

RESEARCH ARTICLE

Overcoming the pitfalls of automatic interpretation of whole genome sequencing data by online tools for the prediction of pyrazinamide resistance in *Mycobacterium tuberculosis*

Tomotada Iwamoto¹*, Yoshiro Murase², Shiomi Yoshida³, Akio Aono², Makoto Kuroda⁴, Tsuyoshi Sekizuka⁴, Akifumi Yamashita⁴, Kengo Kato⁴, Takemasa Takii⁵, Kentaro Arikawa¹, Seiya Kato⁶, Satoshi Mitarai^{2,7}*

1 Department of Infectious Diseases, Kobe Institute of Health, Kobe City, Japan, **2** Bacteriology Division, Department of Mycobacterium Reference and Research, Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Kiyose City, Tokyo, Japan, **3** Clinical Research Center, National Hospital Organization Kinki-chuo Chest Medical Center, Sakai City, Osaka, Japan, **4** Pathogen Genomics Center, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo, Japan, **5** Molecular Epidemiology Division, Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Kiyose City, Tokyo, Japan, **6** Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Kiyose City, Tokyo, Japan, **7** Basic Mycobacteriosis, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki City, Nagasaki, Japan

* These authors contributed equally to this work.

* tomotada_iwamoto@office.city.kobe.lg.jp (TI); mitarai@jata.or.jp (SM)



OPEN ACCESS

Citation: Iwamoto T, Murase Y, Yoshida S, Aono A, Kuroda M, Sekizuka T, et al. (2019) Overcoming the pitfalls of automatic interpretation of whole genome sequencing data by online tools for the prediction of pyrazinamide resistance in *Mycobacterium tuberculosis*. PLoS ONE 14(2): e0212798. <https://doi.org/10.1371/journal.pone.0212798>

Editor: W.C. Yam, The University of Hong Kong, CHINA

Received: October 19, 2018

Accepted: February 9, 2019

Published: February 28, 2019

Copyright: © 2019 Iwamoto et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This research was supported by (Japan Agency for Medical Research and Development) AMED under Grant Number JP18fk0108063 (TI, SM, and SK). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Objectives

Automated online software tools that analyse whole genome sequencing (WGS) data without the need for bioinformatics expertise can motivate the implementation of WGS-based molecular drug susceptibility testing (DST) in routine diagnostic settings for tuberculosis (TB). Pyrazinamide (PZA) is a key drug for current and future TB treatment regimens; however, it was reported that predictive power for PZA resistance by the available tools is low. Therefore, this low predictive power may make users hesitant to use the tools. This study aimed to elucidate why and to uncover the real performance of the tools when taking into account their variation calling lists (manual inspection), not just their automated reporting system (default setting) that was evaluated by previous studies.

Methods

WGS data from 191 datasets comprising 108 PZA-resistant and 83 susceptible strains were used to evaluate the potential performance of the available online tools (TB Profiler, TGS-TB, PhyResSE, and CASTB) for predicting phenotypic PZA resistance.

Results

When taking into consideration the variation calling lists, 73 variants in total (47 non-synonymous mutations and 26 indels) in *pncA* were detected by TGS-TB and PhyResSE, covering

Competing interests: The authors have declared that no competing interests exist.

all mutations for the 108 PZA-resistant strains. The 73 variants were confirmed by Sanger sequencing. TB Profiler also detected all but three complete loss, two large deletion at the 3'-end, and one relatively large insertion of *pncA*. On the other hand, many of the 73 variants were lacking in the automated reporting systems except by TGS-TB; of these variants, CASTB detected only 20. By applying the 'non-wild type sequence' approach for predicting PZA resistance, accuracy of the results significantly improved compared with that of the automated results obtained by each tool.

Conclusion

Users can obtain more accurate predictions for PZA resistance than previously reported by manually checking the results and applying the 'non-wild type sequence' approach.

Introduction

Mycobacterium tuberculosis (MTB) is a slow-growing bacillus and thus requires several weeks to months for phenotypic drug-susceptibility testing (DST). The essential role and ubiquitous use of pyrazinamide (PZA) in tuberculosis (TB) therapy underlines the importance of DST against PZA [1,2]. However, the test for PZA is more complicated and less reliable than those for other anti-TB drugs due to the difficulty of growing TB under the acidic (pH < 6.0) conditions required for the assay [3–5]. As the current phenotypic test is difficult and time consuming, clinical laboratories are eager for a more accurate and rapid assay to detect PZA resistance. Genotypic DST is an ideal approach for bypassing the difficulty, inaccuracy, and lengthy process of phenotypic DST, providing rapid results [6,7]. However, genotypic DST for PZA is extremely challenging because the *pncA* mutations that induce loss of pyrazinamidase (PZase) activity (the major mechanism leading to PZA resistance) are highly diverse and scattered throughout the *pncA* gene [8,9].

Whole-genome sequencing (WGS) can theoretically be used to rule in resistance to all drugs, which would accelerate DST to design appropriate treatment regimens. To assist the interpretation of complex WGS data, several automated online software tools are available for TB DST, including CASTB [10], PhyResSE [11], TB Profiler [12], Mykrobe predictor [13], and TGS-TB [14]. These tools are expected to overcome the barrier of implementing WGS methodologies in routine diagnostic settings, which usually lack bioinformatics expertise. However, it has been reported that the sensitivity and specificity for predicting resistance varies significantly between the different tools and antibiotics tested [15]. Moreover, previous studies that evaluated the predictive accuracy of these automated tools for drug resistance revealed that the accuracy for predicting PZA resistance is low [12,15–17]. This may make users hesitant to use the tools.

