



## 基于基因组序列的脑膜炎奈瑟菌分型方法

袁梦<sup>1</sup>, 张雯<sup>2</sup>, 胡鹏威<sup>1</sup>, 俞慕华<sup>1</sup>, 陈辉<sup>1</sup>, 徐丽<sup>2</sup>, 高源<sup>2</sup>, 朱兵清<sup>2\*</sup>, 段永翔<sup>1\*</sup>, 邵祝军<sup>2</sup>

<sup>1</sup>南山区疾病预防控制中心, 广东 深圳 518054

<sup>2</sup>中国疾病预防控制中心, 北京 102206

**摘要:**【目的】建立并评估 1 种适宜的脑膜炎奈瑟菌(*Neisseria meningitidis*, Nm)基因组分子分型方法。【方法】本研究以 125 株代表性 Nm 菌株的基因组序列为对象, 建立了基于核心基因 SNP 的基因组分型方法, 并与 pubMLST 网站公布的 MLST 和 cgMLST 分型方法进行比较。【结果】基于核心基因 SNP 的基因组分型方法和 cgMLST 方法对 125 株 Nm 菌株的分型结果一致性较高, 两种方法均明显优于 MLST 分型方法。基于 SNP 的基因组分型方法在认识 Nm 菌的种群结构、界定克隆群方面具有优势; cgMLST 分型方法能够对任一菌株进行分型, 但不能进行克隆群的界定和归类。【结论】基于核心基因 SNP 的基因组分型方法和 cgMLST 均明显优于 MLST 分型方法, 未来仍有待进一步整合和提高。

**关键词:** 脑膜炎奈瑟菌, 核心基因 SNP 分型, cgMLST, MLST

脑膜炎奈瑟菌(*Neisseria meningitidis*, Nm)是引起流行性脑脊髓膜炎(简称“流脑”)的病原体, 其感染还可引起严重的脓毒血症<sup>[1]</sup>。根据荚膜多糖的抗原结构, Nm 可分为 12 个血清群; 95%以上的侵袭性病例由 A、B、C、Y、W 和 X 等 6 个血清群感染引起<sup>[2-3]</sup>。采用多位点序列分型(multi-locus sequence typing, MLST)方法, 大部分 Nm 菌株被归入 42 个克隆群(clonal complex, cc), 其中 cc1、

cc5、cc8、cc11、cc32、cc41/44、cc4821、cc269 等属于高致病 cc<sup>[4-9]</sup>。这些分型方法的应用对我们认识中国 Nm 菌株的遗传特征和变异规律起到了至关重要的作用。然而, 大量的研究结果显示, MLST 分型能力有限, 无法识别菌株在短时间内发生的变化<sup>[10-11]</sup>。另外, 在已经定义的 cc 之外, 尚有大量菌株无法进行 cc 归类(UA 菌株), 对这部分菌株, 无法确定其互相之间以及与其他菌株的

**基金项目:** 深圳市“医疗卫生三名工程”引进高层次医学团队项目(SZSM201803081); 南山区卫健局病原生物学、传染病防控医学重点学科建设资助

\*通信作者。朱兵清, Tel: +86-10-58900753, E-mail: zbzhu@126.com; 段永翔, Tel: +86-755-26410928, E-mail: szduanyx@163.com

收稿日期: 2020-06-09; 修回日期: 2020-09-02; 网络出版日期: 2020-10-28

遗传关系(<https://pubmlst.org/neisseria/>)。

近年来,随着基因组测序的普及,许多研究者尝试从基因组水平对菌株进行分型,并且在一些菌株的分型上取得了良好的结果<sup>[12-13]</sup>。这些方法主要基于两种原理——每个核心基因作为一个变异单元(cgMLST)和核心基因(core gene)的 SNP 作为变异单元。目前, Nm 的 cgMLST 方案已在 pubMLST 网站公布,并被用于重要 cc 的菌株进化分析<sup>[10-11]</sup>。本研究基于核心基因 SNP 建立 Nm 的基因组分型方法,并与 cgMLST 方法进行比较,分析二者的优缺点,为后期分型方法的选择和分型结果的解释提供理论依据。

## 1 材料和方法

### 1.1 材料

根据分离年代、分离地点、血清群和基因型,选择 125 株具有代表性的且背景清楚的 Nm 菌株,涵盖 A、B、C、X、W、E 和 cnl 等 7 个血清群/基因群;它们共分为 103 个 ST 型,其中,60 个 ST 型属于 16 个 cc,其余 43 个 ST 型为 UA 菌株。其中 111 株为本科室测序,14 株来源于公共数据库。

### 1.2 DNA 提取、基因组测序和组装

复苏实验室保存的 Nm,挑取单菌落,接种于血琼脂培养基上,置 5% CO<sub>2</sub> 37 °C 条件下培养 18-20 h。参照试剂盒(Promega, 美国)说明书提取细菌基因组 DNA。采用二代测序技术(Illumina HiSeq2000 或 2500 平台)进行基因组测序,这一部分工作由深圳华大基因科技服务有限公司和北京诺禾致源生物信息科技有限公司完成。测序下机的数据先进行过滤,获得 clean data,测序深度大

