微生物学报
Acta Microbiologica Sinica
2021, 61(11): 3653–3666
http://journals.im.ac.cn/actamicrocn
DOI: 10.13343/j.cnki.wsxb.20210095

Research Article 研究报告

结核分枝杆菌耐药相关单碱基突变的计算方法比较

李志远[1,2]，刘飞[2]，曹德民[1,2]，潘元龙[2]，律娜[2]，刘东鑫[3]，贺文从[3]，程城[2]，王亚楠[2]，赵雁林[3]，朱宝利[1,2*]

[1] 中国科学院大学，北京 100049

[2] 中国科学院微生物研究所，病原微生物与免疫学重点实验室，北京 100101

[3] 中国疾病预防控制中心结核病预防控制中心，国家结核病参比实验室，北京 102206

摘要：【目的】耐药结核分枝杆菌(drug-resistant *Mycobacterium tuberculosis*)的产生给结核病(tuberculosis)的治疗带来巨大困难。【方法】使用基于全基因组测序的关联分析探究耐药强相关的单核苷酸多态性(single nucleotide polymorphism，SNP)突变，主要有 GEMMA、phyc、plink。为了阐明其中最优的耐药相关 SNP 计算方法，本研究下载 NCBI 上已有的 1504 株结核分枝杆菌数据，并获取它们对于 3 种常见的一线抗结核治疗药物(isoniazid、rifampicin、ethambutol)的耐药性检验结果。并使用这 3 种耐药相关 SNP 计算方法计算与结核分枝杆菌耐药相关的 SNP；并评估计算得到的耐药相关 SNP 在预测耐药表型的敏感性和特异性。【结果】发现通过 phyc 可以预测到最多的已知耐药相关 SNP 和最少的耐药无关 SNP，而且 phyc 预测的耐药相关 SNP 的敏感性和特异性恒定大于 52.49%。【结论】phyc 在预测结核分枝杆菌耐药相关 SNP 中结果最准确，但考虑到运行时间和表型数据的更新，GEMMA 和 plink 的结果也应作为参考。

关键词：结核分枝杆菌，耐药基因，比较基因组学

结核病是人类公共卫生的重要威胁，并在 2014 年超过艾滋病成为致死人数最多的传染病，全世界约 20 亿人口患潜伏性结核病，并有 5%–10%的患者发病[1–2]。结核分枝杆菌是结核病的致病原[3]，能在干燥状态下存活数周[4]。根据世界卫生组织在 2021 年 1 月 27 日生效的公告，耐多药结核分枝杆菌对 rifampicin 和 isoniazid 都具有耐药性[5]；广泛耐药结核分枝杆菌(XDR-TB)是在满足耐多药结核分枝杆菌的前提下，对 fluoroquinolones 类抗生素以及至少一种 A 类抗生

基金项目：国家重点研发计划(2018YFC1603803)；中国科学院战略性先导科技专项(XDB29020203)；北京市科技计划(Z201100005520041)

*通信作者。Tel/Fax：+86-10-64807362；E-mail：zhubaoli@im.ac.cn

收稿日期：2021-02-09；**修回日期：**2021-05-31；**网络出版日期：**2021-09-14

素 (levofloxacin 、 moxifloxacin 、 bedaquiline 、 linezolid)有耐药性[6]。耐多药结核分枝杆菌致使全球每年新增至少 50 万例患者。根据世界卫生组织的评估，2018 年中国结核病患者和耐药结核分枝杆菌携带者的人数分别为 866000 和 66000 左右，均居世界第二[7]；耐药结核病患者的临床疗程更长，副作用更多，治愈率只有 50%左右[8–9]。

结核分枝杆菌基因组为单拷贝，突变率低，基本不发生水平基因转移和菌株间的基因重组，耐药性主要由耐药基因突变引起(其中主要是 SNP)，适合基于全基因组测序分析建立耐药表型和基因型的联系[10–11]。基于全基因组序列的分析可以有效地挖掘新的耐药相关突变和基因，进而基于这些突变预测临床中菌株的耐药性和亚型[12–14]，替代周期过长的基于培养的耐药性测试和分型实验[15]，为临床诊断和用药提供指导[11]，并且可以结合流行病学数据分析耐药菌的传播规律[13,16]；目前整理的耐药相关突变对一线抗生素中 isoniazid、rifampicin、ethambutol、streptomycin 的耐药性预测的准确度(耐药菌中能被正确预测的百分比)和特异性(携带耐药突变的菌中表型为耐药的百分比)大于 90%[17]，但因数据量有限，对 pyrazinamide 的准确度只能达到 60%左右[18]。

基于全基因组测序的全基因组关联分析(Genome-wide association analysis，GWAS)及其近似的方法已经有效挖掘了大量与抗菌药物耐药有关的突变[19–20]，主要通过对每一个突变在抗性菌株和敏感菌株中的数量的比率进行 Fisher 精确性检验，计算某突变是否和耐药表型显著相关[19]。但是，研究发现群体结构相关的来自不同祖先菌株的变异可能干扰耐药突变的挖掘，而且常规的去除群体不平衡的方法会增大统计学的误差[21]。目前有 2 种算法已被用于去除群体结构有关的变异，GEMMA (genome-wide efficient mixed model association)基于多元线性回归模型，将样本间的遗传相关性视为一个参量，进而去除群体结构带来的假阳性[22]；phyc (phylogenetic convergence test)是基于病原菌受到耐药胁迫后倾向于向某方向收敛进化的原理，通过统计进化树上发生了突变的分支数和未发生突变的分支数在耐药和敏感菌株中的比例，去除了偶然中性进化的干扰，得到了由于收敛进化和正选择获得的抗性突变[23]。另外，计算耐药相关突变的方法还有随机森林算法(random forest)，一种基于多个分类器的机器学习方法[24]，根据耐药菌株和敏感菌株中有显著差别的突变、基因、基因间区构建随机森林分类器；正态分布(normal distribution)算法假设在药物的选择压力下和抗性相关的突变类型在抗性菌株中显著地多[25]，因此选择在抗性菌株中数目显著高于敏感菌株的 SNP、基因、基因间区域视为与抗性相关。

考虑到结核分枝杆菌耐药性积累的严峻性，本研究旨在寻找具有更高准确度和特异性的耐药相关 SNP 计算方法，为临床用药、耐药机制探究、流行病学调查提供指导。

1 材料和方法

1.1 样本分布

本文样本序列和表型均来自于已上传数据 PRJEB10385、ERP002611、PRJNA436454 和已发表文章[19,26–28]，因为只有 isoniazid、rifampicin、ethambutol 三种一线抗生素能找到敏感相关 SNP 的数据，并且缺失表型较少(表 1)，本文针对这 3 种抗生素表型进行分析。另外，菌株耐药表型以耐多药为主(图 1)。

表 1. 样本家系及其表型分布

Table 1. Distribution of the lineage and phenotype of strains

Lineage	Phenotype	Isoniazid	Rifampicin	Ethambutol
Lineage 4	Sensitive	488	549	743
	Resistant	423	362	152
	NA	0	0	14
Lineage 2	Sensitive	137	132	332
	Resistant	456	461	248
	NA	0	0	13

NA: the phenotype is default.