Herein, we wanted to emphasise that a lack of drug resistance-associated variants in the automatically generated plain text report by the online tools does not necessarily mean the tools did not detect mutations in the region of interest, which are possibly causing resistance. Indeed, TB Profiler, TGS-TB, and PhyResSE provide variant calling lists in addition to the plain text report of drug resistance prediction no matter if they are included or not in the pre-defined mutation catalogue for automated reporting systems. Therefore, real performance evaluation of the tools should be performed by two different aspects, one at the level of the automated reporting system and the other at the full potential of its performance for the detection of mutations in a region of interest; all previous reports that evaluated online tool performance analysed the former aspect but did not take into account the latter [15–17].

To clarify the reasons for the reported low accuracy of PZA resistance prediction and to reveal the real performance of the tools, we evaluated the tools using WGS data from 191 datasets on resistant and susceptible strains at the level of automated reporting and the full potential of the tools.

Materials and methods

Specimens

A total of 191 MTB isolates collected from two different isolate stocks were used in this study. To avoid any unambiguous results by heteroresistance due to the existence of a mixed population [18], single colony isolation from the stock culture was performed. Sample set 1 included 112 multidrug-resistant MTB (MDR-TB) isolates from 112 patients, collected through the RYOKEN TB research consortium during 2012–2013 as part of a national drug resistance survey. Sample set 2 included 79 isolates obtained from Kinki-chuo Chest Medical Center located in the western part of Japan between 2001 and 2015; of these, 53 strains were MDR-TB. The isolates were obtained from 77 single patients including two cases where non-MDR-TB developed into MDR-TB in a single patient.

Ethics statement

All of the clinical isolates in this study were fully anonymized before we accessed them. Clinical isolates of *M. tuberculosis* in set 2 were used with approval of the ethical committee of Kinki-chuo Chest Medical Center (reference 2015/493). Clinical isolates in set 1 were obtained through the survey followed Japanese national guidelines for epidemiological research and ethical approval was not required for this laboratory study based on *M. tuberculosis* isolates.

Phenotypic DST of PZA

For all 191 strains, PZA susceptibility was measured using MGIT AST PZA (Becton Dickinson, Franklin Lakes, NJ) according to manufacturer's instructions. The critical concentration used in this study was 100 µg/mL. Extended analysis using the PZase assay by Wayne's method [19] was performed for the 16 strains that showed discrepant results between MGIT AST PZA and mutations in *pncA* or *rpsA* to determine PZA susceptibility. This analysis was also performed for all strains in sample set 1 (n = 112) to supplement MGIT AST PZA testing. Minimum inhibitory concentration (MIC) measurement for PZA was performed for 7 of the 16 strains whose PZA susceptibility could not be determined even after the PZase assay. MIC was determined [18] from PZA concentrations of 12.5, 25, 50, 100, 200, 300, 400, and 800 µg/mL in MGIT culture media, using the MGIT EpiCenter system. The MIC for *M. tuberculosis* H37Rv (control for PZA susceptibility) and *M. bovis* BCG Tokyo (control for PZA resistance) was found to be 50 µg/mL and > 800 µg/mL, respectively. We considered MIC values ≤ 100 as PZA susceptible.

WGS data analysis

Genomic DNA of each isolate was prepared from MTB grown on Ogawa medium using an Isoplant Kit (Wako Pure Chemical Industries Ltd, Tokyo, Japan). A DNA-seq library was prepared with the QIAseq FX DNA library kit (Qiagen, Hilden, Germany) using 50 ng TB genomic DNA, followed by paired-end sequencing using Illumina MiSeq (Illumina, San Diego, CA) and the MiSeq Reagent Kit v3 (600 cycles) according to manufacturer's instructions. The read data were then analysed using TB Profiler (online version, September 2018) [12], TGS-TB v2 [14], PhyResSE v1.0 [11], and CASTB v1.5 [10]. In addition to the text report obtained

through automated interpretation (default result), we manually inspected the tool variation calling lists for *pncA*, *rpsA*, *panD*, and their promoter regions (manual result); this was not performed with CASTB as it does not provide a variation calling list. The synonymous *pncA* Gln122Gln mutation and the known phylogenetic informative mutations [12] *rpsA* Met432Thr (for LAM lineage) and *rpsA* Arg212Arg (for Beijing lineage) detected in this study were excluded from further analysis. Phenotypic PZA susceptibility testing (MGIT testing) complemented with the PZase assay and MIC measurement were used as the standard reference for evaluating the online tools' prediction powers for PZA resistance.

Sanger sequencing for *pncA*

To validate the mutation call by online tools, we performed sanger sequencing for all mutation variants of *pncA* detected in this study as previously described [20].

Statistical analysis

All statistical analyses were performed with EZR [21], a graphical user interface for R.

Read data deposit

Raw sequence reads of all 191 MTB strains subjected to WGS were deposited in the DDBJ Sequence Read Archive (DRA; <https://www.ddbj.nig.ac.jp/dra/index.html>) under the study accession number DRA006844 (S1 Table).

Results

Variation calling lists by online tools

Mutation variants detected in the 191 strains are summarised in Table 1.

All mutations in *panA* detected through automated interpretation and manual inspection by three online tools (TB Profiler, TGS-TB, and PhyResSE) were validated by Sanger sequencing. Forty-seven single-nucleotide variants (SNVs) were detected by all three tools and were consistent with Sanger sequencing results. Regarding the indels, TGS-TB and PhyResSE called 26 variations that were consistent with Sanger sequencing, whereas TB Profiler showed different calls for 5 indel variants and no calls for 6 indels (three complete losses of *pncA*, two large deletions at the 3'-end of *pncA*, and one relatively large insertion in *pncA*); details of these discrepancies are listed in S2 Table. Furthermore, eight SNVs and one indel in *rpsA* as well as one SNV in the *panD* promoter were detected by all three tools; S1 Fig shows an image of the

Table 1. Mutation variants and detection by online tools.