于 100×。使用 SOAPdenovo 1.04<sup>[14-15]</sup>短序列组装软件对 clean data 进行组装。拼接后基因组 contig 少于 100 个。

### 1.3 核心基因 SNP 分型方法

**1.3.1 基因预测和核心基因鉴定:**采用 Glimmer 3 软件<sup>[16]</sup>进行基因预测,并通过比对 COG、KEGG、NT、NR 和 SwissProt 等数据库进行基因的功能注释。对以上预测的基因,采用 OrthoMCL (<http://orthomcl.org/orthomcl/>)进行氨基酸水平的聚类分析,相似度较高(在 50%以上的长度上同源性高于 50%)且存在于所有菌株基因组中的单拷贝基因即为核心基因。

**1.3.2 单核苷酸多态性 (single nucleotide polymorphism, SNP)检测:**采用 SOAPsnip<sup>[14]</sup>和 MUMmer 中 NUCmer<sup>[17]</sup>软件,以 Nm 8013(GCA\_000026965.1)的完整基因组为参比序列,进行菌株间基因组序列比较,查找核心基因和核心基因 SNP。位于核心基因中、不在重复区域且 2 个 SNP 间距 $\geq 20$ bp 的 SNP 被定义为核心基因 SNP。

**1.3.3 群体结构分析:**采用 STRUCTURE 2.2 程序<sup>[18]</sup>,对 125 个 Nm 基因组筛查出的核心基因 SNP 开展群体结构分析,使用的迭代次数为 5,参数为 Burn-in=200000 和 Repts=500000。分别基于假设有 1 个群( $K=1$ )依次增加到 30 个群( $K=30$ ),依次计算,重复 5 次,之后运用 Structure HARVEST<sup>[19]</sup>计算不同假设下所获得的  $\Delta K$  值,选择  $\Delta K$  值最高者为最终的分群方案。利用 CLUMPP<sup>[20]</sup>软件将 STRUCTURE 运行得到的对应  $K$  值的多次重复结果进行合并,并据此评估对各株菌进行分组的情况。

**1.3.4 系统发育分析:**采用 MEGA 6 软件中的临

近法, 对核心基因 SNP 构建系统发育树。树的构建选择的 bootstrap 值为 1000。

#### 1.4 cgMLST 分型

向 pubMLST 网站 Nm 基因组数据库提交已组装的基因组序列。采用网站内 Genome comparator 工具, 选择 Scheme 为“*N. meningitidis* cgMLST v1.0”, 参数设置为“最小同源性  $\geq 70\%$ , 最短匹配长度  $\geq 50\%$ ”, 将基因组序列与核心基因数据库 (1605 个基因) 比对, 寻找基因组中的核心基因及其等位基因号, 并获得菌株间的差异矩阵。采用 Splits Tree (版本 4.14.8) 图像化展示菌株间差异, 即为菌株的系统发育树。

## 2 结果和分析

### 2.1 核心基因 SNP 分型

**2.1.1 核心基因和核心 SNP:** 在 125 株 Nm 菌株的基因组中, 1049 个基因为所有菌株共有, 即为核心基因。在这些基因中共筛选出 4263 个符合

要求的核心 SNP 位点, 用于群体结构和系统发育分析。

**2.1.2 群体结构:** 采用 STRUCTURE<sup>[18]</sup> 和 Structure HARVEST<sup>[19]</sup> 分析流程, 分别将 125 个 Nm 基因组基于分为 1–30 个群的假设分别进行群体结构计算。结果显示, Nm 分为 20 个群 ( $K=20$ ) 时  $\Delta K(81.30)$  最高, 其他分群方案  $\Delta K$  值为 0.53–74.35, 因此, 选定 20 个祖先群为最终的分型方案, 并采用该分型方案进行后续群特异性 SNP 和基因的鉴定, 此处将每一个群定义为基因组群 (genome typing group, GTG)。基于 Clumpup 软件, 计算 20 个 GTG 中的 125 株菌株的最高构成比 (the highest proportion of putative ancestral population, HPPAP), 如图 1 所示, 该值为 0.2018–1。按照 HPPAP 大于 0.4 则将该菌划分为该 GTG 的标准<sup>[12]</sup>, 114 株菌株可归入 18 个 GTG, 每个 GTG 包括 2–18 株菌株, 剩余 11 株菌株由于每个 GTG 的 HPPAP 都低于 0.4 而不能归入任何一个 GTG (图 1)。

