

Wiskunde

Christian Jongeneel

Grafentheorie brengt virus in kaart

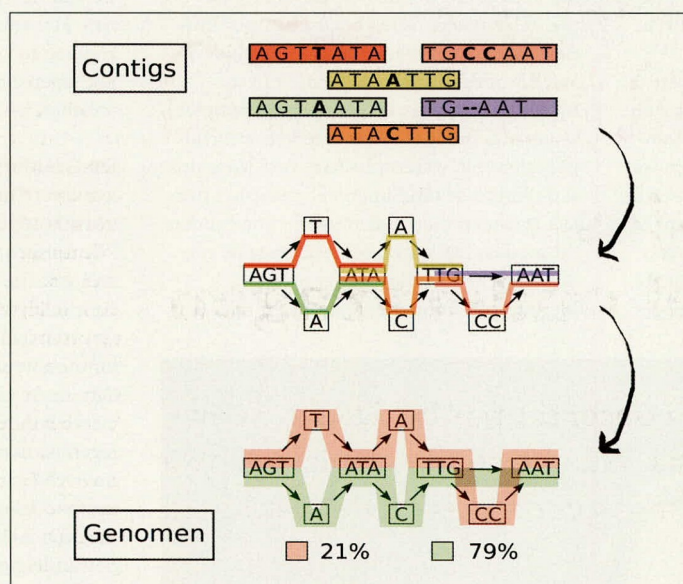
Vanwege de grote genetische variatie is het lastig om het genoom van een virus in kaart te brengen.

Onderzoekster Jasmijn Baaijens puzzelde voor haar promotie de stukjes wiskundig in elkaar.

Virussen hebben compact DNA. Eenmaal in een lichaam vermenvuldigen ze zich razendsnel, waarbij veel varianten ontstaan. Het DNA van een enkel virus sequencen lukt niet, het is altijd het gemiddelde materiaal van een heleboel exemplaren. Bovendien levert sequencen kleine stukjes op, zogeheten *reads*, die later aan elkaar geplakt moeten worden tot een genoom. Bij een mens weet je dat er precies twee kopieën zijn van ieder chromosoom, maar bij een hoeveelheid virusmateriaal weet je niet hoeveel gelijke kopieën er zijn. Bovendien zijn er mutaties. Dat maakt het reconstrueren van het genoom ingewikkeld.

Jasmijn Baaijens, die in september bij het Centrum voor Wiskunde en Informatica in Amsterdam promoveerde op genoomreconstructie van virussen, gebruikt de grafentheorie om hier helderheid in te scheppen. Een graaf is een wiskundige term voor een verzameling punten die op

'Verschillende groepen gaan deze methode daadwerkelijk gebruiken'



Een variatiegraaf brengt de variatie tussen de contigs in beeld. Door het oplossen van een optimalisatieprobleem worden de contigs vervolgens gecombineerd tot complete virusgenomen.

de een of andere manier onderling verbonden kunnen zijn: een knoop. 'In de overlapgraaf die ik gemaakt heb, vormt iedere read een knoop', vertelt Baaijens. 'De pijlen in de graaf geven aan dat de reads overlappen en dus waarschijnlijk van dezelfde kopie van het virus afkomstig zijn.' In zo'n graaf zijn clusters van reads te onderscheiden, die grote overlap hebben met elkaar. Baaijens bedacht algoritmes om die clusters te vinden en met elkaar te combineren tot langere stukjes virtueel DNA, zogeheten *contigs*. Daarmee bouwde ze een nieuwe overlapgraaf. Dat proces herhaalde ze tot er een compleet genoom uit kwam. De methode werkt, maar heeft wel beper-

kingen. Het is namelijk lastig om de puzzel volledig op te lossen, dat wil zeggen de virusgenomen als één stuk te reconstrueren. Vaak blijven er nog meerdere contigs per genoom over. Baaijens: 'Maar je wilt eigenlijk de complete genomen met elkaar kunnen vergelijken. Daarom heb ik ook een tweede methode ontworpen, waarmee je de contigs verder in elkaar puzzelt.'

Baaijens' aanpak loste ook een ander probleem op, namelijk dat je bij sequencing niet zeker weet of een bepaalde mutatie echt is, of het gevolg van een uitleesfout van de apparatuur. In de praktijk vergelij-

ken onderzoekers de uitkomsten daarom doorgaans met een bekend referentiegenoom om de leesfouten eruit te halen. Bij nieuwe virussen bestaat die voorkennis niet, maar in de enorme hoeveelheid data blijkt je met een slim algoritme wel een onderscheid te kunnen vinden tussen eenmalige leesfouten en zeldzame mutaties.

Rekentijd

Naast het bedenken van de algoritmes stak Baaijens veel tijd in de optimalisatie ervan. Met name het bouwen van de graaf kost namelijk veel rekentijd. Immers, voor iedere read neemt de rekentijd aanzienlijk toe, omdat die met alle andere vergeleken moet worden om de mate van overlap te bepalen. Ook de clustering van de reads tot contigs vergde na het *proof of concept* nog efficiëntere algo-

ritmen om in de praktijk bruikbaar te zijn.

'We hadden tijdens het onderzoek contact met verschillende groepen die onze methodes daadwerkelijk gaan gebruiken om virussen te reconstrueren', zegt Baaijens. 'Het is aan microbiologen om uit de wiskundige resultaten bijvoorbeeld de genen en mutaties te halen die kenmerkend zijn voor dit bepaalde virus.'

Inmiddels werkt Baaijens aan Harvard University, waar ze zich over bacteriën buigt. 'Enerzijds is het probleem vergelijkbaar, maar bacteriën hebben een langer genoom en gedragen zich totaal anders. Dus er zijn nog uitdagingen genoeg.' ●