Gene	Mutation	No. of variation	TB Profiler			TGS-TB			PhyRes			CASTB	
			Default	Manual	None	Default	Manual	None	Default	Manual	None	Default	None
<i>pncA</i>	SNV	47	35	12	0	42	5	0	33	14	0	10	37
	Indel*	26	0	20**	6	26	0	0	0	26	0	10	16
<i>rpsA</i>	SNV***	8	0	8	0	1	7	0	1	7	0	0	8
	Indel	1	0	1	0	0	1	0	0	1	0	0	1
<i>panD</i> promoter	SNV	1	0	1	0	0	1	0	0	1	0	0	1

None, not detected by either default or manual inspection.

*Variation of the three complete and two 3' end deletions were discriminated based on their regions of deletion.

**Five deletions which showed discrepancy with TGS-TB and PhyRes were included (see S2 Table).

***Met432Thr is not included since it is a LAM defined SNP.

<https://doi.org/10.1371/journal.pone.0212798.t001>

Table 2. Association between PZA resistant conferring mutations and MGIT PZA susceptibility testing.

	MGIT (set 1, n = 112)		MGIT (set 2, n = 79)	
	R	S	R	S
Default mutations by TB-Profiler	35	5	21	4
Mutations other than default by TB Profiler	25	7	17	0
No mutations	0	40	0	37

PZA, pyrazinamide; R, resistant; S, sensitive

<https://doi.org/10.1371/journal.pone.0212798.t002>

detection of these variants by TGS-TB. In total, 83 variants (47 SNVs and 26 indels in *pncA*, 8 SNVs and 1 indel in *rpsA*, and 1 SNV in the *panD* promoter) were detected in this study and used for further analysis to evaluate the online tools' prediction powers for PZA resistance. All strains with variants and phenotypic PZA susceptibility information are listed in [S1 Table](#).

Determination of PZA susceptibility

Of the 191 strains analysed, 98 demonstrated resistance by the MGIT test with mutations in *pncA* and its promoter (60 in set 1 and 38 in set 2) and were judged PZA resistant ([Table 2](#)). The 77 strains that were deemed susceptible by MGIT and without mutations (40 in set 1 and 37 in set 2) were judged PZA susceptible. These classifications correspond with PZase analysis results ([S1 Table](#)).

Meanwhile, 16 strains of 191 showed discrepant results between genotypic and phenotypic results ([Table 2](#)); nine strains were found negative by PZase analysis (the major mechanism leading to PZA resistance) with mutations in the *pncA* gene and were thus characterised as PZA resistant ([Table 3](#)).

Of the remaining seven strains that were further analysed by MICs, six were deemed PZA susceptible and one was PZA resistant ([Table 3](#)). None of the 191 tested isolates possessed

Table 3. Extended analysis of the 16 *M. tuberculosis* strains for determining PZA susceptibility.

Strain ID	Source	Default mutations by TB Profiler	Mutations other than default by TB Profiler in candidate genes	MGIT PZA	PZase	MIC (µg/mL)	Conclusion
126	set 2	<i>pncA</i> (Lys96Gln)	none	S	neg.	N/A	R
135	set 2	<i>pncA</i> (Asp63Gly)	none	S	neg.	N/A	R
131	set 2	<i>pncA</i> (Gly78Ser)	none	S	neg.	N/A	R
020	set 1	<i>pncA</i> (Leu182Ser)	none	S	neg.	N/A	R
011	set 1	<i>pncA</i> (T-11C promoter)	none	S	neg.	N/A	R
018	set 1	<i>pncA</i> (Lys96Asn)	none	S	neg.	N/A	R
017	set 1	<i>pncA</i> (Met175Ile)	none	S	±	200	R
019	set 1	<i>pncA</i> (Leu4Ser)	<i>rpsA</i> (Asp15Ala)	S	neg.	N/A	R
129	set 2	<i>pncA</i> (Gly97Asp)	<i>rpsA</i> (1834845: TCGCCGCCGCCG /TCGCCGCCG), <i>panD</i> promoter (A-13G)	S	neg.	N/A	R
052	set 1	none	<i>pncA</i> (Ser67Stop)	S	neg.	N/A	R
049	set 1	none	<i>pncA</i> (Val44Ala)	S	±	100	S
048	set 1	none	<i>rpsA</i> (1834845: TCGCCGCCGCCG /TCGCCGCCG)	S	pos.	50	S
047	set 1	none	<i>rpsA</i> (Asp123Ala)	S	pos.	50	S
042	set 1	none	<i>rpsA</i> (Thr42Ala)	S	pos.	100	S
043	set 1	none	<i>rpsA</i> (Asp27Asn)	S	pos.	100	S
041	set 1	none	<i>rpsA</i> (Tyr109His), <i>pncA</i> (Stop187Gly)	S	±	50	S

R, resistant; S, sensitive; N/A., not applicable; neg., negative; pos., positive; ±, indeterminate

<https://doi.org/10.1371/journal.pone.0212798.t003>

mutations in the *panD* gene. Overall, the tested isolates were phenotypically classified into 108 PZA-resistant and 83 PZA-susceptible isolates.

Mutations in PZA resistance-conferring genes

All 108 PZA-resistant strains possessed mutations in the *pncA* gene which consists of 44 different non-synonymous mutations, 26 different indels, and one mutation in the promoter region (S3 Table). None of the 71 mutations were found in the 83 PZA-susceptible strains. Among these mutations, 61 genetic variants had single *pncA* mutations per strain, one variant showed double *pncA* mutations, nine variants exhibited double mutations in both *pncA* and *rpsA* genes, and one variant had triple mutations in *pncA*, *rpsA*, and the *panD* promoter. All 26 indels in the *pncA*-containing region detected in this study led to PZA resistance, whereas two non-synonymous mutations in *pncA* (Stop187Gly, Val44Ala) did not confer PZA resistance. None of the 108 PZA-resistant strains possessed mutations in either *rpsA* or *panD* alone, which are putative PZA resistance-conferring genes. Among the 83 PZA-susceptible isolates, one possessed a *pncA* Val44Ala mutation, four exhibited single mutations in *rpsA* (Thr42Ala, Asp27Asn, Asp123Ala, or deletion at 1824845), and one showed double mutations in *pncA* Stop187Gly and *rpsA* Tyr109His (Table 2).

pncA mutations in this study were found distributed across the entire gene (Table 4), as previously reported [7,22,23].

