

Groeiende rol statistiek bij ontwikkeling nieuwe groenterassen



Als je in het groentevak in de supermarkt het aantal tomatensoorten telt, kom je makkelijk tot tien: trostomaten, minitros-, snack-, kers-, vlees-, roma-, gele, oranje, groene... Wat komt er bij kijken om nieuwe tomaten- en groenterassen te kweken? Marc Rutten, statisticus bij Bayer CropScience Vegetable Seeds, legt uit welke rol wiskunde hierbij speelt.

Naast kantoorgebouwen staan op de site van Bayer CropScience in Nunhem een aantal grote kassen. Hier, en op andere locaties verspreid over de wereld, worden zaden ontwikkeld voor de professionele tuinbouw die wereldwijd worden geproduceerd, zo'n 1200 rassen van 28 verschillende groentes. De site in Nunhem huisvest ook een R&D-centrum, waar wetenschappers de plantenveredelaars helpen vernieuwende groenterassen te ontwikkelen. Statistiek is hierbij een onmisbaar hulpmiddel.

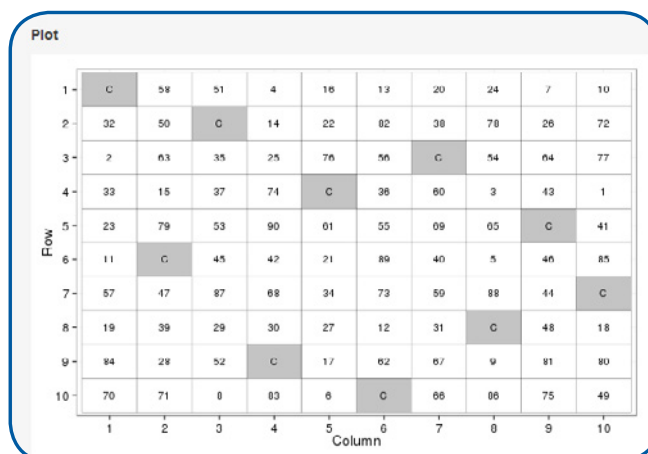
Slimme layout

"Een belangrijke formule bij gewassenveredeling", vertelt Rutten, "*is phenotype = genotype + milieu (-invloeden)*. Groenten worden veredeld op het fenotype, dus op het uiterlijk van de plant. Bij een tomaat is dit bijvoorbeeld de vorm, de glans, de kleur, de vorm van een cluster – dit is de manier waarop ze aan een tak zitten –, de smaak, de opbrengst en nog veel meer. En voor de groentelers zijn opbrengst en resistentie tegen ziektes hele belangrijke eigenschappen."

Voor een deel worden deze eigenschappen door het genotype bepaald, dus door de erfelijke eigenschappen. Voor een ander deel worden ze veroorzaakt door milieu-invloeden zoals de hoeveelheid licht die een plant krijgt, de hoeveelheid voedsel, warmte enzovoort. Als je nieuwe soorten kweekt, wil je graag weten hoeveel van de variatie in een fenotype je kunt toeschrijven aan het genotype. "We noemen dit de *heritability*. Dit geeft een indicatie van de vooruitgang die je uiteindelijk kunt bereiken in een bepaalde eigenschap", aldus Rutten.

Om daar achter te komen worden proeven uitgevoerd waarbij vele variëteiten van een groentegewas worden getoetst op hun gebruikswaarde voor de tuinbouw en groenteketen. Rutten helpt bij het ontwerp van deze experimenten. "Het is allereerst belangrijk om *confounding* te voorkomen", legt Rutten uit. "In de statistiek is dit de term voor een versturende variabele, zoals het effect van de locatie. Groentes zoals wortel, prei en ui worden buiten, in de volle grond gekweekt. Het komt dan wel voor dat een stuk land wat lager ligt en daardoor vochtiger is, of dat een deel van een proefvak meer wind vangt. Afhankelijk van de locatie zal een plant dan beter of slechter groeien. Als alle planten van één variëteit op een slechte locatie staan en alle planten van een andere variëteit op een gunstige locatie, weet je niet in welke mate prestatieverschillen die je meet, veroorzaakt worden door het genotype of de milieu-invloeden."

"Ook in de kas speelt locatie een rol, bijvoorbeeld als een plant vlak bij een ventilator staat of aan de buitenkant van een rij. Bij het inzaaien van proefvakken zaaien we daarom commerciële hybriden mee. De prestatie van deze planten ligt extreem goed vast. Als deze referentieplanten zich verschillend ontwikkelen, kun je dit toeschrijven aan zo'n versturende variabele. Met behulp van data-analyse kun je hiervoor corrigeren. We hebben daarom een layout tool ontwikkeld om per experiment de beste plaats voor die referentieplanten te bepalen."

