugs. I here review the role of the WGS analysis in epidemiological investigations of tuberculosis.

Key words: Tuberculosis, Molecular epidemiology, Whole genome analysis

¹Department of Mycobacteriology, Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, ²Graduated School of Pharmaceutical Sciences, Nagoya City University

Correspondence to: Takemasa Takii, Department of Mycobacteriology, Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, 3-1-24, Matsuyama, Kikyose-shi, Tokyo 204-8533 Japan. (E-mail: t-takii@jata.or.jp)