1.2 序列质量控制及比对

从 NCBI 的 SRA 数据库下载得到 fastq 文件后，使用 Trimmomatic 和 fastp 的默认参数[29]去除低质量及引物序列。使用 BWA[30]将序列比对到参考基因组 H37Rv(RefSeq:NC_000962.3)的序列上，使用 Speedseq[31]定位突变。使用 vcftools[32]将序列合并成 vcf 格式，将序列上与参考基因组一致的序列从”.”转换成”0/0”，按照缺失率<0.01、最小等位基因频率高于 0.001 的条件筛选 SNP (表 2)。使用 fasttree 依据最大似然法原理构建进化树[33]，使用 itol 绘制和注释进化树[34]。

1.3 耐药相关 SNP 计算

计算使用的电脑 CPU 配置均为 Intel Xeon Gold 6126 (2.6GHz 12C)。依据测序数据去除 PEPPE 重复序列，使用 GEMMA[22]实现 GEMMA 算法分析，选取 *P* 值>0.05 作为阈值；使用 plink

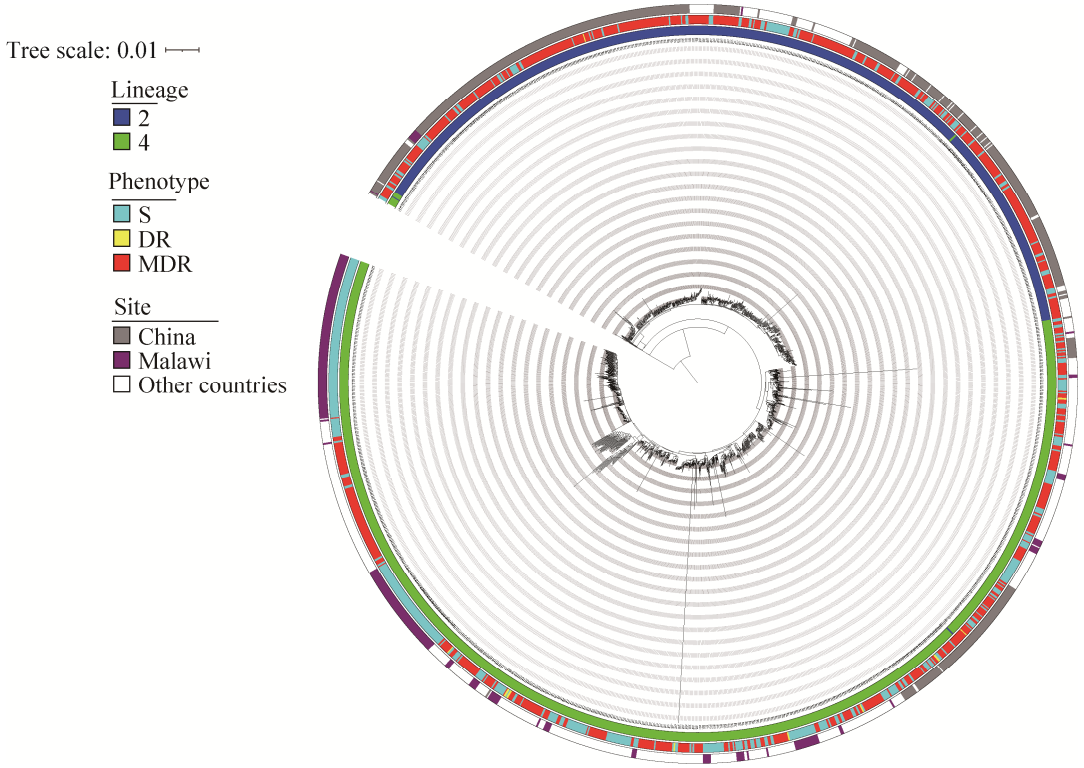

图 1. 样本表型在进化树上的分布

Figure 1. Distribution of the lineage and phenotype of strains on phylogenetic tree. In the inner annotation circle, indigo bar means susceptible strains, yellow bar means drug-resistant strains, red bar means multi drug-resistant strains. In the middle circle, blue and green bar means strains from lineage 2 and 4. In the outer circle, grey, purple and white bar means the strains acquired in China, Malawi and other countries. *M. canettii* is added to set up the root of phylogenetic tree.

软件实现基于 *t* 分布的 wald 检验[35]，选取 *P* 值>0.01 和 OR 值>1 作为阈值；使用 R 语言的 ape 包构建最大简约法进化树[36]，使用“phangorn”包中的 pratchet 和 ancestral.pars 工具执行祖源序列重建[37]，而后基于此前的方法进行 phyc 分析[23]，选取 *P* 值<0.05 作为阈值。执行完这些计算工具后，只保留在已知和 isoniazid、rifampicin、ethambutol 耐药相关的基因上出现的 SNP (表 2)。

2 结果和分析

2.1 样本基因组系统发育分析

基于 1504 株结核分枝杆菌突变频次大于 2 的 SNP 构建系统发育树，添加 *M. canettii* 基因组作为外群建立树根；本研究发现地理位置较为接近的基因组数据在进化树上成簇分布，比如样本较多的 Malawi 地区的内部菌株在进化树上的位置较近；另外，家系 2 在我国分布广泛，而家系 4 在我国分布相对较少。可见不同地区和家系的菌株进化距离较远，携带很多与耐药表型无关的突变(图 1)。

2.2 3 种抗生素杀伤机制及其已知耐药相关 SNP

根据已有实验和测序分析的文章，确认和耐药有直接关联的基因及其 SNP，它们主要来自 TBProfiler 预测软件的突变数据集[38]、Tim walker 团队预测 MTB 耐药性的文章对应的突变[39](组成了两款预测软件 Mykrobe 中的突变数据库)、Tim walker 团队自己整理的突变数据集(目前未公开)，本研究取 3 个数据的交集作为已知耐药相关 SNP；考虑到只有 isoniazid、rifampicin、ethambutol 在数据集中具有耐药无关突变，本研究取这 3 种抗生素进行评估(表 2)。isoniazid 可以通过 katG 基因编码的过氧化氢酶催化形成异烟酸，进而和 NADH 结合成复合体，作为底物结构类似物阻断脂肪酸合成酶执行催化功能，抑制细胞壁的合成[40–43]。rifampicin 可以通过疏水性的细胞膜与 DNA 依赖的 *rpoB* 编码的 RNA 聚合酶 β 亚基结合，抑制了 RNA 聚合酶活性，进而阻止 DNA 转录合成 RNA 并破坏细胞生理功能[44–46]。ethambutol 可以作为阿拉伯糖结构类似物与 *embA*、*embB*、*embC* 基因编码的阿拉伯糖转移酶结合，阻止了组成细胞壁必需的阿拉伯半乳聚糖的合成[47–49]。这 3 种抗生素耐药相关 SNP 大多与激活抗生素有关的酶结构改变、抗生素靶点的改变、抗生素靶向蛋白表达量的改变有关(表 2)。