All 45 non-synonymous mutations in *pncA*—except for three mutations (Gly78Ser, Thr103Ser, Val163Gly)—detected in this study from 108 PZA-resistant strains were included in either of five comprehensively validated mutation lists as high confidence mutations [22–26] (Table 4). Although these three mutations were not on the list, the strains carrying these mutations showed negative results by PZase analysis and resistance via the MGIT test (S1 Table). Two strains harbouring the Stop187Gly and Val44Ala mutations in *pncA* were determined as PZA-susceptible by the MGIT test and neither mutation was listed in the five published lists (Table 4). The MIC values for these strains were not greater than 100 µg/mL, whereas *M. tuberculosis* H37Rv and *M. bovis* BCG Tokyo showed MICs of 50 µg/mL and > 800 µg/mL, respectively.

Sensitivity and specificity of PZA resistance prediction by online tools

Using WGS data from 191 strains, we analysed the sensitivity and specificity of PZA resistance predictions by different tools available online (TB profiler, TGS-TB, PhyResSE, and CASTB). To evaluate tool performance at the automated reporting system level and at their full potential for mutation detection, we categorised the mutations detected by TB profiler, TGS-TB, and PhyResSE as either default or manual (S3 Table and Table 1); default indicates the automated interpretation report and manual indicates all mutations detected by the tools other than default.

Remarkable differences between the tools were found in their detection of indels. TGS-TB returned all 26 indel variants in the *pncA* gene through default settings, whereas the other two tools did not (Table 1). The default detection rate of SNVs in *pncA* varied among the tools; TGS-TB exhibited the best detection rate (42/47), followed by TB Profiler (35/47), PhyResSE (33/47), and CASTB which detected only 10 SNVs and 10 indels (Table 1). As a result, the default sensitivity of the tools was low (60.1%, 49.1%, and 31.5% for TB Profiler, PhyResSE, and CASTB, respectively), except for TGS-TB (97.2%; Table 5) which is consistent with previous studies [12,15–17].

This drastically changed when we evaluated the detection rate with manual settings. PhyResSE detected all 26 indels while TB Profiler missed 6 indels (Table 1); of these 6 indels, two

Table 4. Non-synonymous mutations in *pncA* and their presence in the published lists.

Mutation	PZA sensitivity		<i>pncA</i> mutations in the published lists				
	no. of R	no. of S	Miotto [22]	Yadon [23]	Manson [24]	Miotto [25]	Farhat [26]
<i>pncA</i> (Met1Leu)	1	0	no	yes	no	no	no
<i>pncA</i> (Ala3Glu)	6	0	yes	yes	no	yes	no
<i>pncA</i> (Leu4Ser)	1	0	no	yes	no	yes	yes
<i>pncA</i> (Gln10Arg)	1	0	yes	yes	no	yes	yes
<i>pncA</i> (Asp12Ala)	3	0	yes	no	no	yes	yes
<i>pncA</i> (Asp12Glu)	1	0	yes	yes	no	no	yes
<i>pncA</i> (Cys14Stop)	1	0	yes	yes	no	yes	no
<i>pncA</i> (Cys14Tyr)	1	0	no	yes	no	no	no
<i>pncA</i> (Tyr34Asp)	8	0	no	no	no	yes	yes
<i>pncA</i> (Tyr34Stop)	1	0	yes	no	yes	yes	no
<i>pncA</i> (His51Gln)	1	0	yes	no	yes	yes	no
<i>pncA</i> (Pro54Leu)	3	0	yes	yes	no	yes	yes
<i>pncA</i> (His57Gln)	1	0	yes	yes	no	no	no
<i>pncA</i> (Leu159Pro)	1	0	no	yes	no	no	no
<i>pncA</i> (Asp63Gly)	1	0	yes	yes	no	yes	yes
<i>pncA</i> (Ser67Stop)	1	0	no	yes	no	no	yes
<i>pncA</i> (Trp68Arg)	1	0	yes	yes	no	yes	yes
<i>pncA</i> (Thr76Pro)	1	0	yes	no	no	yes	yes
<i>pncA</i> (Gly78Ser)	3	0	no	no	no	no	no
<i>pncA</i> (Lys96Asn)	2	0	yes	yes	no	yes	no
<i>pncA</i> (Lys96Gln)	1	0	yes	yes	no	no	no
<i>pncA</i> (Gly97Asp)	3	0	yes	yes	no	yes	yes
<i>pncA</i> (Thr100Pro)	1	0	yes	no	no	no	yes
<i>pncA</i> (Tyr103Ser)	1	0	no	no	no	no	no
<i>pncA</i> (Tyr103Stop)	1	0	yes	yes	yes	yes	yes
<i>pncA</i> (Ser104Ile)	2	0	yes	no	no	no	no
<i>pncA</i> (Leu116Arg)	1	0	yes	yes	no	yes	no
<i>pncA</i> (Trp119Cys)	1	0	yes	no	no	no	no
<i>pncA</i> (Val125Gly)	1	0	yes	no	no	no	no
<i>pncA</i> (Val128Gly)	1	0	yes	no	no	yes	yes
<i>pncA</i> (Gly132Ala)	1	0	yes	yes	no	yes	no
<i>pncA</i> (Asp136Gly)	1	0	no	no	no	no	yes
<i>pncA</i> (Cys138Arg)	1	0	yes	yes	no	no	no
<i>pncA</i> (Val139Ala)	2	0	no	yes	no	yes	yes
<i>pncA</i> (Val139Met)	1	0	no	yes	no	no	yes
<i>pncA</i> (Gln141Pro)	3	0	yes	no	no	yes	yes
<i>pncA</i> (Ala146Thr)	1	0	no	yes	no	no	no
<i>pncA</i> (Val155Met)	1	0	yes	yes	no	no	no
<i>pncA</i> (Val163Gly)	1	0	no	no	no	no	no
<i>pncA</i> (Ser164Pro)	1	0	yes	yes	no	no	no
<i>pncA</i> (Leu172Pro)	1	0	yes	yes	no	yes	yes
<i>pncA</i> (Met175Ile)	2	0	yes	yes	no	yes	no
<i>pncA</i> (Val180Phe)	1	0	yes	yes	no	yes	yes
<i>pncA</i> (Leu182Ser)	2	0	yes	yes	no	no	yes
<i>pncA</i> (Stop187Gly)	0	1	no	no	no	no	no

