

# Puzzelen met bacterieel DNA

4-6-2021

De waarneembare eigenschappen (kenmerken) van organismen zijn het resultaat van de expressie van veel genen die aanwezig zijn binnen de cellen van organismen. De samenstelling, volgorde en expressie van deze genen maakt een organisme uniek; we noemen dit het 'genotype' van een organisme. Genotypen kunnen worden vastgesteld met technieken die gebruikmaken van zogenaamd 'Next Generation Sequencing (NGS)'. In de afgelopen 10 jaar is er een explosie geweest in het aantal medische microbiologische laboratoria en het aantal onderzoekers dat gebruikmaakt van deze NGS-techniek. De informatie die met NGS wordt verkregen, kan worden gebruikt om volledige genoomsequenties van pathogene bacteriën in kaart te brengen - deze bacteriën zijn bijvoorbeeld geïsoleerd uit patiënten met levensbedreigende infecties of tijdens uitbraken van antimicrobiële resistente infecties binnen ziekenhuizen of verpleeginstellingen.

NGS-technieken genereren een grote hoeveelheid sequentiefragmenten die nauwkeurig als puzzelstukjes in een bepaalde volgorde moeten worden gelegd (geassembleerd) om er zinvolle conclusies aan te kunnen verbinden. Er zijn verschillende bioinformatica softwareprogramma's die gebruikt kunnen worden om een compleet beeld van deze 'puzzelstukjes' te maken, maar die verschillende softwareprogramma's zijn niet allemaal even accuraat en dit kan mogelijk gevolgen hebben voor het bepalen van het genotype van een infectie-veroorzakende bacterie.

Dr. Nikolas Strepis probeert uit te zoeken welk softwareprogramma of combinatie van softwareprogramma's het meest accurate en complete beeld genereert van de samenstelling en volgorde van genen van verschillende ziekmakende bacteriën die bijvoorbeeld binnen ziekenhuizen voorkomen. Hij gebruikt hiervoor vele bacteriële stammen en verschillende softwareprogramma's - bijvoorbeeld het commerciële QIAGEN CLC Genomics-softwarepakket<sup>1</sup> of met open source programma SPADES<sup>2</sup> met de bijbehorende Unicycler-software<sup>3</sup>. Elk softwareprogramma heeft zijn eigen voor- en nadelen en daarom onderzocht dr. Strepis wat de meest accurate softwareprogramma's zijn die gebruikt kunnen worden binnen het medisch microbiologisch laboratorium.

Zijn onderzoek laat zien dat de verkregen resultaten vaak niet identiek zijn, zelfs niet als dezelfde bacteriële sequenties worden gebruikt. Deze verschillen zouden er mogelijk toe kunnen leiden dat bacteriën niet correct worden geïdentificeerd, of dat hun verspreiding binnen een ziekenhuis verkeerd wordt ingeschat, of dat het aantal of types antimicrobiële resistentiegenen die worden gedragen door de bacterie, niet correct zijn. Het is dus mogelijk dat de geassembleerde genomen niet een goede weerspiegeling is van de oorspronkelijke bacteriën. Met betrouwbare analyse van het genoom kan de verspreiding van bacteriële ziekteverwekkers binnen ziekenhuizen goed in de gaten gehouden worden.

1. <https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-genomics-workbench/>
2. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-477. doi:10.1089/cmb.2012.0021
3. Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13(6): e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>

Dr. N. Strepis is postdoctoraal bio-informaticus bij de afdeling Medische Microbiologie &

Infectieziekten, Erasmus Universitair Medisch Centrum (Erasmus MC), Rotterdam.

Dr. Strepis verklaart geen belangenconflicten te hebben en hij heeft geen banden met de hardware- of softwareontwikkelaars die in dit artikel worden genoemd.