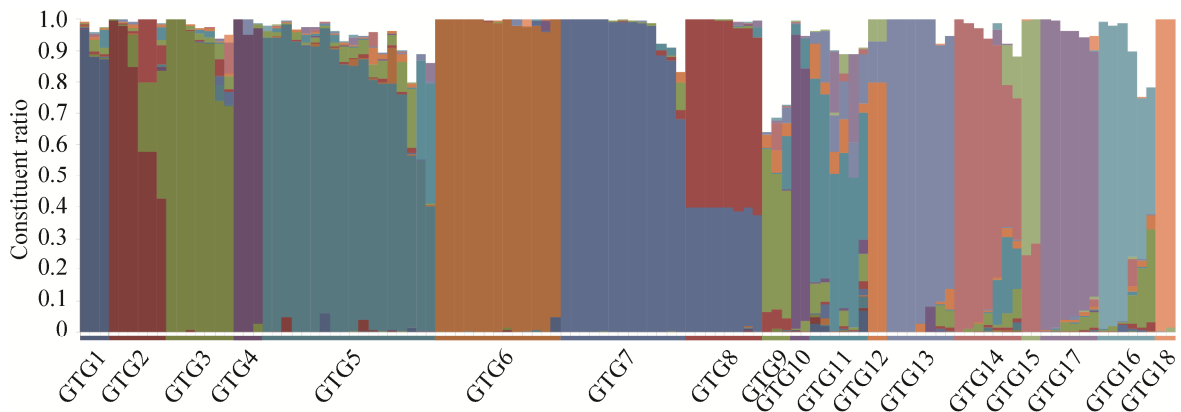


图 1. 18 个 GTG 在 114 株 Nm 菌株中的构成比

Figure 1. The constituent ratio of 18 GTGs of 114 *N. meningitidis* strains. Strains are arrayed along the horizontal ordinate. The proportion of different GTGs in each strain is shown on the vertical coordinate. Different GTGs are indicated by colors.

在 18 个 GTG 的 114 株 Nm 菌株中, 91 株菌株(79.8%) HPPAP  $\geq 0.6$ , 其中, 60 株菌株 HPPAP 高于 0.9, 18 株 HPPAP 为 0.8–0.9, 8 株为 0.7–0.8。GTG4、GTG6 和 GTG18 的所有菌株 HPPAP 均高于 0.9, 另有 8 个 GTG 的所有菌株 HPPAP 均高于 0.6。HPPAP 低的菌株主要在 GTG9 和 GTG11 中, 而这 2 个 GTG 包括的 cc 也较其他 GTG 多。

**2.1.3 核心基因 SNP 分型与 MLST 分型的比较:** 将以上核心基因分型结果与 MLST 分型结果进行比较后发现二者吻合度较高, 但核心基因分型对 MLST 明显具有补充作用(表 1)。18 个 GTG 中, 7 个 GTG 分别与 7 个 cc 对应, 6 个 GTG 仅包括 UA 菌株, 5 个 GTG 对应 2 个以上的 cc (包括 UA)。18 个 GTG 的 HPPAP 以及与 cc 的对应关系见表 1。

值得注意的是, 42 株 UA 菌株显示出明显的序列多态性, 其中部分菌株单独归入一个特定的 GTG, 部分菌株与其他 cc 菌株归入同一个 GTG。有 11 株菌株无法归入任何 GTG, 这些菌株中有 10 株为 UA 菌株。

**2.1.4 基于 SNP 的系统发育树:** 为了验证以上核心基因分型的可靠性, 采用以上 4263 个 SNP 对 125 株菌株进行了系统发育分析。从系统发育树上可以看出(图 2), 核心基因分型结果与系统发育一致性较好。除了 GTG9 和 GTG11, 其他 GTG 的菌株在系统发育树上各自成簇, 相互之间界限清晰。反过来, 16 个 cc 中的 13 个分别被归入单一的 GTG 中, 另外 3 个[cc4821(GTG5 和 GTG6)、cc11(GTG12 和 GTG13)和 cc41/44(GTG14 和

表 1. 125 株 Nm 菌株的基因组群(GTG)分类结果  
Table 1. The classification results of GTG of 125 *N. meningitidis* strains

GTG	Strains	HPPAP	Clonal complex (cc)
1	3	0.8744–0.9659	UA
2	6	0.4290–0.9966	UA
3	7	0.7226–1.0000	UA
4	3	0.9401–1.0000	UA
5	18	0.4039–0.9418	cc4821, UA
6	13	0.9515–1.0000	cc4821
7	13	0.6833–1.0000	cc4, cc5
8	8	0.5677–0.6000	cc1
9	3	0.4111–0.5210	cc53, cc213, UA
10	2	0.8114–0.9354	UA
11	6	0.4043–0.6546	cc18, cc174, cc175, cc269, UA
12	2	0.7968–0.8000	cc11
13	7	0.8069–1.0000	cc8, cc11, UA
14	7	0.4488–0.9978	cc41/44
15	2	0.7146–0.7522	cc41/44
16	6	0.4063–0.9802	UA
17	6	0.7866–0.9918	cc32
18	2	0.9838–0.9970	cc198
NA*	11	0.2018–0.3794	cc103, UA

\* 11 strains cannot be assigned any GTG due to the low HPPAP (<0.4).