(Continued)

Table 4. (Continued)

Mutation	PZA sensitivity		<i>pncA</i> mutations in the published lists				
	no. of R	no. of S	Miotto [22]	Yadon [23]	Manson [24]	Miotto [25]	Farhat [26]
<i>pncA</i> (Val44Ala)	0	1	no	no	no	no	no

R, resistant; S, sensitive

<https://doi.org/10.1371/journal.pone.0212798.t004>

were large deletions at the 3' region of *pncA*, three were complete losses of *pncA*, and one was a large fragment insertion at 2288788 (S3 Table). Five indels that were miscalled by TB Profiler (S2 Table) were treated as the call for PZA resistance prediction by TB Profiler. All mutations in *rpsA* and the *panD* promoter in our sample set were detected by the three tools through searches beyond the default report (Table 1). When we deemed the strains with these indels and non-synonymous mutations as PZA resistant (interpret 'non-wild type sequence' as PZA resistant), the three online tools demonstrated remarkably high sensitivity and specificity in predicting PZA resistance (Table 5). Specificity slightly decreased when manual results were compared with those of default (Table 5). This is due to the inclusion of mutations in *rpsA* whose association with PZA resistance is debatable [27,28].

Discussion

Our analysis of online tool performance in detecting PZA resistance demonstrated that the tools can predict resistance with more sensitivity than previously reported when all detected variants (non-synonymous mutations and indels) are taken into account (Table 5). In fact, all or most mutations in the *pncA* gene of 108 PZA-resistant strains were detected by TB Profiler, TGS-TB, and PhyResSE but many variants were missing in the automated interpretation for plain text report. As a result, the accuracy of PZA resistance prediction by online tools differed

Table 5. Sensitivity and specificity of the online tools in predicting PZA resistance.

			PZA susceptibility		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
			R (n = 108)	S (n = 83)				
Default setting	TGS-TB	R	105	1	0.972 (0.921–0.991)	0.988 (0.935–0.998)	0.991 (0.949–1.00)	0.965 (0.900–0.993)
		S	3	82				
	TB Profiler	R	65	0	0.601 (0.510–0.694)	1 (0.935–1.000)	1.000 (0.918–1.000)	0.659 (0.569–0.741)
		S	43	83				
	PhyResSE	R	53	1	0.491 (0.396–0.585)	0.988 (0.935–0.998)	0.981 (0.901–1.000)	0.599 (0.511–0.681)
		S	55	82				
	CASTB	R	34	1	0.315 (0.227–0.402)	0.988 (0.935–0.998)	0.971 (0.851–0.999)	0.526 (0.444–0.606)
		S	74	82				
Manual setting	TGS-TB	R	108	6	1 (0.950–1.000)	0.928 (0.872–0.983)	0.947 (0.889–0.980)	1.000 (0.931–1.000)
		S	0	77				
	TB Profiler	R	102	6	0.944 (0.883–0.979)	0.928 (0.849–0.973)	0.944 (0.883–0.979)	0.928 (0.849–0.973)
		S	6	77				
	PhyResSE	R	108	6	1 (0.950–1.000)	0.928 (0.872–0.983)	0.947 (0.889–0.980)	1.000 (0.931–1.000)
		S	0	77				

R, resistance; S, Sensitive; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value
All strains with indels or mutations detected by manual setting are classified as PZA resistance.

<https://doi.org/10.1371/journal.pone.0212798.t005>

substantially between the default report and manual inspection (Table 5). Thus, the fundamental cause for the reported low sensitivity of PZA resistance prediction is the tool interpretation pipeline, which is built upon pre-defined mutation catalogues and not due to the inability of software algorithms in detecting genetic variants.

The main aim of the online tools is to make complex WGS data analysis accessible for routine use in clinical settings where there is usually a lack in bioinformatics expertise. The advantage of these tools is automated interpretation for drug resistance upon pre-defined mutation catalogues. However, this study emphasised that proper knowledge of drug resistance-conferring mutations with some background in bioinformatics is still needed to avoid misinterpretations, i.e., a lack of drug resistance-associated variants in the automated report does not necessarily mean that there are no mutations in the region of interest causing resistance. Users should be aware that three online tools (TB Profiler, TGS-TB, PhyResSE) used in this study provide variant calling lists in addition to the plain text report of drug resistance prediction no matter if they are included or not in the pre-defined mutation catalogue for automated reporting systems. The previously reported low accuracy of PZA resistance prediction can be improved by a manual inspection of the list (Table 5). Although the tools do not guide users for this adequately, the variation calling list provided by the tools is a straightforward approach.

