

耐药结核分枝杆菌基因组信息挖掘

张建中

(《微生物学通报》编委会 北京 100101)

由于结核分枝杆菌的耐药性问题,使结核病这个古老的传染病死灰复燃,并成为全球性的严重公关问题。从1998年首个结核分枝杆菌(H37Rv)全基因组完成图的获得^[1],到近年来采用新一代测序技术对多个菌株进行高通量的基因组测序与分析,对结核分枝杆菌进化及耐药机制的认识不断深入。

关于结核分枝杆菌菌株的基因组比较分析有多篇报道,包括对中国12个省来源的161株结核分枝杆菌的全基因组测序和系统分析^[2],以及对世界各地不同型别的123株结核分枝杆菌的全基因组测序分析^[3]等,但比较组研究结果,所发现新候选耐药基因重复并不多,反映出当前数据挖掘能力的不足:采用不同的菌株背景和不同的数据挖掘策略,会有不同的发现。本刊2014年第5期刊登了林楠、汪世华等的论文“两株超级广泛耐药结核分枝杆菌全基因组比较分析”^[4],作者通过对两株不同型别泛耐药结核菌的全基因组分析,并与NCBI中22株菌全基因组数据^[5]整合分析,寻找可能的耐药机制和相应补偿机制,分析思路经典、深入,并发现已知的单耐药基因突变集合不能完全解释菌株的广泛耐药性,值得一读;也进一步提示了当今基因组数据挖掘技术不足的现状。如何充分挖掘基因组数据所蕴含的功能信息,是基因组学研究的重要内容;对数据处理、算法设计和功能信息挖掘等技术的探索和研究,应受到充分重视。

该研究通过对两株耐药菌株的全基因组比较分析,试图解释两个型别的广泛耐药菌株的突变情况,由于目前信息挖掘能力及所选择菌株的代表性所限,所得“突变位点中型别相关的和非型别相关的SNPs”仅是一个相对概念,需更多的实验验证。

关键词: 结核分枝杆菌, 基因组, 数据挖掘

参 考 文 献

- [1] Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence[J]. Nature, 1998, 393(6685): 537-544.
- [2] Zhang H, Li D, Zhao L, et al. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance[J]. Nature Genetics, 2013, 45(10): 1255-1260.
- [3] Farhat MR, Shapiro BJ, Kieser KJ, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*[J]. Nature Genetics, 2013, 45(10): 1183-1189.
- [4] 林楠, 周杰, 周盈, 等. 两株超级广泛耐药结核分枝杆菌全基因组比较分析[J]. 微生物学通报, 2014, 41(5): 1011-1019.
- [5] Comas I, Chakravarti J, Small PM, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved[J]. Nature Genetics, 2010, 42(6): 498-503.

Genomic data mining of drug-resistant *Mycobacterium tuberculosis*

ZHANG Jian-Zhong

(The Editorial Board of Microbiology China, Beijing 100101, China)

Keywords: *Mycobacterium tuberculosis*, Genome, Data mining