

De herkomst van meidoornspanners op Walcheren: DNA-onderzoek anno 2016

Ken Kraaijeveld, Mark Lammers, Jesse Kerkvliet
Jolanda Essens, Dwin Grashof, Jens Zwart
Johan den Dunnen, Jan Goedbloed

TREFWOORDEN

Next generation sequencing, *Theria primaria*, verwantschap

Entomologische Berichten 77 (1): 18-22

In 2011 werd een populatie van de meidoornspanner, *Theria primaria*, ontdekt op het Zeeuwse eiland Walcheren. Gezien de geïsoleerde ligging van deze populatie ten opzichte van andere meidoornspannerpopulaties en de vleugelloosheid van de vrouwelijke vlinders, rees de vraag waar deze populatie vandaan kwam. In samenwerking met de stichting *LeveDNA!* werd een project opgezet om met behulp van de nieuwste DNA-technologie een antwoord te vinden op deze vraag. Het complete DNA, ca. 537 miljoen letters (nucleotiden), van vier meidoornspanners, afkomstig uit populaties in Utrecht, Frankrijk, Engeland en Walcheren werd afgelezen. In totaal vonden we 130.000 één-letterverschillen (puntmutaties) tussen het DNA van de vier meidoornspanners. De verschillen tussen alle vier de meidoornspanners waren bijna even groot, maar de Walcherse meidoornspanner leek het meest verwant aan die uit Frankrijk. Daarmee lijkt het onwaarschijnlijk dat de populatie op Walcheren recent is aan komen waaien uit één van de andere populaties. Eerder is deze populatie een overblijfsel van een populatie met een veel grotere verspreiding die zich uitstreckte tot in Frankrijk.

Introductie

Meidoornspanners op Walcheren

De vlinder- en libellenwerkgroep Zeeland deed van 2006 tot en met 2012 uitgebreid onderzoek naar het voorkomen van nachtvlinders in Zeeland. Dit met het doel om samen met Stichting Het Zeeuws Landschap een provinciale nachtvlinderatlas uit te brengen. Daartoe voerden vrijwilligers inventarisaties uit met verschillende methoden. Zo onderzocht Jan Goedbloed bij Dishoek op Walcheren onder andere welke vlinders te lokken waren met smeer (een mengsel van banaan, suiker en alcohol). Tijdens zijn avondronde op 22 januari 2011 zag hij op een takje net onder één van zijn smeerplekken twee kleine spannertjes zitten (figuur 1). Hij ging er in eerste instantie van uit dat het kleine wintervlinders, *Operophtera brumata* (Linnaeus), waren, maar de vlindertjes bleken een band op de vleugels te hebben met daarin een donkerder vlek. Het waren meidoornspanners, *Theria primaria* (Haworth), een zeldzame soort die alleen bekend was uit het oosten des lands. Onmiddellijk werd een zoekactie op touw gezet. De soort bleek voor te komen aan zowel de zuidwestkust als aan de noordkust van Walcheren, maar niet in andere delen van de delta (figuur 2). De dichtstbijzijnde populaties vonden zich voor zover bekend op de Utrechtse heuvelrug, in Noord-Frankrijk en Engeland (figuur 3).

Gezien de geïsoleerde ligging van de populatie meidoornspanners op Walcheren, rees de vraag waar deze vandaan komt. Net als bij verschillende andere soorten spanners die in herfst, winter of vroege voorjaar actief zijn, hebben de vrouwtjes van

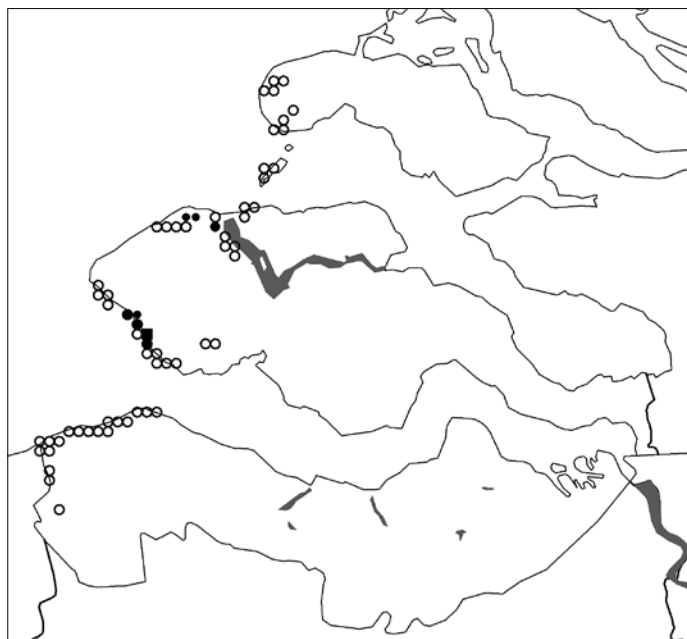
de meidoornspanner sterk gereduceerde vleugels (Heppner 1991). Mogelijk zorgen ze er zo voor dat de beperkte hoeveelheid energie optimaal gebruikt wordt voor de voortplanting. Een ogenschijnlijk nadeel is dat de volwassen vrouwtjes geen rol kunnen spelen bij de verspreiding. Deze soorten verspreiden zich mogelijk doordat de jonge rupsen, gezeten op de uiteinden van takjes, een spinseldraad produceren en zich door de wind laten meevoeren (Roff 1990). Ze maken dan deel uit van het 'luchtplankton'. Het is dus mogelijk dat de meidoornspanners Walcheren hebben bereikt via windverspreiding vanuit een andere populatie, vanwege de gangbare heersende windrichting waarschijnlijk ten zuiden of westen van Walcheren. Een tweede mogelijkheid is dat de soort is meegekomen met plantmateriaal dat gebruikt is om Walcheren te herbepplanten na de inundatie in 1944 en dat dit afkomstig was uit Oost-Nederland of uit een ander Europees land. Ten derde zouden meidoornspanners kunnen zijn meegekomen op drijvend plantenmateriaal dat werd aangevoerd met rivierwater. Als laatste kan de Walcherse populatie een overblijfsel zijn van een vroegere populatie met een veel ruimere verspreiding.