A recent genome-wide association study (GWAS) indicated that PZA resistance prediction can be enhanced by the inclusion of insertions and deletions [29], which was also confirmed by our data. It is clear that the superiority of TGS-TB over the other tools regarding the default setting was mainly due to the inclusion of indels for reporting PZA resistance. Our data showed a higher ratio of strains with structural variants by indels in PZA resistance (29.6%) compared with that of the previously reported values (16%, 19.6%, and 22.7%) for global samples [7], the Chinese province Zhejiang [20], and the Chinese city Hanzhou [30], respectively. This higher ratio was observed for both sample set 1 (28.3%) and set 2 (30.0%) in our study, thereby reflecting the situation in Japan. These data imply that the impact of structural variants on PZA resistance is more significant in Japan where the ancient Beijing genotype is predominant compared with that of other regions [31,32]. Further studies are needed to clarify this distribution.

In this study, we interpreted strains with any non-synonymous mutations or indels in *pncA* as genotypic PZA-resistant strains; this ‘non-wild type sequence’ approach was implemented in the Genoscholar PZA-TB II line probe assay (PZA-LPA2; NIPRO Corporation, Osaka, Japan) [28]. One concern with this approach is the overestimation of resistant strains caused by non-resistance-associated mutations, including synonymous and lineage-specific mutations. Unlike the line probe assay, WGS-based DST can exempt them from mutations conferring PZA resistance, but with a lack of reliable lists as a drawback. Yadon et al. [23] identified many non-synonymous amino acid substitutions that do not confer PZA resistance by using a comprehensive library of *pncA* mutations in *M. tuberculosis* created by random PCR mutagenesis. Although these mutations were created *in vitro*, such a list would be helpful in identifying exemptions from the ‘non-wild type sequence’ approach. To evaluate the impact of non-resistance-associated mutations on this approach in clinical settings, we examined recent studies performed in clinical settings under well-controlled conditions. The degree of adverse effects by non-resistance-associated mutations varied but was low, i.e., one mutation out of 400 PZA-susceptible isolates (1/400) [3], 5/156 [20], 0/57 [28], and 2/83 in this study. This implies that the occurrence of such mutations inside a patient’s body is rare. Although the overestimation of resistant strains is unavoidable by the ‘non-wild type sequence’ approach, it can effectively screen PZA-susceptible strains and help reduce the need for culture-based DST, which is labour-intensive, requires facilities with biosafety measures, and is usually less reliable for PZA

than other key drugs [33]. Inclusion of other PZA resistance-associated genes such as *panD* [34] may improve the accuracy of screening for PZA-susceptible strains, although it was not detected in this study. In our sample set, inclusion of mutations in *rpsA* as PZA resistance-conferring mutations decreased specificity but did not increase sensitivity of WGS-based tools for PZA resistance prediction (Table 5). It remains debatable if *rpsA* is associated with PZA resistance [27,28]. Moreover, to draw conclusions on its use for the prediction of PZA resistance, MIC measurement for a large number of strains with mutations in *rpsA* should be performed.

It is of note that detection of a complete region of interest deletion, such as for *pncA* in this study and *katG* for isoniazid resistance [35], is troublesome. TGS-TB and PhyResSE can detect complete deletions; however, we had to check the mapping status of PhyResSE results, which is not straightforward from the variation calling list and would likely be missed. TGS-TB was much more user friendly in this regard. TB profiler did not report the deletion as the tool has no function to visualise mapping status. To make the tools more user friendly, the developers should consider reporting the intactness of the region of interest as 'intact and wild-type' and including it in the automated reporting system, especially for *pncA* and possibly *katG* as well.

The study limitation is that we did not evaluate heteroresistant strains of mixed populations that appear in clinical settings [18]. How to deal with phenotype/genotype discrepant results, mutations conferring low levels of resistance, and heteroresistance are the practical questions that need to be addressed by WGS data analysis [36,37]. In conclusion, our findings demonstrated that users can obtain more accurate predictions for PZA resistance than previously reported by carefully checking the results manually and applying the 'non-wild type sequence' approach. It would be worthwhile for developers to include all *pncA* mutations without synonymous and lineage-specific mutations into their automated reporting system for PZA resistance prediction as already implemented by TGS-TB for indels. This makes online PZA resistance prediction tools more suitable for clinical setting where there is usually a lack in bioinformatics expertise.

Supporting information

S1 Fig. Capture image of mapping results by TGS-TB.
(PPTX)

S1 Table. Strains analyzed in this study and their profiles of PZA resistance.
(XLSX)

S2 Table. Variants in *pncA* those showed discrepant mutation calls among the online Tools.
(XLSX)

S3 Table. Variation of the mutations detected in this study.
(XLSX)

Acknowledgments

We would like to thank Editage (www.editage.jp) for English language editing.

Author Contributions

Conceptualization: Tomotada Iwamoto, Yoshiro Murase, Satoshi Mitarai.

Data curation: Tomotada Iwamoto, Shiomi Yoshida, Makoto Kuroda, Tsuyoshi Sekizuka, Akifumi Yamashita, Kengo Kato, Takemasa Takii, Kentaro Arikawa, Satoshi Mitarai.

Formal analysis: Tomotada Iwamoto, Yoshiro Murase, Shiomi Yoshida, Akio Aono, Tsuyoshi Sekizuka, Takemasa Takii, Kentaro Arikawa, Satoshi Mitarai.

Funding acquisition: Tomotada Iwamoto, Seiya Kato, Satoshi Mitarai.

Investigation: Tomotada Iwamoto, Yoshiro Murase.

Methodology: Shiomi Yoshida, Tsuyoshi Sekizuka, Akifumi Yamashita, Kengo Kato.

Project administration: Seiya Kato, Satoshi Mitarai.

Resources: Satoshi Mitarai.

Software: Makoto Kuroda.

Supervision: Makoto Kuroda, Seiya Kato, Satoshi Mitarai.

Writing – original draft: Tomotada Iwamoto.