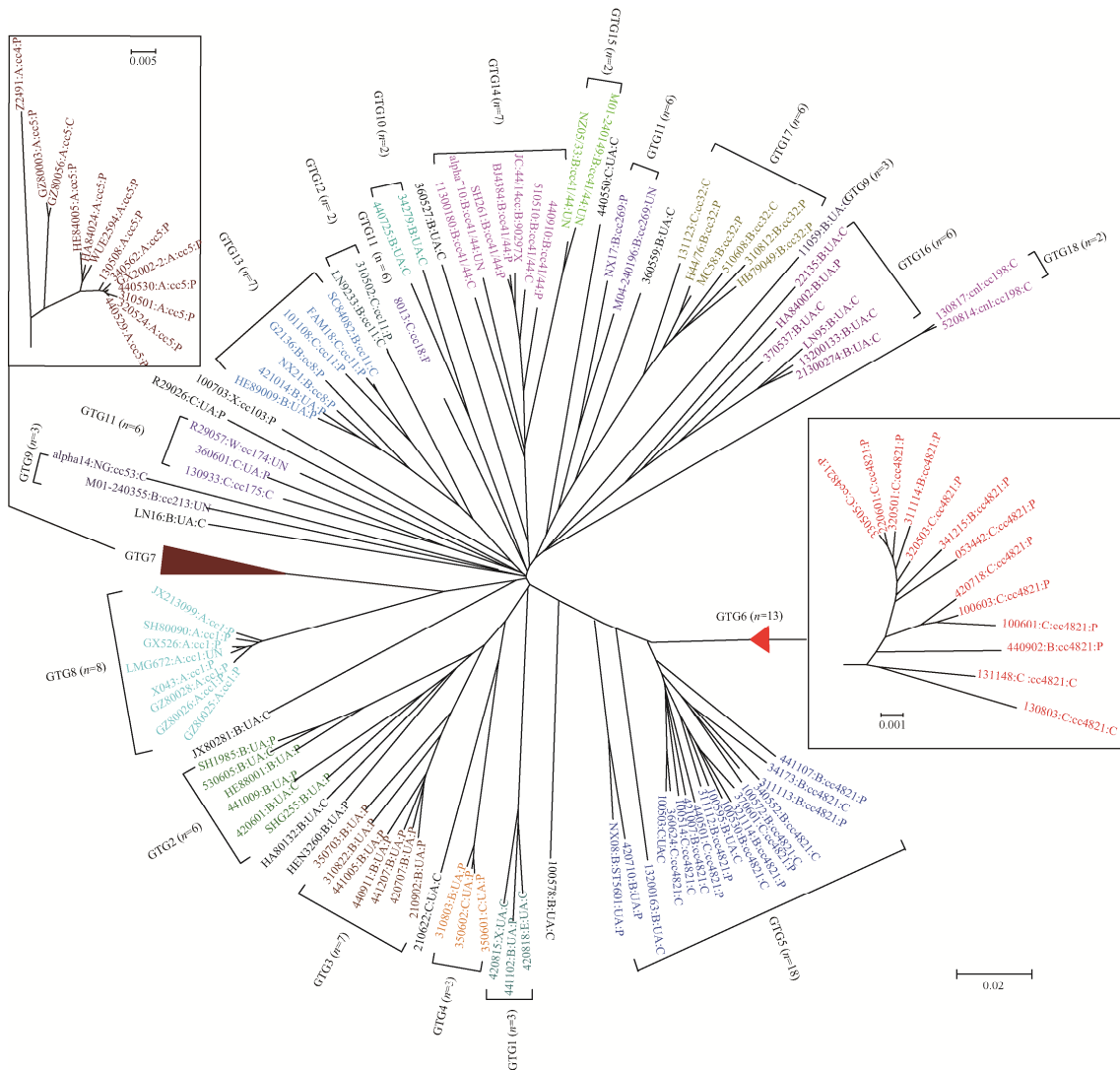


图 2. 基于 SNP 构建的 Nm 菌株系统发育树

Figure 2. The phylogenetic tree of *N. meningitidis* constructed based on SNPs. The information of strain includes the identification number, Serogroup clonal complex and strain source (P: patient; C: carrier) which were separated by a colon. The genome typing groups of all strains marked with different colors are labeled on the tree.

GTG15)]分别被归入 2 个分支相邻的 GTG 中。值得注意的是,同 1 个 GTG 中菌株之间的遗传距离可长于它们与其他 GTG 菌株间的距离,而且 UA 菌株(如 GTG4 中)之间的距离可相当近,甚至比同一个 cc (如 cc4821)中菌株之间的距离更近。除 GTG9 和 GTG11 中有 3 株菌株分散于系统发育树不同处外,其余同一个 cc 的菌株均聚集在树上的

同一处。

**2.1.5 GTG 特异 SNP 和基因:** 寻找 GTG 特异的 SNP 和/或基因可以快速确定菌株的遗传学归类。对 18 个 GTG 的全基因组序列(包括核心和非核心)进行比较,发现 14 个 GTG 具有特异性 SNP,其中 5 个 GTG 还含有特异性的基因。其余 4 个 GTG 中未发现特异基因或 SNP,但可以通过 SNP 组合