2.3 3 种不同的耐药相关 SNP 计算方法比较

在耐药相关 SNP 的显著性检验结果中，对于位于异烟肼、利福平、乙胺丁醇相关的 *katG*、*rpoB*、*embA* 基因上的已知耐药相关 SNP，*katG*Ser315Gly、Trp191Arg、*rpoB* Gln432Lys 突变只在 phyc 方法的结果中显示和耐药显著相关，而 GEMMA 和 plink 则显示其与耐药不显著相关。在 *embA*-43G>C、Met306Val 以及 *rpoB* Asp435Gly、

表 2. 三种抗生素杀菌机制和耐药相关 SNP

Table 2. Mechanisms of three antibiotics killing bacteria and their resistant-related SNPs

Antibiotics	Resistant-related genes	Resistant-related SNPs	Resistant-unrelated SNPs	Reference
Isoniazid	*fabG1*, *inhA*, *katG*, *kasA*, *ahpC*, *ndh*	323	100	[40–43]
Rifampicin	*rpoB*, *rpoC*	145	61	[44–46]
Ethambutol	*embA*, *embC*, *embR*, *embB*	193	321	[47–49]

His445Leu、Ser441Leu、Ile491Phe、Ser450Trp、Gln432Lys 中只有 plink 无法被计算出和耐药显著相关，但 phyc 得到的 *P* 值依然远小于 GEMMA (表 3)。另外，本研究还新发现 *pks12* Ile2261Val 与异烟肼耐药有关，发现 *serA1* Ala302Gly 和 *pyrG* Gly576Ala 突变与乙胺丁醇耐药有关，发现 *fusA1* Lys359Arg 和 Lys359Glu 与利福平耐药有关。

本研究通过统计 3 种方法计算出的耐药相关突变在耐药菌和敏感菌中占据的比例，评估这些 SNP 在预测菌株耐药表型中的准确度和特异性，发现其在预测 isoniazid 耐药相关 SNP 中表现差异不大，但 GEMMA 和 phyc 的敏感度在 rifampicin 和 ethambutol 中显著高于 plink (图 2，表 3)，并且 GEMMA 和 phyc 可以发现更多的已知耐药相关 SNP；phyc 计算得到的乙胺丁醇相关的 SNP 在保证特异性最高的情况下保证了最高的准确性 (图 2，表 3)。

通过比较 3 种计算方法的原理和计算时间，发现 GEMMA 和 plink 的原理较为接近，都是将耐药表现型视为基因型和随机参数的函数，当 SNP 在耐药菌株中频率显著高于敏感菌株时，则视为与耐药有关。不同的是 GEMMA 还计算了遗传矩阵用于去除平衡群体分层带来的耐药无关突变。而 phyc 计算原理中考虑了抗生素积累耐药中的收敛进化，计数进化树上形成的子分支簇是否多为耐药菌，若包含某 SNP 的进化分支构成的簇

表 3. 不同的计算方法中算出的已知和新耐药相关 SNP

Table 3. Significance test of known and novel resistant-related SNPs

Antibiotics	Gene	SNP type	Mutation frequency in resistant strains/%	Mutation frequency in susceptibe strains/%	Phyc	GEMMA	Plink
Known resistant-related SNPs							
Isoniazid	*katG*	p.Trp191Arg	0.910	0.160	0.00590	NA	0.00759
		p.Ser315Gly	0.429	0	0.00479	NA	NA
Rifampicin	*rpoB*	p.Asp435Gly	1.94	0	2.11E-10	0.011	NA
		p.His445Leu	1.94	0	1.37E-08	0.011	NA
		p.Ser441Leu	0.97	0	8.97E-07	0.005	NA
		p.Ser450Trp	1.09	0	3.61E-06	0.006	NA
		p.Gln432Lys	0.459	0.147	0.0152	NA	NA
		p.Ile491Phe	0.573	0	0.0153	0.003	NA
Ethambutol	*embA*	c.-43G>C	2.75	0	5.69E-09	5.730E-08	NA
		p.Met306Val	33.5	4.460	7.34E-47	3.346E-54	3.99E-87
Novel resistant-related SNPs							
Isoniazid	*pks12*	Ile2261Val	6.37	0.410	0.0599	8.71E-28	3.12E-13
Rifampicin	*fusA1*	Lys359Arg	2.92	0.599	1.42E-11	7.40E-5	1.99E-8
		Lys359Glu	3.04	0.699	2.12E-11	1.05E-4	3.52E-8
Ethambutol	*serA1*	Ala302Gly	3.00	0.783	2.57E-8	2.00E-5	1.67E-3
	pyrG	Gly576Ala	14.0	8.690	2.84E-5	4.84E-4	7.22E-6

NA: the *P* value is not significant or the OR value in plink is less than 1.