Writing – review & editing: Tomotada Iwamoto, Yoshiro Murase, Akio Aono, Makoto Kuroda, Seiya Kato, Satoshi Mitarai.

References

1. Horne DJ, Pinto LM, Arentz M, Lin SY, Desmond E, Flores LL, et al. Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line and second-line antituberculosis drugs. *J Clin Microbiol.* 2013; 51: 393–401. <https://doi.org/10.1128/JCM.02724-12> PMID: 23152548.
2. Zignol M, Dean AS, Alikhanova N, Andres S, Cabibbe AM, Cirillo DM, et al. Population-based resistance of *Mycobacterium tuberculosis* isolates to pyrazinamide and fluoroquinolones: results from a multicountry surveillance project. *The Lancet Infectious Diseases.* 2016; 16: 1185–1192. [https://doi.org/10.1016/S1473-3099\(16\)30190-6](https://doi.org/10.1016/S1473-3099(16)30190-6) PMID: 27397590
3. Werngren J, Alm E, Mansjo M. Non-pncA Gene-Mutated but Pyrazinamide-Resistant *Mycobacterium tuberculosis*: Why Is That? *J Clin Microbiol.* 2017; 55: 1920–1927. <https://doi.org/10.1128/JCM.02532-16> PMID: 28404681.
4. Hoffner S, Angeby K, Sturegard E, Jonsson B, Johansson A, Sellin M, et al. Proficiency of drug susceptibility testing of *Mycobacterium tuberculosis* against pyrazinamide: the Swedish experience. *Int J Tuberc Lung Dis.* 2013; 17: 1486–1490. <https://doi.org/10.5588/ijtld.13.0195> PMID: 24125455.
5. Zhang Y, Permar S, Sun Z. Conditions that may affect the results of susceptibility testing of *Mycobacterium tuberculosis* to pyrazinamide. *J Med Microbiol.* 2002; 51: 42–49. <https://doi.org/10.1099/0022-1317-51-1-42> PMID: 11800471.
6. Schon T, Miotto P, Koser CU, Viveiros M, Bottger E, Cambau E. *Mycobacterium tuberculosis* drug-resistance testing: challenges, recent developments and perspectives. *Clin Microbiol Infect.* 2017; 23: 154–160. <https://doi.org/10.1016/j.cmi.2016.10.022> PMID: 27810467.
7. Ramirez-Busby SM, Valafar F. Systematic review of mutations in pyrazinamidase associated with pyrazinamide resistance in *Mycobacterium tuberculosis* clinical isolates. *Antimicrob Agents Chemother.* 2015; 59: 5267–5277. <https://doi.org/10.1128/AAC.00204-15> PMID: 26077261.
8. Zhang Y, Yew WW. Mechanisms of drug resistance in *Mycobacterium tuberculosis*. *Int J Tuberc Lung Dis.* 2009; 13: 1320–1330. PMID: 19861002.
9. Singh P, Mishra AK, Malonia SK, Chauhan DS, Sharma VD, Venkatesan K, et al. The paradox of pyrazinamide: an update on the molecular mechanisms of pyrazinamide resistance in *Mycobacteria*. *J Commun Dis.* 2006; 38: 288–298. PMID: 17373362.
10. Iwai H, Kato-Miyazawa M, Kirikae T, Miyoshi-Akiyama T. CASTB (the comprehensive analysis server for the *Mycobacterium tuberculosis* complex): A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. *Tuberculosis (Edinb).* 2015; 95: 843–844. <https://doi.org/10.1016/j.tube.2015.09.002> PMID: 26542225.
11. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a Web Tool Delineating *Mycobacterium tuberculosis* Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. *J Clin Microbiol.* 2015; 53: 1908–1914. <https://doi.org/10.1128/JCM.00025-15> PMID: 25854485.