对其进行鉴定。GTG18 的特异性基因和 SNP 明显多于其他 GTG，与系统发育树中显示其为独立进化分支的结果吻合。

### 2.2 cgMLST 分型

1605 个核心基因中，5 个基因在所有菌株中完全相同，1600 个基因存在差异；另有 111 个基因序列与数据库中两个或更多的基因位点匹配。

保留菌株间存在差异且仅与唯一位点匹配的 1489 个基因，用于计算菌株间的遗传距离和构建系统发育树(图 3)。从图中可以看出，该系统发育树的拓扑结构与基于 SNP 构建的系统发育树几乎一致，菌株之间的遗传关系亦高度相似，那些属于同一个 GTG 但遗传距离较远的菌株在该树上距离亦较远。

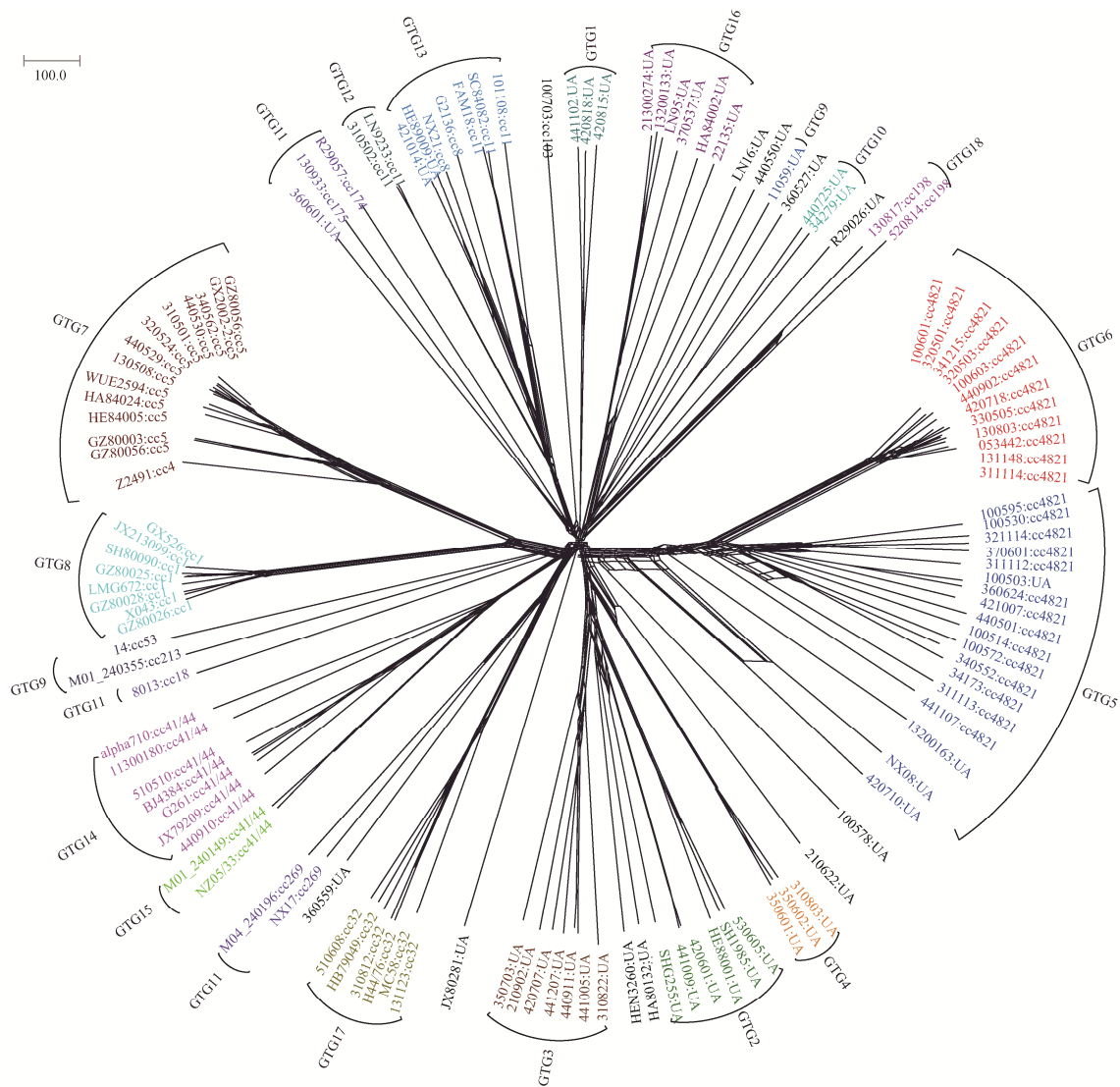


图 3. 基于 cgMLST 构建的 Nm 菌株系统发育树

Figure 3. The phylogenetic tree of *N. meningitidis* constructed based on cgMLST. The information of strain included the identification number and clonal complex which were separated by a colon. The genome typing groups of all strains were labeled and the color of each GTG was in accordance with that in Figure 2.