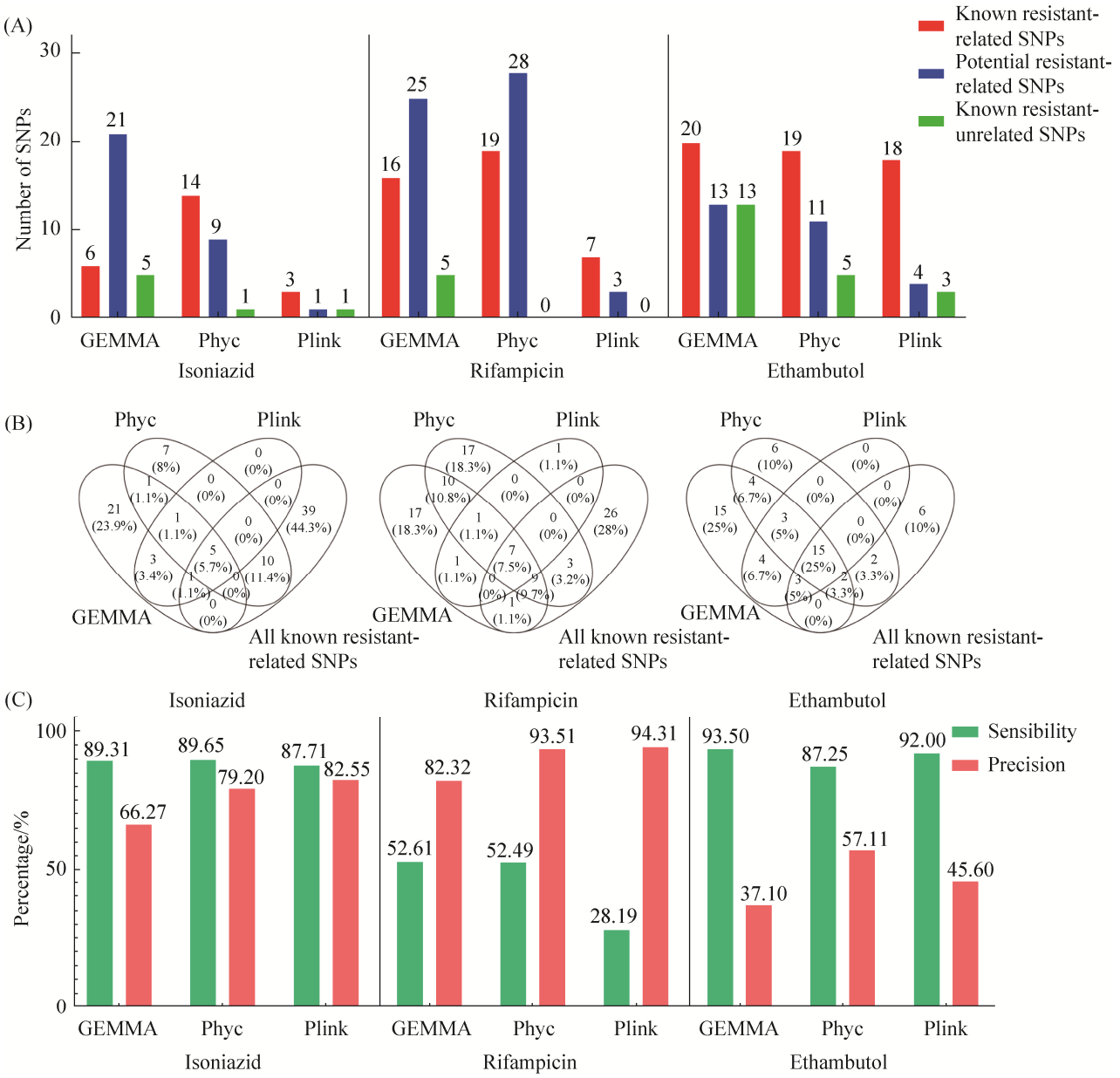

图 2. GEMMA、phyc、plink 方法计算所得耐药 SNP 及其敏感性和精确性

Figure 2. The sensitivity and accuracy of drug-resistant SNP calculated by GEMMA, phyc and plink methods. A: Number of SNPs acquired by GEMMA, phyc and plink method; B: Number of SNPs acquired by GEMMA, phyc and plink method and their overlap with all resistant-related SNPs in database; C: Sensitivity and precision of the SNPs acquired by GEMMA, phyc and plink method. Sensitivity is the proportion of the strains with resistant-related SNPs in all resistant strains. Precision is the proportion of resistant strains in the strains with the resistant-related SNPs.

多为耐药表型，则某 SNP 与耐药有关。这也决定了 phyc 计算中只能将表现型视为耐药或非耐药菌，难以统计连续值表现型，而 GEMMA 和 plink 可以；另外，phyc 在计算中需要使用频次大于 2 或 3 的 SNP 构建进化树和大量随机抽样，这使得其消耗的时间远大于 GEMMA 和 plink (表 4)。

表 4. 不同方法原理的比较

Table 4. Comparison of three methods for the calculation of resistant-related SNPs

Method	Method of eliminating population stratification	Calulation time	Phenotype
Plink	No	10 min	DST/MIC
GEMMA	Calculating sequence similarity matrix	10 min	DST/MIC
Phyc	Using convergence as the signal of positive selection and the specificity of convergence to cases	144 h	DST

MIC: the result of minimum inhibitory concentration. DST means the result of drug susceptibility testing.

3 讨论

本文通过整理1504株结核分枝杆菌的序列及表型数据，使用 3 种常用的方法计算耐药相关SNP，比较了可以找到的已知耐药相关 SNP、已知耐药无关SNP的数目，以及用这些SNP重新推断原菌株耐药表型时的敏感性和特异度，发现phyc 在计算异烟肼、利福平、乙胺丁醇耐药相关SNP时可以算出更多已知耐药相关SNP或基本持平，而且phyc计算得到的耐药相关SNP具有更稳定的敏感性和特异度(恒定大于 52.49%)；虽然GEMMA 和 plink 得到的耐药相关 SNP 对应的敏感性和特异度有时也最高，但是也会出现低于40%的情况。由于GEMMA也考虑了遗传关联性，排除了一些潜在的群体分层等带来的假阳性，因此在乙胺丁醇中找到的已知耐药相关 SNP 数和phyc找到的SNP数持平(图2)；而plink虽然也可以在特异性中表现接近GEMMA和phyc，但由于没有考虑基因组的关联，只是对等位基因频率统计学分析，计算出的已知耐药相关SNP数均最少(图2，表3)。

另外，GEMMA和plink会将大量已知和耐药显著相关的SNP计算为和耐药不显著相关，比如*katG* Ser315Gly 、*rpoB* Gln432Lys 、*rpoB* Glu460Gly、*embA* -43G>C、*rpoB* Ser450Trp、*rpoB* Ile491Phe。实验和测序证据证明已知这些SNP均和耐药的发生有关联，其中*katG* Ser315Gly位于激活异烟肼的酶的活性 SNP，已有研究说明其与异烟肼耐药的形成显著相关[50]；rpoB Gln432Lys、Glu460Gly、Ile491Phe等位于RNA聚合酶的编码区，可能与 RNA 聚合酶的结构有关[51]；embA-43G>C位于乙胺丁醇靶向的阿拉伯糖转移酶的启动子区，可能通过提高表达强度干扰抗生素的抑制[52]；已有的全基因组关联分析已经算出本研究提供的耐药相关SNP与耐药有关[14,39]。phyc的表现更好可能与其在计算时直接计数进化形成的耐药菌株簇并且把进化距离较近且表现型相同的冗余分支合并，而非如同 plink 和 GEMMA 单纯统计全部耐药和敏感菌株携带SNP的数目多少[23]。

本研究预测的新耐药相关SNP大多与代谢过程或细胞膜、细胞壁的合成有关(表 3)。*serA1* 在结核分支杆菌中负责丝氨酸代谢，在生长中必不可少，SNP可能通过改变蛋白结构影响酶活，进而影响丝氨酸代谢[53]。*pyrG* 是重要的CTP合成酶[54]，此前文章通过全基因组关联分析和热耗散模型推测 *pyrG* 可能与耐药有关[55]。*fusA1* 是锚定蛋白靶向的延伸因子(argyrin B targets elongation factor G)，其氨基酸突变(如P443L)等被发现可以造成外排泵、生物膜相关基因表达量改变，并且与多种革兰氏阴性致病菌的氨基糖苷类、妥布霉素等抗生素耐药有关[56–57]，目前没有关于导致结核分枝杆菌耐药的报道。*pks12* 表达聚酮合酶，负责催化