Stichting *LeveDNA!*

In augustus 2012 schreef de eerste auteur van dit artikel een column in dit tijdschrift (Kraaijeveld 2012), waarin hij beargumenteerde dat de nieuwste DNA-technologie dermate goedkoop was geworden dat 'Jan-de-amateur-entomoloog' aan de slag kon met zijn DNA-vragen. De stichting *LeveDNA!* zou



1. Meidoornspanner, *Theria primaria*. Foto: J. Goedbloed
1. Early moth, *Theria primaria*.



2. Voorkomen van de meidoornspanner, *Theria primaria*, in Zeeland. Dichte cirkel = vondst; open cirkel = wel gezocht, niet gevonden.
2. Distribution of the early moth, *Theria primaria*, in the province of Zeeland. Filled circles = records; open circles = sampled areas without early moth records.

kunnen bemiddelen. Jan Goedbloed voelde zich aangesproken en nam contact op. Een plan werd gesmeed: hij zou meidoornspanners verzamelen uit Walcheren, Engeland, Utrecht en Frankrijk, Ken Kraaijeveld zou zorgen dat hieruit DNA geïsoleerd werd, wat vervolgens op het Leiden Genome Technology Center (onder leiding van Johan den Dunnen) afgelezen zou worden. Het aflezen van de lettervolgorde van het DNA ('sequencing'), werd in dit geval gedaan met behulp van de nieuwste technologie, de zogenaamde 'next generation sequencing' (zie www.LeveDNA.nl voor uitleg over deze techniek). Ten slotte zouden de gegevens geanalyseerd worden door studenten bioinformatica van de Hogeschool Leiden (Jesse Kerkvliet, Jolanda Essens, Dwin Grashof en Jens Zwart). Het pas opgerichte instituut Generade werd bereid gevonden de sequencing te financieren. Generade heeft tot doel projecten te realiseren waarbij studenten van de Hogeschool Leiden helpen vragen met behulp van genomics technologie (DNA) te beantwoorden. Het meidoornspannerproject kon mooi dienen als een testcase voor dit model. Mark Lammers hielp met de fylogenetische analyse. Dit artikel is het resultaat van dit project.

Methoden

Verzamelde monsters

Naast meidoornspanners verzameld bij Dishoek op Walcheren (51.476465-3.518862) werden in januari 2014 exemplaren verzameld bij Les Hudions, Machiennes – Wadignies-Hamange, Frankrijk (50.395305-3.288213), Littleworth, Faringdon, Oxfordshire, Engeland (51.671837-1.548475) en op de Grebbeberg, Utrecht (51.950759-5.601651). De vlinders werden verzameld door te zoeken met een zaklamp, en gedood en geconserveerd in epjes met alcohol 70%.