### 3 讨论

随着高通量测序技术的发展,细菌基因组已逐渐应用到部分菌种的鉴定分型领域<sup>[12,21-22]</sup>,其与传统分型方法相结合,具有分辨率高的优点。本研究建立了基于核心基因的 Nm 的分型方法,同时采用 cgMLST 基因组分型方法,对同一批 125 株菌株进行了分子分型,并与传统的 MLST 方法的分型结果进行了比较。结果显示,基于基因组序列的分型方法与 MLST 的分型结果一致性较高,同时在以下方面又具有明显的优势:(1) 在分辨 UA 菌株方面,基因组分型方法相较于 MLST 方法更灵敏。在定义 UA 菌株间以及 UA 与其他 cc 菌株间的关系上,本研究中的大部分 UA 菌株在基因组系统发育树上成簇分布,显示出克隆性。分析菌株的 MLST 位点信息,发现除了 GTG9 和 GTG11 外,每个 GTG 中的菌株具有至少一个相同的 MLST 位点,说明这些菌株确实具有相关性,MLST 之所以未能将其归为相应的 cc,原因是 MLST 分析的位点有限而对 cc 的定义又比较严格(4 个以上的位点相同方可归入该 cc)。GTG9 和 GTG11 比较特别,其 3 株菌株分散于系统发育树上多处,且 MLST 位点也没有同源性,因为这些菌株归入相应 GTG 的 HPPAP 均较低,因此推测它们不是这两个 GTG 的代表菌株,或者这些菌株在进化过程中发生了较多的重组,导致菌株间差异较大。(2) 基于基因组序列的分型方法能够对现有的 cc 进行更准确的界定。在基因组序列分型中,cc4821 分为 2 个明显不同的 GTG(GTG5 和 GTG6)或簇,已有研究结果显示,两簇菌株之间生物学特征差异显著<sup>[23-24]</sup>,这说明 cc4821 确实是由 2 个遗传分支组成,对 cc4821 菌株如果仅仅按照 MLST 结果进行归类将会掩盖许多有价值的信息。反过

来,有些 cc 具有紧密的遗传关系,但在 MLST 分型中被界定为独立的 cc,如 cc8 和 cc11(二者的中心 ST 型有 4 个位点相同),基因组分型结果显示这 2 个 cc 同属于 GTG13,在基因组分型上没有明显的差异。

本研究中的两种基因组分型方法对菌株的分型结果几乎一致,但二者分型原理不同,潜在的适用范围也有差异。基于 SNP 的基因组分型方法如同 MLST 方法一样,先对基因组型别(GTG)进行界定,确定每个型别的特异基因或 SNP 位点,然后根据待检菌株中这些特异基因和位点信息,对菌株进行 GTG 归类。该方法的优点是分型标准固定,便于不同批次或不同实验室间结果的比较。而且,还可以根据每个 GTG 组特异的 SNP 位点或基因设计简单的 PCR 方法对 GTG 进行鉴定,从而提高分型效率和降低分型费用。本研究建立的 Nm 核心基因分型方法的缺点是,仍有可能有未知的 GTG 未被涵盖。Nm 种群结构复杂,目前公布的 cc 已超 40 个,另外还有相当一部分菌株不能归入任何一个 cc;而且,近年来 Nm 新的 cc 或亚群不断出现<sup>[5,11,25]</sup>,在这一种属中选取代表性菌株建立基因组分型系统是相当困难的。本研究发现,不能进行 GTG 归类的菌株具有多个祖先群的成分但比例都不高,这说明在本研究鉴定的 20 个 GTG 之外,可能还有其他的 GTG 未被识别出来。此外,Nm 具有较高的转化和重组能力<sup>[26]</sup>,可致使菌株快速变化,对于基于特异的 SNP 位点或基因开发的用于 GTG 鉴定的 PCR 方法,如果变化发生于 GTG 特异性位点,有可能导致 GTG 的误判。

相比起来,cgMLST 方法纳入了所有的核心基因,且以每个核心基因作为一个变异单元,既

避免了利用少数几个基因进行分型带来的不稳定性, 又避免由重组带来的大量 SNP 的影响。cgMLST 是一个开放的分型系统, 可以对任何新的 Nm 菌株进行分型。但是, 目前 cgMLST 分型系统中还没有对不同的 cgMLST 型别进行 cc 归类的方案, 因此, 只能通过构建系统发育树, 根据菌株间的遗传距离, 判断某个菌株与其他菌株的关系。然而, 从本研究中构建的系统发育树来看, 有的 cc 内部菌株间的遗传距离较大, 有的又相当小, 不同的 GTG 之间也是这种情况, 这说明仅仅根据菌株间的遗传距离对菌株进行归类是不准确的。