合成结核分枝杆菌细胞壁必需的脂类(如结核菌醇 dimycocerosyl phthiocerol)，已被证明与固有耐药有关[58]。*CaeA* 编码和细胞膜形成有关的蛋白酶，被证明与结核分枝杆菌固有耐药有关。

考虑到 phyc 方法适用表型数据有限和计算时间过长，其并非可以完全替代 GEMMA 和 plink。同等配置下 phyc 计算时间需要 1 周左右，远长于 GEMMA 和 plink (表 4)，而且在耐药菌株样本量达到 879 的异烟肼中，敏感性、特异性和计算出的已知耐药相关 SNP 数目区别不大(图 2)，所以如果样本总量超过了 1000，可以在等待 phyc 计算结果的同时，先使用 GEMMA、plink 初步算出频率较高的耐药相关 SNP。虽然 phyc 在敏感性和特异性上表现均较好，但考虑到其在分析中必须先根据临界浓度定义进化树上的耐药分支和敏感分支[23,59]，目前无法考虑到最小抑菌浓度(minimum inhibitory concentration，MIC)的连续值数据。此前研究中使用的药物敏感性测试(drug susceptibility testing，DST)主要是基于临界抗生素浓度的药物敏感性测试，它曾是世界卫生组织和部分结核分枝杆菌有关研究采用的耐药性检测标准[20,60–62,48]；但是世界卫生组织和研究者认为临界抗生素浓度的设定基于经验，缺乏科学依据[63–64]，“决定耐药性的临界浓度通常非常接近实现抗分枝杆菌活性所需的最低抑菌浓度，这增加了敏感和耐药菌株误分类的可能性，并导致药物敏感性测试结果的可重复性差”[65]；后来采用的微孔板法最小抑菌浓度测定虽然更昂贵但更加准确[66–67]。因此本研究认为若有最小抑菌浓度的数据，使用 phyc 时还需要用 GEMMA、plink 软件加以验证，或者根据最小抑菌浓度大小将耐药菌分为低水平耐药和高水平耐药，并在 phyc 软件定义耐药分支时将其分为高水平耐药分支和低水平耐药分支。

本文中部分已知耐药相关 SNP 在 3 种计算方法的结果中均显示与耐药无关(图 2)，而且计算得到的新耐药相关突变较少(表 3)，可能是由于耐药菌样本量大小有限，而且耐药突变的计算容易受到群体分层的干扰[68]，系统发育树显示 Malawi 地区的菌株在进化关系上较为接近，可能和它们祖先菌株有关的 SNP 被误认为与耐药有关(图 1)，因此导致了全基因组关联分析存在偏差。考虑到耐药基因常常受到更高的进化选择压力，在之后的研究中可以结合进化选择压力(非同义突变除以同义突变率)批量检索受到正选择压力的基因或区域，进而降低预测耐药相关突变的假阳性，进一步确认找到的 SNP 是耐药相关 SNP[21]；其次，多个 SNP 的突变以及菌株遗传背景对耐药性的共同影响也应被重视[69]，本研究发现和 ethambutol 强相关的突变在敏感菌中也有出现(图 2)，文献中也指出单个 SNP 的突变(如 *embB* 第 306 个氨基酸)并不会确保一定获得高水平的 ethambutol 耐药性[70]，即使是和异烟肼耐药性显著相关的 *katG* 的第 315 个氨基酸也存在携带突变的敏感菌[39,71]；也可以使用基于蛋白结构建模的计算方法探究突变对蛋白结构的影响[72]。

综上所述，计算结核分枝杆菌耐药相关 SNP 并保证最高准确性的合理策略是：尽可能下载已公开的结核分枝杆菌基因组和表现型数据，尽量保证来自不同地区或者家系的样本数目平衡，使用最小抑菌浓度测定表型，计算耐药相关性大小时先运行 phyc 方法，在等待计算结果时使用 GEMMA 和 plink 计算；使用已知耐药相关和耐药无关 SNP 确定合理的 *P* 值阈值和突变优势比，尽力避免假阳性和假阴性，同时计算新耐药相关基

因和突变的进化选择压力大小；对进化关系和表型有差异的组之间进行主成分分析，去除取样时潜在的群体分层现象和家系相关 SNP 对耐药相关 SNP 预测结果的干扰。

参 考 文 献

[1] Lv L, Li C, Zhang X, Ding N, Cao T, Jia X, Wang J, Pan L, Jia H, Li Z, Zhang J, Chen F, Zhang Z. RNA profiling analysis of the serum exosomes derived from patients with active and latent *Mycobacterium tuberculosis* infection. *Frontiers in Microbiology*, 2017, 8: 1051.

[2] Gonzalo-Asensio, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, Laval F, Daffé M, Martin C, Brosch R, Guilhot C. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(31): 11491–11496.

[3] Gordon SV, Parish T. Microbe profile: *Mycobacterium tuberculosis*: humanity's deadly microbial foe. *Microbiology*, 2018, 164(4): 437–439.

[4] Parish T, Stoker NG. *Mycobacteria*: bugs and bugbears (two steps forward and one step back). *Molecular Biotechnology*, 1999, 13(3): 191–200.

[5] GT Programme, Meeting report of the WHO expert consultation on the definition of extensively drug-resistant tuberculosis. 2021, World Health Organization.

[6] Palomino JC, Martin A. Drug resistance mechanisms in *Mycobacterium tuberculosis*. *Antibiotics*: *Basel*, *Switzerland*, 2014, 3(3): 317–340.

[7] Global tuberculosis report 2019. World Health Organization: 2019.

[8] HF Roh, Kim J, Nam SH, Kim JM, Pulmonary resection for patients with multidrug-resistant tuberculosis based on survival outcomes: a systematic review and meta-analysis. *European Journal of Cardio-thoracic Surgery*, 2017, 52(4): 673–678.

[9] Marks SM, Flood J, Seaworth B, Hirsch-Moverman Y, Armstrong L, Mase S, Salcedo K, Oh P, Graviss EA, Colson PW, Armitige L, Revuelta M, Sheeran K, Consortium TTES. Treatment practices, outcomes, and costs of multidrug-resistant and extensively drug-resistant tuberculosis, United States, 2005–2007. *Emerging Infectious Diseases*, 2014, 20(5): 812–821.

[10] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998, 393(6685): 537–544.

[11] Coll F, McNerney R, Preston MD, Guerra-Assun{c}ao JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, Perdigao J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, Pain A, Clark TG. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Medicine*, 2015, 7(1): 51.

[12] Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nature Communications*, 2014, 5: 4812.

