



计算机技术的进步帮助我们监测流感新动向

在 一场新的全球性大流感可能爆发之际，全球卫生主管部门都必须仔细谨慎地制定应对策略。其中，制定测序策略的目的就是探究致病病毒株的危险性有多大，以及它们可能会变得有多危险。在应对今年春天爆发的甲型H1N1流感时，科学家借助了各种各样的计算机工具来预测病毒的感染力。

最新的分析工具可以同时好几株流感病毒株的遗传信息进行比对，而不再像以前那样，只是重点关注病毒基因组序列中几个突变位点的统计学意义。有些专家认为，这些更加高级的算法能帮助我们更加准确地估计流感大疫情的进展。

虽然科学家早已开始使用计算机来分析流感病毒数据，但还是有一部分研究人员选择使用人工比对的方式将新发现的流感病毒毒株与以往的毒株进行比较，他们希望凭借这种方法来预测新毒株的性状。Wendy Barclay是英国伦敦帝国理工学院（Imperial College）流感病毒学系主任，他将甲型H1N1病毒株的基因组序列与现有数据库中的数据进行比对后发现，这次流感病毒株血凝素（hemagglutinin）H1氨基酸序列中有两处关键的突变位点，这两个突变位点与造成1918年世界流感大流行的那株病毒株中发现的突变位点一模一样。之前，研究人员对1918年流感病毒株的研究曾发现，这两处氨基酸位点的突变有助于病毒与人体上呼吸道细胞之间的粘附，这样病毒就不易侵入人体肺组织当中（*Science* 315, 655–659; 2007）。这说明本次甲型H1N1流感病毒也可能极易传播，但是致病性不会有甲型H5N1禽流感病毒那么强，因为甲型H5N1禽流感病毒更倾向于与肺组织细胞粘附（*Science* 312, 399; 2006及*Nature* 440,435–436; 2006）。

正是有了美国资助的流感病毒基因组测序计划（Influenza Genome Sequencing Project）和

中国科学院北京基因组研究所（Beijing Institute of Genomics）的流感病毒数据库这类项目的开展，流感病毒的相关数据才得以不断积累。现在，GenBank网络数据库中已收录了上万条流感病毒基因组数据，科研工作者可以非常方便地利用这些资源，将新发现的病毒序列与以往的病毒序列进行比对。

但是，美国疾病预防控制中心（US Centers for Disease Control and Prevention）的流感病毒专家Jackie Katz认为，随着流感病毒数据越来越多，人们就越来越需要计算机的帮助才能发现流感病毒的感染力特征。比如，计算机可以发现好几株流感病毒中共进化的突变位点以及与高感染力有关的突变位点等等。

以前的计算机比对算法重点关注流感病毒基因组中单个位点的突变情况。统计学分析表明，这些突变位点都与流感病毒的毒力相关。不过，最新的比对算法主要是综合评价那些单独从统计学角度看起来与病毒毒力无关的突变位点，但是将这些位点结合起来分析就会发现，它们都与病毒毒力相关。Jonathan Allen和Tom Slezak都是美国能源部Lawrence Livermore国家实验室（Energy's Lawrence Livermore National Laboratory）的计算机专家，他们合作开发了一种比对算法，这种算法就能发现那些只有通过“彼此协作”才能造成危害的突变位点。

就在刚刚过去的2009年4月，Allen和Slezak报

道了他们对20世纪出现的2,100多株流感病毒样品进行分析的结果，这些病毒样品还包括造成1918年、1957年和1968年全球流感大流行的3株病毒。Allen等人发现，这3株造成流感大疫情的病毒株中有34个保守的氨基酸位点（参见*BMC Microbiol.* 9, doi:10.1186/1471-2180-9-77; 2009），但这些保守位点中只有一半，即17个保守位点可以在今年春天爆发的这次H1N1流感病毒株中发现。

Allen认为他们开发的这套软件工具能够帮助科研人员从病毒基因组中发现与流感病毒感染力相关的位点。不过，虽然这套软件能够对基因组数据进行筛查，但它还是无法对其它一些重要的因素进行分析，例如流感病毒如何与宿主体免疫系统发生相互作用等这些对于病毒毒力预测非常重要的因素。

Allen说道：“从计算机研究的角度来说，如何进一步缩小预测候选位点结果的范围还是一个需要长期研究的课题。”

除了病毒基因组测序数据之外，科研人员还运用流行病学等资料来预测今年这次H1N1流感病毒感染力，并运用于病毒传播扩散速度的数学模型当中。该方法曾经预测出这株H1N1流感病毒会在墨西哥造成4至1,000名患者的死亡（参见*Science* doi:10.1126/science.1176062; 2009）。不过，由于科研人员不可能获得轻症患者甚至是无症状患者的资料，所以该方法还是存在着很多不确定因素。

除了运用计算机方法分析之外，还有科研人员尝试通过检测病毒在组织或动物模型体内的状况来进行研究。科学家使用反向遗传学（reverse genetics）技术（即在病毒中引入突变，然后观察这些突变会对病毒造成何种影响）来开展这类研究。Barclay认为，这类试验具有非常重要的作用。他指出，你无法确定根据基因组序列预测出的结果是否真的会表现出生物学意义。

William Gallaher是美国路易斯安那州立大学（Louisiana State University）的荣誉退休教授，他提醒我们不要盲目相信计算机预测的结果。因为计算机构建的模型中都是人为设定的假设，计算机只会基于这些假设进行分析。William Gallaher还说：“要保证结果的正确性，先得保证这些假设的正确性。”

原文检索

Kirsten Dorans. (2009) Computational tools evolve to reveal patterns in flu data. *Nature Medicine*, 15(6): 587.



图片说明：H1N1流感病毒已经引起了全世界的恐慌，我们需要为大疫情爆发做好准备。

图片来源：Newscom