总体来说, 基于 SNP 的基因组分型方法和 cgMLST 分型方法要明显优于传统的 MLST 分型方法, 但二者适用范围也不一样。基于 SNP 的基因组分型方法在认识 Nm 的种群结构、界定 cc 方面具有优势。cgMLST 分型方法能够计算任何两株菌株之间的遗传距离, 但不能进行 cc 的界定和归类, 目前 cgMLST 方法主要用于对单一 cc 菌株的变异分析<sup>[10-11]</sup>。目前这两种方法可结合应用在 Nm 的分型工作中。未来我们计划进一步整合和优化两种分型方法, 进一步提高其适用性、实用性和准确性。

## 参 考 文 献

- [1] Harrison LH, Pelton SI, Wilder-Smith A, Holst J, Safadi MAP, Vazquez JA, Taha MK, LaForce FM, von Gottberg A, Borrow R, Plotkin SA. The global meningococcal initiative: recommendations for reducing the global burden of meningococcal disease. *Vaccine*, 2011, 29(18): 3363–3371.
- [2] Abio A, Neal KR, Beck CR. An epidemiological review of changes in meningococcal biology during the last 100 years. *Pathogens and Global Health*, 2013, 107(7): 373–380.
- [3] Li JH, Li YX, Shao ZJ, Li L, Yin ZD, Ning GJ, Xu L, Luo HM. Prevalence of meningococcal meningitis in China from 2005 to 2010. *Vaccine*, 2015, 33(8): 1092–1097.
- [4] Harrison LH, Trotter CL, Ramsay ME. Global epidemiology of meningococcal disease. *Vaccine*, 2009, 27 Suppl 2: B51–B63.
- [5] Caugant DA, Kristiansen PA, Wang X, Mayer LW, Taha MK, Ouédraogo R, Kandolo D, Bougoudogo F, Sow S, Bonte L. Molecular characterization of invasive meningococcal isolates from countries in the African meningitis belt before introduction of a serogroup A conjugate vaccine. *PLoS ONE*, 2012, 7(9): e46019.
- [6] Achtman M. Microevolution and epidemic spread of serogroup a *Neisseria meningitidis*-a review. *Gene*, 1997, 192(1): 135–140.
- [7] Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou JJ, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt B G. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Science of the United States of America*, 1998, 95(6): 3140–3145.
- [8] Shao ZJ, Li W, Ren J, Liang XF, Xu L, Diao BW, Li MC, Lu MJ, Ren HY, Cui ZG, Zhu BQ, Dai ZW, Zhang L, Chen X, Kan B, Xu JG. Identification of a new *Neisseria meningitidis* serogroup C clone from Anhui province, China. *The Lancet*, 2006, 367(9508): 419–423.
- [9] Law DKS, Lorange M, Ringuette L, Dion R, Giguère M, Henderson AM, Stoltz J, Zollinger WD, De Wals P, Tsang RSW. Invasive meningococcal disease in Quebec, Canada, due to an emerging clone of ST-269 serogroup B meningococci with serotype antigen 17 and serosubtype antigen P1.19 (B:17:P1.19). *Journal of Clinical Microbiology*, 2006, 44(8): 2743–2749.
- [10] Lucidarme J, Hill DMC, Bratcher HB, Gray SJ, du Plessis M, Tsang RSM, Vazquez JA, Taha MK, Ceyhan M, Efron AM, Gorla MC, Findlow J, Jolley KA, Maiden MCJ, Borrow R. Genomic resolution of an aggressive, widespread, diverse and expanding meningococcal serogroup B, C and W lineage. *Journal of Infection*, 2015, 71(5): 544–552.
- [11] Zhu BQ, Lucidarme J, Bai XL, Guo PB, Zhang AY, Borrow R, Gao WY, Xu L, Gao Y, Shao ZJ. Comparative genomic analyses of Chinese serogroup W ST-11 complex *Neisseria*