[13] Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, Farhat MR, Guthrie JL, Laukens K, Miotto P, Ofori-Anyinam B, Dreyer V, Supply P, Suresh A, Utpatel C, van Soolingen D, Zhou Y, Ashton PM, Brites D, Cabibbe AM, Jong BC, de Vos M, Menardo F, Gagneux S, Gao Q, Heupink TH, Liu QY, Loiseau C, Rigouts L, Rodwell TC, Tagliani E, Walker TM, Warren RM, Zhao YL, Zignol M, Schito M, Gardy J, Cirillo DM, Niemann S, Comas I, Van Rie A. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nature Reviews Microbiology*, 2019, 17(9): 533–545.

[14] Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *The Lancet Infectious Diseases*, 2013, 13(2): 137–146.

[15] Tagliani E, Cabibbe AM, Miotto P, Borroni E, Toro JC,

Mansjö M, Hoffner S, Hillemann D, Zalutskaya A, Skrahina A, Cirillo DM. Diagnostic performance of the new version (v2.0) of GenoType MTBDRslAssay for detection of resistance to fluoroquinolones and second-line injectable drugs: a multicenter study. *Journal of Clinical Microbiology*, 2015, 53(9): 2961–2969.

[16] Satta G, Lipman M, Smith GP, Arnold C, Kon OM, McHugh TD. *Mycobacterium tuberculosis* and whole-genome sequencing: how close are we to unleashing its full potential? *Clinical Microbiology and Infection*, 2018, 24(6): 604–609.

[17] Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, Bradley P, Cabibbe AM, Cancino-Muñoz I, Caulfield MJ, Chaiprasert A, Cirillo DM, Clifton DA, Comas I, Crook DW, De Filippo MR, de Neeling H, Diel R, Drobniewski FA, Faksri K, Farhat MR, Fleming J, Fowler P, Fowler TA, Gao Q, Gardy J, Gascoyne-Binzi D, Gibertoni-Cruz AL, Gil-Brusola A, Golubchik T, Gonzalo X, Grandjean L, He G, Guthrie JL, Hoosdally S, Hunt M, Iqbal Z, Ismail N, Johnston J, Khanzada FM, Khor CC, Kohl TA, Kong C, Lipworth S, Liu Q, Maphalala G, Martinez E, Mathys V, Merker M, Miotto P, Mistry N, Moore DAJ, Murray M, Niemann S, Omar SV, Ong RT, Peto TEA, Posey JE, Prammananan T, Pym A, Rodrigues C, Rodrigues M, Rodwell T, Rossolini GM, Sánchez Padilla E, Schito M, Shen X, Shendure J, Sintchenko V, Sloutsky A, Smith EG, Snyder M, Soetaert K, Starks AM, Supply P, Suriyapol P, Tahseen S, Tang P, Teo YY, Thuong TNT, Thwaites G, Tortoli E, van Soolingen D, Walker AS, Walker TM, Wilcox M, Wilson DJ, Wyllie D, Yang Y, Zhang H, Zhao Y, Zhu B. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *The New England Journal of Medicine*, 2018, 379(15): 1403–1415.

[18] Mahé P, El Azami M, Barlas P, Tournoud M. A large scale evaluation of TBProfiler and mykrobe for antibiotic resistance prediction in *Mycobacterium tuberculosis*. *PeerJ*, 2019, 7: e6857.

[19] Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, Abdallah AM, Alghamdi S, Alsomali M, Ahmed AO, Portelli S, Oppong Y, Alves A, Bessa TB, Campino S, Caws M, Chatterjee A, Crampin AC, Dheda K, Furnham N, Glynn JR, Grandjean L, Minh Ha D, Hasan R, Hasan Z, Hibberd ML, Joloba M, Jones-López EC, Matsumoto T, Miranda A, Moore DJ, Mocillo N, Panaiotov S, Parkhill J, Penha C, Perdigão J, Portugal I, Rchiad Z, Robledo J, Sheen P, Shesha NT, Sirgel FA, Sola C, Oliveira Sousa E, Streicher EM, Helden PV, Viveiros M, Warren RM, McNerney R, Pain A, Clark TG. Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nature Genetics*, 2018, 50(2): 307–316.

[20] Companion Handbook to the WHO Guidelines for the Programmatic Management of Drug-Resistant Tuberculosis. 2014: *Geneva*.

[21] Chen PE, Shapiro BJ. The advent of genome-wide association studies for bacteria. *Current Opinion in Microbiology*, 2015, 25: 17–24.

[22] Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*, 2014, 11(4): 407–409.

[23] Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PKC, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nature Genetics*, 2013, 45(10): 1183–1189.

[24] Liaw AWM. Classification and regression by randomforest. *R News*, 2002, 2: 18–22.

[25] Zhang HT, Dongfang L, Zhao LL, Fleming J, Lin N, Wang T, Liu ZY, Li CY, Galwey N, Deng JY, Zhou Y, Zhu YF, Gao YR, Wang T, Wang SH, Huang YF, Wang M, Zhong Q, Zhou L, Chen T, Zhou J, Yang RF, Zhu GF, Hang HY, Zhang J, Li FB, Wan KL, Wang J, Zhang XN, Bi LJ. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nature Genetics*, 2013, 45(10): 1255–1260.

[26] Sobkowiak B, Glynn JR, Houben R, Mallard K, Phelan JE, Guerra-Assunção JA, Banda L, Mzembe T, Viveiros M, McNerney R, Parkhill J, Crampin AC, Clark TG. Identifying mixed *Mycobacterium tuberculosis* infections from whole genome sequence data. *BMC Genomics*, 2018, 19(1): 613.

[27] Zeng X, Kwok JS, Yang KY, Leung KS, Shi M, Yang Z, Yam

WC, Tsui SK. Whole genome sequencing data of 1110 *Mycobacterium tuberculosis* isolates identifies insertions and deletions associated with drug resistance. *BMC Genomics*, 2018, 19(1): 365.

[28] Huang H, Ding N, Yang T, Li C, Jia X, Wang G, Zhong J, Zhang J, Jiang G, Wang S, Zong Z, Jing W, Zhao Y, Xu S, Chen F. Cross-sectional whole-genome sequencing and epidemiological study of multidrug-resistant *Mycobacterium tuberculosis* in China. *Clinical Infectious Diseases*, 2019, 69(3): 405–413.

[29] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014, 30(15): 2114–2120.

[30] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009, 25(14): 1754–1760.

[31] Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, Marth GT, Quinlan AR, Hall IM. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nature Methods*, 2015, 12(10): 966–968.

[32] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA. The variant call format and VCFtools. *Bioinformatics*, 2011, 27(15): 2156–2158.

[33] Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE*, 2010, 5(3): e9490.

[34] Liu F, Hu YF, Wang Q, Li HM, Gao GF, Liu CH, Zhu BL. Comparative genomic analysis of *Mycobacterium tuberculosis* clinical isolates. *BMC Genomics*, 2014, 15(1): 1–13.