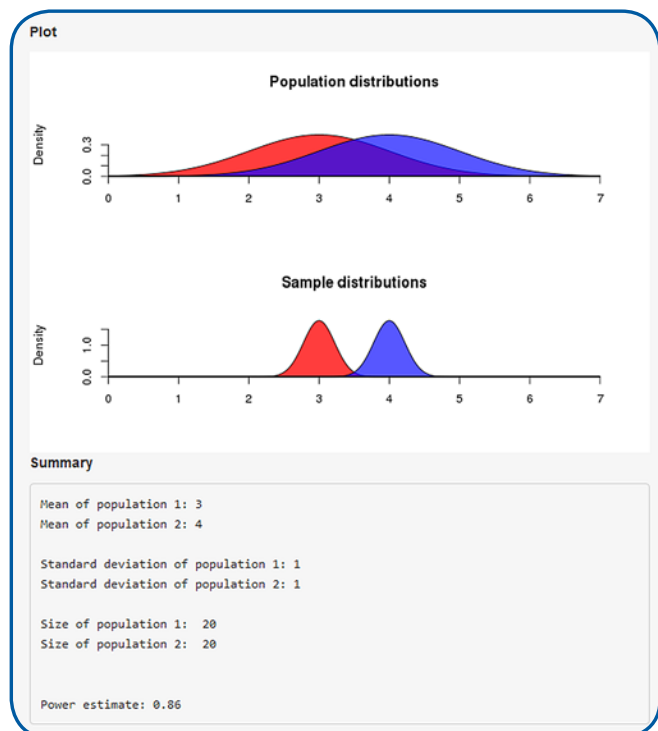


Tool waarmee de layout van een proefvak wordt geoptimaliseerd; deze tool zorgt ervoor dat de referentieplanten (de grijze posities) goed verdeeld zijn om voor locatie-effecten te kunnen corrigeren

Zo min mogelijk planten

De ontwikkelaars die nieuwe rassen kweken heten veredelaars. Voor zijn experimenten, heeft een veredelaar een deel van een proefkas of proefakker tot zijn beschikking. Deze ruimte wil hij natuurlijk optimaal benutten. Hoeveel planten heb je nodig om gefundeerde uitspraken te kunnen doen? Op basis van eerdere proeven is vaak bekend welke spreiding je kunt verwachten in bepaalde eigenschappen, zoals de opbrengst van een gewas. Hoe meer metingen je doet hoe nauwkeuriger je dergelijke parameters kunt bepalen.

“In een experiment vergelijken we bijvoorbeeld twee varianten van een gewas. Een veredelaar onderzoekt per experiment vaak meerdere kenmerken. Met een door ons ontwikkelde tool kunnen zij berekenen hoeveel planten er nodig zijn om aan te kunnen tonen dat die kenmerken in de twee populaties statistisch van elkaar verschillen. Deze tool is gebaseerd op een methode die poweranalyse heet. Soms is de conclusie dat er zoveel planten gekweekt moeten worden dat de beschikbare ruimte niet groot genoeg is. Een veredelaar kan dan kijken of een grotere opzet met meer ruimte mogelijk is of beslissen om één of meerdere kenmerken later te onderzoeken.”



In de powertool kun je de parameters van verdelingen en populatiegroottes invoeren om de power van een experiment te toetsen

Tijdens het experiment worden meetgegevens verzameld door mensen die erin getraind zijn om de juiste kenmerken vast te leggen. Na afloop gaat Rutten met deze gegevens aan de slag: “Het is de kunst om de effecten die je wilt onderzoeken boven water te halen. We doen dit onder andere met lineaire regressie modellen. Hierin verwerken we ook de locatiegegevens van een plant zodat we locatie-effecten kunnen uitfilteren. Op basis van mijn analyses beoordelen de veredelaars de mogelijkheden van nieuwe varianten van een gewas.”



Moderne varianten van tomaten en sla

Er zijn constant nieuwe varianten nodig. Enerzijds om in te spelen op de vraag naar nieuwe smaken, vormen of kleuren, anderzijds om antwoord te bieden aan (nieuwe) ziektes die de gewassen bedreigen. Hiervoor wordt vaak teruggegrepen op oude varianten van een plant die nog dicht bij de oervorm liggen. Voor sommige eigenschappen is bekend op welke plek op het DNA de verantwoordelijke genen zijn vastgelegd. Bij het kweken van nieuwe varianten worden eigenschappen uit een oud ras ingekruist in het DNA van een veredelingsras. Dit gebeurt bijvoorbeeld met een techniek die backcrossing wordt genoemd. Hierbij worden een oud ras met een nieuw ras gekruist. “We maken hierbij gebruik van wiskunde die ontwikkeld is door de evolutiebioloog Haldane”, legt Rutten uit. “Met behulp van DNA-markers en mapping functies kunnen we bepalen welke planten wel en welke niet geschikt zijn om te kruisen voor een volgende generatie. Zo kunnen we het aantal planten dat je moet kweken om het gewenste DNA-profiel te krijgen tot een minimum beperken.”

“Door de resultaten die we boeken, zien steeds meer veredelaars de toegevoegde waarde.”

Laatbloeier

Rutten noemt zichzelf een laatbloeier op het gebied van wiskunde. “Op de middelbare school vond ik het een moeilijk vak. Toen ik later in mijn opleiding met statistiek in aanraking kwam, raakte ik geboeid. Ik vind het fascinerend hoe je fenomenen die je met je eigen ogen hebt gezien door middel van analyse verder kunt ontdoen van ruis. In de wereld van groenteveredeling krijgt statistiek een steeds prominentere rol. Door de resultaten die we boeken, zien steeds meer veredelaars de toegevoegde waarde. Proeven blijven nodig, maar ik geloof in de kracht van analyse en ben ervan overtuigd dat we steeds meer vanaf papier gewassen kunnen veredelen.”