12. Coll F, McNerney R, Preston MD, Guerra-Assuncao JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med.* 2015; 7: 51. <https://doi.org/10.1186/s13073-015-0164-0> PMID: 26019726.
13. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun.* 2015; 6: 10063. <https://doi.org/10.1038/ncomms10063> PMID: 26686880.
14. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, Kato S, et al. TGS-TB: Total Genotyping Solution for *Mycobacterium tuberculosis* Using Short-Read Whole-Genome Sequencing. *PLoS One.* 2015; 10: e0142951. <https://doi.org/10.1371/journal.pone.0142951> PMID: 26565975.
15. Schleusener V, Koser CU, Beckert P, Niemann S, Feuerriegel S. *Mycobacterium tuberculosis* resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools. *Sci Rep.* 2017; 7: 46327. <https://doi.org/10.1038/srep46327> PMID: 28425484.
16. van Beek J, Haanpera M, Smit PW, Mentula S, Soini H. Evaluation of whole genome sequencing and software tools for drug susceptibility testing of *Mycobacterium tuberculosis*. *Clin Microbiol Infect.* 2018. <https://doi.org/10.1016/j.cmi.2018.03.041> PMID: 29653190.
17. Macedo R, Nunes A, Portugal I, Duarte S, Vieira L, Gomes JP. Dissecting whole-genome sequencing-based online tools for predicting resistance in *Mycobacterium tuberculosis*: can we use them for clinical decision guidance? *Tuberculosis (Edinb).* 2018; 110: 44–51. <https://doi.org/10.1016/j.tube.2018.03.009> PMID: 29779772.
18. Mitarai S, Kato S, Ogata H, Aono A, Chikamatsu K, Mizuno K, et al. Comprehensive multicenter evaluation of a new line probe assay kit for identification of *Mycobacterium* species and detection of drug-resistant *Mycobacterium tuberculosis*. *J Clin Microbiol.* 2012; 50: 884–890. <https://doi.org/10.1128/JCM.05638-11> PMID: 22205814.
19. Wayne LG. Simple pyrazinamidase and urease tests for routine identification of mycobacteria. *Am Rev Respir Dis.* 1974; 109: 147–151. <https://doi.org/10.1164/arrd.1974.109.1.147> PMID: 4203284.
20. Xia Q, Zhao LL, Li F, Fan YM, Chen YY, Wu BB, et al. Phenotypic and genotypic characterization of pyrazinamide resistance among multidrug-resistant *Mycobacterium tuberculosis* isolates in Zhejiang, China. *Antimicrob Agents Chemother.* 2015; 59: 1690–1695. <https://doi.org/10.1128/AAC.04541-14> PMID: 25583712.
21. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant.* 2013; 48: 452–458. <https://doi.org/10.1038/bmt.2012.244> PMID: 23208313.
22. Miotto P, Cabibbe AM, Feuerriegel S, Casali N, Drobniewski F, Rodionova Y, et al. *Mycobacterium tuberculosis* pyrazinamide resistance determinants: a multicenter study. *MBio.* 2014; 5: e01819–01814. <https://doi.org/10.1128/mBio.01819-14> PMID: 25336456.
23. Yadon AN, Maharaj K, Adamson JH, Lai YP, Sacchetti JC, Joerger TR, et al. A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide. *Nat Commun.* 2017; 8: 588. <https://doi.org/10.1038/s41467-017-00721-2> PMID: 28928454.
24. Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE 3rd, et al. Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into the emergence and spread of multidrug resistance. *Nat Genet.* 2017; 49: 395–402. <https://doi.org/10.1038/ng.3767> PMID: 28092681.
25. Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, et al. A standardised method for interpreting the association between mutations and phenotypic drug resistance in *Mycobacterium tuberculosis*. *Eur Respir J.* 2017; 50. <https://doi.org/10.1183/13993003.01354-2017> PMID: 29284687.
26. Farhat MR, Sultana R, Iartchouk O, Bozeman S, Galagan J, Sisk P, et al. Genetic Determinants of Drug Resistance in *Mycobacterium tuberculosis* and Their Diagnostic Value. *Am J Respir Crit Care Med.* 2016; 194: 621–630. <https://doi.org/10.1164/rccm.201510-2091OC> PMID: 26910495.
27. Rosen BC, Dillon NA, Peterson ND, Minato Y, Baughn AD. Long-Chain Fatty Acyl Coenzyme A Ligase FadD2 Mediates Intrinsic Pyrazinamide Resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother.* 2017; 61. <https://doi.org/10.1128/AAC.02130-16> PMID: 27855077.
28. Bouzouita I, Cabibbe AM, Trovato A, Draoui H, Ghariani A, Midouni B, et al. Is sequencing better than phenotypic tests for the detection of pyrazinamide resistance? *Int J Tuberc Lung Dis.* 2018; 22: 661–666. <https://doi.org/10.5588/ijtld.17.0715> PMID: 29862951.
29. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat Genet.* 2018; 50: 307–316. <https://doi.org/10.1038/s41588-017-0029-0> PMID: 29358649.
30. Liu W, Chen J, Shen Y, Jin J, Wu J, Sun F, et al. Phenotypic and genotypic characterization of pyrazinamide resistance among multidrug-resistant *Mycobacterium tuberculosis* clinical isolates in Hangzhou, China. *Clin Microbiol Infect.* 2017. <https://doi.org/10.1016/j.cmi.2017.12.012> PMID: 29288021.

31. Iwamoto T, Yoshida S, Suzuki K, Wada T. Population structure analysis of the *Mycobacterium tuberculosis* Beijing family indicates an association between certain sublineages and multidrug resistance. *Antimicrob Agents Chemother.* 2008; 52: 3805–3809. <https://doi.org/10.1128/AAC.00579-08> PMID: 18694954.
32. Iwamoto T, Fujiyama R, Yoshida S, Wada T, Shirai C, Kawakami Y. Population structure dynamics of *Mycobacterium tuberculosis* Beijing strains during past decades in Japan. *J Clin Microbiol.* 2009; 47: 3340–3343. JCM.01061-09 [pii] <https://doi.org/10.1128/JCM.01061-09> PMID: 19710282.
33. Chedore P, Bertucci L, Wolfe J, Sharma M, Jamieson F. Potential for erroneous results indicating resistance when using the Bactec MGIT 960 system for testing susceptibility of *Mycobacterium tuberculosis* to pyrazinamide. *J Clin Microbiol.* 2010; 48: 300–301. <https://doi.org/10.1128/JCM.01775-09> PMID: 19923479.
34. Shi W, Chen J, Zhang S, Zhang W, Zhang Y. Identification of Novel Mutations in LprG (rv1411c), rv0521, rv3630, rv0010c, ppsC, and cyp128 Associated with Pyrazinoic Acid/Pyrazinamide Resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother.* 2018; 62. <https://doi.org/10.1128/AAC.00430-18> PMID: 29686155.
35. Kandler JL, Mercante AD, Dalton TL, Ezewudo MN, Cowan LS, Burns SP, et al. Validation of Novel *Mycobacterium tuberculosis* Isoniazid Resistance Mutations Not Detectable by Common Molecular Tests. *Antimicrob Agents Chemother.* 2018; 62. <https://doi.org/10.1128/AAC.00974-18> PMID: 30082293.
36. Satta G, Atzeni A, McHugh TD. *Mycobacterium tuberculosis* and whole genome sequencing: a practical guide and online tools available for the clinical microbiologist. *Clin Microbiol Infect.* 2017; 23: 69–72. <https://doi.org/10.1016/j.cmi.2016.09.005> PMID: 27642177.
37. Rhoads DD. Lowering the Barriers to Routine Whole-Genome Sequencing of Bacteria in the Clinical Microbiology Laboratory. *J Clin Microbiol.* 2018; 56. <https://doi.org/10.1128/JCM.00813-18> PMID: 29950328.