[35] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 2007, 81(3): 559–575.

[36] Popescu AA, Huber KT, Paradis E. Ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, 2012, 28(11): 1536–1537.

[37] Schliep KP. Phangorn: phylogenetic analysis in R. *Bioinformatics*, 2011, 27(4): 592–593.

[38] Phelan JE, O'Sullivan DM, Machado D, Ramos J, Oppong YEA, Campino S, O'Grady J, McNerney R, Hibberd ML, Viveiros M, Huggett JF, Clark TG. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Medicine*, 2019, 11(1): 1–7.

[39] Walker TM, Kohl TA, Omar SV, Hedge J, del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CLC, Bowden R, Drobniewski FA, Allix-Béguec C, Gaudin C, Parkhill J, Peto TEA. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *The Lancet Infectious Diseases*, 2015, 15(10): 1193–1202.

[40] Ramaswamy SV, Reich R, Dou SJ, Jasperse L, Pan X, Wanger A, Quitugua T, Graviss EA. Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 2003, 47(4): 1241–1250.

[41] Vilchèze C, Jacobs WR Jr. Resistance to isoniazid and ethionamide in *Mycobacterium tuberculosis*: genes, mutations, and causalities. *Microbiol Spectr*, 2014, 2(4): Mgm2-0014-2013.

[42] Parida SK, Axelsson-Robertson R, Rao MV, Singh N, Master I, Lutckii A, Keshavjee S, Andersson J, Zumla A, Maeurer M. Totally drug-resistant tuberculosis and adjunct therapies. *J Intern Med*, 2015, 277(4): 388–405.

[43] Pasca MR, Guglierame P, De Rossi E, Zara F, Riccardi G. *mmpL7* gene of *Mycobacterium tuberculosis* is responsible for isoniazid efflux in *Mycobacterium smegmatis*. *Antimicrob Agents Chemother*, 2005, 49(11): 4775–4777.

[44] Miotto P, Zhang Y, Cirillo DM, Yam WC. Drug resistance mechanisms and drug susceptibility testing for tuberculosis. *Respirology*, 2018, 23(12): 1098–1113.

[45] He L, Wang X, Cui P, Jin J, Chen J, Zhang W, Zhang Y. ubiA (Rv3806c) encoding DPPR synthase involved in cell wall synthesis is associated with ethambutol resistance in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)*, 2015, 95(2): 149–154.

[46] Brandis G, Wrande M, Liljas L, Hughes D. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. *Molecular Microbiology*, 2012, 85(1): 142–151.

[47] Colangeli R, Helb D, Sridharan S, Sun J, Varma-Basil M, Hazbón MH, Harbacheuski R, Megjugorac NJ, Jacobs WR, Jr., Holzenburg A, Sacchettini JC, Alland D. The *Mycobacterium tuberculosis iniA* gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. *Molecular Microbiology*, 2005, 55(6): 1829–1840.

[48] Clark TG, Mallard K, Coll F, Preston M, Assefa S, Harris D, Ogwang S, Mumbowa F, Kirenga B, O'Sullivan DM, Okwera A, Eisenach KD, Joloba M, Bentley SD, Ellner JJ, Parkhill J, Jones-López EC, McNerney R. Elucidating emergence and transmission of multidrug-resistant tuberculosis in treatment experienced patients by whole genome sequencing. *PLoS ONE*, 2013, 8(12): e83012.

[49] Ramaswamy SV, Amin AG, Göksel S, Stager CE, Dou SJ, El Sahly H, Moghazeh SL, Kreiswirth BN, Musser JM, Molecular genetic analysis of nucleotide polymorphisms associated with ethambutol resistance in human isolates of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 2000, 44(2): 326–336.

[50] Suarez J, Ranguelova K, Schelvis JPM, Magliozzo RS. Antibiotic resistance in *Mycobacterium tuberculosis*: peroxidase intermediate bypass causes poor isoniazid activation by the S315G mutant of m. tuberculosis catalase-peroxidase (KatG). *Journal of Biological Chemistry*, 2009, 284(24): 16146–16155.

[51] Lai LY, Hsu LY, Weng SH, Chung SE, Ke HE, Lin TL, Hsieh PF, Lee WT, Tsai HY, Lin WH, Jou R, Wang JT. A glutamine insertion at *Codon* 432 of RpoB confers rifampicin resistance in *Mycobacterium tuberculosis*. *Frontiers in Microbiology*, 2020, 11: 583194.

[52] Giri A, Gupta S, Safi H, Narang A, Shrivastava K, Kumar Sharma N, Lingaraju S, Hanif M, Bhatnagar A, Menon B, Alland D, Varma-Basil M. Polymorphisms in Rv3806c (ubiA) and the upstream region of embA in relation to ethambutol resistance in clinical isolates of *Mycobacterium tuberculosis* from North India. *Tuberculosis*, 2018, 108: 41–46.

[53] Haufroid M, Wouters J. Targeting the serine pathway: a promising approach against tuberculosis? *Pharmaceuticals*, 2019, 12(2): 66.

[54] Ejalonibu MA, Elrashedy AA, Lawal MM, Kumalo HM, Mhlongo NN. Probing the dual inhibitory mechanisms of novel thiophenecarboxamide derivatives against *Mycobacterium tuberculosis* PyrG and PanK: an insight from biomolecular modeling study. *Journal of Biomolecular Structure and Dynamics*, 2020: 1–13.

[55] Cui ZJ, Zhang WT, Zhu Q, Zhang QY, Zhang HY. Using a heat diffusion model to detect potential drug resistance genes of *Mycobacterium tuberculosis*. *Protein & Peptide Letters*, 2020, 27(8): 711–717.

[56] Bolard A, Plésiat P, Jeannot K. Mutations in gene fusA1 as a novel mechanism of aminoglycoside resistance in clinical strains of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 2018, 62(2): e01835–17. DOI:10.1128/aac.01835-17.

[57] Scribner MR, Santos-Lopez A, Marshall CW, Deitrick C, Cooper VS. Parallel evolution of tobramycin resistance across species and environments. *mBio*, 2020, 11(3).

[58] Philalay JS, Palermo CO, Hauge KA, Rustad TR, Cangelosi GA. Genes required for intrinsic multidrug resistance in *Mycobacterium avium*. *Antimicrobial Agents and Chemotherapy*, 2004, 48(9): 3412–3418.

[59] Alam MT, Petit RA III, Crispell EK, Thornton TA, Conneely KN, Jiang YX, Satola SW, Read TD. Dissecting vancomycin-intermediate resistance in *Staphylococcus aureus* using genome-wide association. *Genome Biology and Evolution*, 2014, 6(5): 1174–1185.