- meningitidis* isolates. *Journal of Infection*, 2020, 80(1): 54–60.
- [12] Chen C, Zhang W, Zheng H, Lan RT, Wang HY, Du PC, Bai XM, Ji SB, Meng Q, Jin D, Liu K, Jing HQ, Ye CY, Gao GF, Wang L, Gottschalk M, Xu JG. Minimum core genome sequence typing of bacterial pathogens: a unified approach for clinical and public health microbiology. *Journal of Clinical Microbiology*, 2013, 51(8): 2582–2591.
- [13] Meijis AP, Hengeveld PD, Dierikx CM, Maassen CBM, de Greeff SC, de Haan A, Bosch T, van Duijkeren E. Prolonged carriage of (livestock-associated) MRSA in individuals without professional livestock contact. *Journal of Antimicrobial Chemotherapy*, 2020, 75(6): 1405–1409.
- [14] Li RQ, Li YR, Fang XD, Yang HM, Wang J, Kristiansen K, Wang J. SNP detection for massively parallel whole-genome resequencing. *Genome Research*, 2009, 19(6): 1124–1132.
- [15] Luo RB, Liu BH, Xie YL, Huang WH, Yuan JY, He GZ, Chen YX, Pan Q, Liu YJ, Tang JB, Wu GX, Zhang H, Shi YJ, Liu Y, Yu C, Wang B, Lu Y, Han CL, Cheung DW, Yiu SM, Peng SL, Zhu XQ, Liu GM, Liao XK, Li YR, Yang HM, Wang J, Lam TW. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 2012, 1(1): 18.
- [16] Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics*, 2007, 23(6): 673–679.
- [17] Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. *Genome Biology*, 2004, 5(2): R12.
- [18] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*, 2000, 155(2): 945–959.
- [19] Earl DA, Vonholdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 2012, 4(2): 359–361.
- [20] Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 2007, 23(14): 1801–1806.
- [21] Zhou HJ, Liu WB, Qin T, Liu C, Ren HY. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Klebsiella pneumoniae*. *Front Microbiology*, 2017(8): 371.
- [22] Kan B, Zhou HJ, Du PC, Zhang W, Lu X, Qin T, Xu JG. Transforming bacterial disease surveillance and investigation using whole-genome sequence to probe the trace. *Frontiers of Medicine*, 2018, 12(1): 23–33.
- [23] Zhu B, Xu Z, Du P, Xu L, Sun X, Gao Y, Shao Z. Sequence type 4821 clonal complex serogroup B *Neisseria meningitidis* in China, 1978–2013. *Emerging Infectious Diseases*, 2015, 21(6): 925–932.
- [24] Guo PB, Zhu BQ, Liang H, Gao WY, Zhou GL, Xu L, Gao Y, Yu JX, Zhang MJ, Shao ZJ. Comparison of pathogenicity of invasive and carried meningococcal isolates of ST-4821 complex in China. *Infection and Immunity*, 2019, 87(12): e00584–19.
- [25] Kristiansen PA, Diomandé F, Wei SC, Ouédraogo R, Sangaré L, Sanou I, Kandolo D, Kaboré P, Clark TA, Ouédraogo AS, Absatou KB, Ouédraogo CD, Hassan-King M, Thomas JD, Hatcher C, Djingarey M, Messonnier N, Préziosi MP, LaForce M, Caugant DA. Baseline meningococcal carriage in Burkina Faso before the introduction of a meningococcal serogroup A conjugate vaccine. *Clinical and Vaccine Immunology*, 2011, 18(3): 435–443.
- [26] Cui YJ, Yang XW, Didelot X, Guo CY, Li DF, Yan YF, Zhang YQ, Yuan YT, Yang HM, Wang J, Wang J, Song YJ, Zhou DS, Falush D, Yang R. Epidemic clones, oceanic gene pools, and eco-Id in the free living marine pathogen *Vibrio parahaemolyticus*. *Molecular Biology and Evolution*, 2015, 32(6): 1396–1410.

# Whole genome sequence-based typing methods for *Neisseria meningitidis*

Meng Yuan<sup>1</sup>, Wen Zhang<sup>2</sup>, Pengwei Hu<sup>1</sup>, Muhua Yu<sup>1</sup>, Hui Chen<sup>1</sup>, Li Xu<sup>2</sup>, Yuan Gao<sup>2</sup>, Bingqing Zhu<sup>2\*</sup>, Yongxiang Duan<sup>1\*</sup>, Zhujun Shao<sup>2</sup>

<sup>1</sup> Nanshan Center for Disease Control and Prevention, Shenzhen 518054, Guangdong Province, China

<sup>2</sup> Chinese Center for Disease Control and Prevention, Beijing 102206, China

**Abstract:** **[Objective]** To find a suitable molecular typing method for *Neisseria meningitidis*. **[Methods]** We used 125 representative *N. meningitidis* strains to develop a new typing method, named core genome single nucleotide polymorphism (SNP) typing, based on the whole genome sequencing. This method was compared to the MLST and cgMLST typing that have been published on pubMLST website. **[Results]** The genotyping results of 125 *N. meningitidis* strains isolates were consistent between the core genome SNP typing and egMLST, both of which are significantly better than MLST. The core genome SNP typing identified the population structure and defined the clonal complexes. Comparatively, cgMLST could type any strain but cannot define or assign its clonal complex. **[Conclusion]** Core genome SNP typing showed consistent results with cgMLST and both methods are significantly better than MLST.

**Keywords:** *Neisseria meningitidis*, core genome SNP typing, cgMLST, MLST

(本文责编: 张晓丽)

---

Supported by the Sanming Project of Medicine in Shenzhen (SZSM201803081) and by the Medical Key Subject for Pathogen Biology and Prevention and Control of Infectious Diseases of Nanshan District Municipal Health Commission

\*Corresponding authors. Bingqing Zhu, Tel: +86-10-58900753, E-mail: zbqzhu@126.com; Yongxiang Duan, Tel: +86-755-26410928, E-mail: szduanyx@163.com

Received: 9 June 2020; Revised: 2 September 2020; Published online: 28 October 2020