[60] Desjardins CA, Cohen KA, Munsamy V, Abeel T, Maharaj K, Walker BJ, Shea TP, Almeida DV, Manson AL, Salazar A, Padayatchi N, O'Donnell MR, Mlisana KP, Wortman J, Birren BW, Grosset J, Earl AM, Pym AS. Genomic and functional analyses of *Mycobacterium tuberculosis* strains implicate ald in D-cycloserine resistance. *Nature Genetics*, 2016, 48(5): 544–551.

[61] van Ingen J, Simons S, de Zwaan R, van der Laan T, Kamst-van Agterveld M, Boeree MJ, van Soolingen D. Comparative study on genotypic and phenotypic second-line drug resistance testing of *Mycobacterium tuberculosis* complex isolates. *Journal of Clinical Microbiology*, 2010, 48(8): 2749–2753.

[62] Torrea G, Coeck N, Desmaretz C, Van De Parre T, Van Poucke T, Lounis N, de Jong BC, Rigouts L. Bedaquiline susceptibility testing of *Mycobacterium tuberculosis* in an

automated liquid culture system. *Journal of Antimicrobial Chemotherapy*, 2015, 70(8): 2300–2305.

[63] Schon T, Jureen P, Giske CG, Chryssanthou E, Sturegard E, Werngren J, Kahlmeter G, Hoffner SE, Angeby KA. Evaluation of wild-type MIC distributions as a tool for determination of clinical breakpoints for *Mycobacterium tuberculosis*. *Journal of Antimicrobial Chemotherapy*, 2009, 64(4): 786–793.

[64] Organization W. Technical report on critical concentrations for TB drug susceptibility testing of medicines used in the treatment of drug-resistant TB. 2018.

[65] Ängeby K, Juréen P, Kahlmeter G, Hoffner S, Schön T. Challenging a dogma: antimicrobial susceptibility testing breakpoints for *Mycobacterium tuberculosis*. *Bulletin of the World Health Organization*, 2012, 90(9): 693–698.

[66] Colangeli R, Jedrey H, Kim S, Connell R, Ma S, Chippada Venkata UD, Chakravorty S, Gupta A, Sizemore EE, Diem L, Sherman DR, Okwera A, Dietze R, Boom WH, Johnson JL, Mac Kenzie WR, Alland D. Bacterial factors that predict relapse after tuberculosis therapy. *The New England Journal of Medicine*, 2018, 379(9): 823–833.

[67] Farhat MR, Freschi L, Calderon R, Ioerger T, Snyder M, Meehan CJ, de Jong B, Rigouts L, Sloutsky A, Kaur D, Sunyaev S, van Soolingen D, Shendure J, Sacchettini J, Murray M. GWAS for quantitative resistance phenotypes in *Mycobacterium tuberculosis* reveals resistance genes and regulatory regions. *Nature Communications*, 2019, 10: 2128.

[68] Saber MM, Shapiro BJ. Benchmarking bacterial genome-wide association study methods using simulated genomes and phenotypes. *Microbial genomics*, 2020, 6(3): e000337.

[69] Farhat MR, Freschi L, Calderon R, Ioerger T, Snyder M, Meehan CJ, de Jong B, Rigouts L, Sloutsky A, Kaur D, Sunyaev S, van Soolingen D, Shendure J, Sacchettini J, Murray M. GWAS for quantitative resistance phenotypes in *Mycobacterium tuberculosis* reveals resistance genes and regulatory regions. *Nature Communications*, 2019, 10(1): 1–11.

[70] Safi H, Sayers B, Hazbón MH, Alland D. Transfer of embB *Codon* 306 mutations into clinical *Mycobacterium tuberculosis* strains alters susceptibility to ethambutol, isoniazid, and rifampin. *Antimicrobial Agents and Chemotherapy*, 2008, 52(6): 2027–2034.

[71] Otto-Knapp R, Vesenbeckh S, Schönfeld N, Bettermann G, Roth A, Bauer T, Rüssmann H, Mauch H. Isoniazid minimal inhibitory concentrations of tuberculosis strains with *kat*G mutation. *The International Journal of Tuberculosis and Lung Disease*, 2016, 20(9): 1275–1276.

[72] Phelan J, Coll F, McNerney R, Ascher DB, Pires DEV, Furnham N, Coeck N, Hill-Cawthorne GA, Nair MB, Mallard K, Ramsay A, Campino S, Hibberd ML, Pain A, Rigouts L, Clark TG. *Mycobacterium tuberculosis* whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. *BMC Medicine*, 2016, 14(1): 1–13.

Evalutation of three methods searching resistant-related mutations in *mycobacterium tuberculosis* genome

Zhiyuan Li[1,2], Fei Liu[2], Demin Cao[1,2], Yuanlong Pan[2], Na Lü[2], Dongxin Liu[3], Wencong He[3], Cheng Cheng[2], Yanan Wang[2], Yanlin Zhao[3], Baoli Zhu[1,2*]

[1] University of Chinese Academy of Sciences, Beijing 100049, China

[2] CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

[3] National Tuberculosis Reference Laboratory, National Center for Tuberculosis Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Abstract: [Objective] The emergence of durg-resistant *Mycobacterium tuberculosis* brought a tough challenge to tuberculosis treatment. **[Methods]** Considering the lack of homologous recombination in *Mycobacterium tuberculosis*'s genome, the mutation strongly related to resistance could be efficiently confirmed by genome-wide association analysis. However, many resistance-related mutations had yet been found by existing methods. To calculate resistant-related mutations, researchers commonly used some methods similar to genome-wide association analysis (GWAS), which mainly included genome-wide efficient mixed model association (GEMMA), phylogenetic convergence test (phyc), plink. To find the better method among the three methods when calculating resistant-related SNPs on non-mobile antibiotic resistance gene, the genomes of 1504 *M. tuberculosis* strains from Hunan province and National Center of Biotechnology Information was obtained with their phenotypes of three first-line antibiotics (isoniazid, rifampicin, ethambutol). The three methods were performed to calculate the association between phenotype and known or novel SNPs related to resistance, and their sensibility and specificity were evaluated by the resistant-related SNPs got by the three methods. **[Results]** Phyc was able to search more known resistance-related SNPs with the sensibility and specificity higher than 52.49%. **[Conclusion]** Phyc is the most accurate in predicting the SNP related to drug resistance of *Mycobacterium tuberculosis*, but considering the update of running time and phenotypic data, the results of GEMMA and plink should also be used as a reference.

Keywords: *Mycobacterium tuberculosis*, resistance gene, comparative genomics

(本文责编：李磊)

Supported by the National Key R&D Program of China (2018YFC1603803), by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB29020203) and by the Beijing Municipal Science & Technology Commission (Z201100005520041)

*Corresponding author. Tel/Fax: +86-10-64807362; E-mail: zhubaoli@im.ac.cn

Received: 9 February 2021; Revised: 31 May 2021; Published online: 14 September 2021