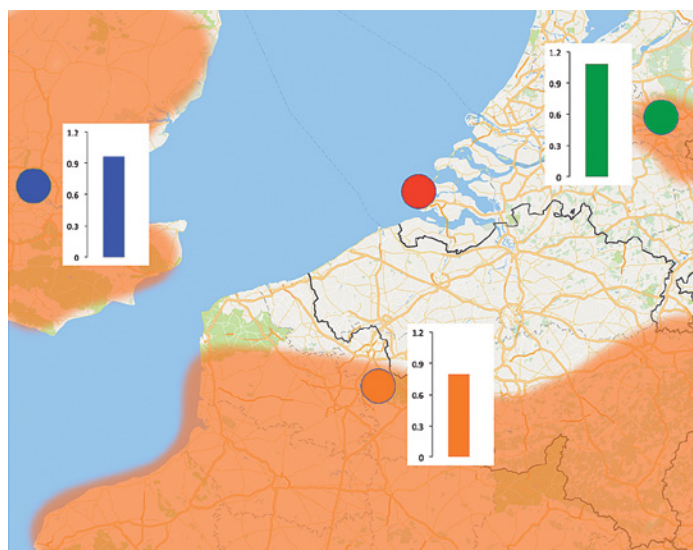
DNA-extractie en sequencing

Eén meidoornspanner uit elk van de vier verzamellocaties werd

gebruikt voor de DNA-analyse. Andere, tegelijk verzamelde, exemplaren bevinden zich in de collectie van J. Goedbloed. Uit elk van de vier meidoornspanners werd DNA geïsoleerd met de 'blood and tissue kit' van Qiagen (Qiagen, Valencia, CA). Het DNA werd vervolgens klaar gemaakt voor sequencing volgens bestaande standaard protocollen. Eerst werd het DNA in willekeurige stukken gebroken met ultrasone geluidstrillingen (instrument van Covaris Inc., USA). Aan weerszijden van de DNA-fragmenten werden stukjes synthetisch DNA geplakt van een bekende lettervolgorde (Illumina adapters). Vervolgens werden stukken kleiner dan 600 en stukken groter dan 700 DNA-letters verwijderd. De overgebleven stukken werden afgelezen op een Illumina Hiseq 2000 via de zogenaamde 'paired-end'-modus. Dit laatste betekent dat van ieder fragment de eerste 100 en de laatste 100 letters werden gelezen, de zogenaamde 'reads'. Tussen deze twee stukken zat dus een gat van ongeveer 400 letters waarvan we de lettervolgorde niet wisten.

Data-analyse

Het monster waarvan de meeste reads beschikbaar waren, werd gebruikt om het totale DNA (het 'genoom') van de meidoornspanner te reconstrueren. Dit is een enorme puzzel, een zogenaamde 'de novo assembly' en werd uitgevoerd met behulp van de software Platanus (standaard instellingen; zie Kajitani *et al.* 2014). Bij dit puzzelen wordt de overlap tussen losse stukjes gebruikt om steeds langere stukken ('contigs') van de DNA-volgorde te verkrijgen. Bijvoorbeeld: de twee stukjes AAGGcctt en ccttGAAG worden gecombineerd tot AAGGccttGAAG. Deze contigs werden vervolgens met elkaar verbonden tot 'scaffolds', gebruik makende van het feit dat we wisten dat twee reads van 100 letters afkomstig van hetzelfde DNA-fragment bij elkaar hoorden. Alleen scaffolds langer dan 2.000 letters werden gebruikt voor de uiteindelijke analyse. De losse DNA-volgordes van alle vier de meidoornspanners werden vergeleken met de set lange scaffolds met behulp van de software Bowtie2 (Langmead & Salzberg 2012). Scaffolds afkomstig van het mitochondriale



3. Geografische posities van de verschillende populaties van de meidoornspanner, *Theria primaria*, ten opzichte van Walcheren. De oranje gebieden geven globaal het verspreidingsgebied van de meidoornspanner weer (gebaseerd op verschillende gegevens van internet, zoals Waarneming.nl, Waarnemingen.be, en Lepinet.fr). De cirkels markeren de vier bemonsterde populaties. De grafiekjes geven de fylogenetische afstand ten opzichte van de meidoornspanner uit Walcheren weer.

3. Geographical location of Walcheren in relation to the distribution of the early moth, *Theria primaria*, in Northwestern Europe (in orange). The coloured circles mark the sampled populations. The bar plots indicate the phylogenetic distance relative to the early moth from Walcheren.

DNA werden verwijderd. Tenslotte werden alle één-letterverschillen (puntmutaties) opgespoord, gebruik makend van Samtools (Li et al. 2009) en Varscan (Koboldt et al. 2012). De resulterende lijst met verschillen werd vervolgens gefilterd op de volgende criteria: de positie moest in alle vier de monsters minimaal tien keer zijn gezien en ieder monster moest homozygoot zijn, dat wil zeggen de twee chromosomen moesten op die plek dezelfde letter hebben. De software SNPhylo (Lee et al. 2014) werd gebruikt om aan de hand van alle overgebleven

verschillen de verwantschappen tussen de vier meidoornspanners te bepalen. Deze software gebruikt 'bootstrapping' om statistische significantie en betrouwbaarheidsintervallen voor de lengte van de takken van de resulterende fylogenetische boom te berekenen.

Resultaten

Tabel 1 geeft een overzicht van de DNA-data voor elk van de vier meidoornspanners. De meidoornspanner uit Frankrijk gaf de beste opbrengst en daarom werd daarmee een de novo assembly uitgevoerd. Het resultaat daarvan staat samengevat in tabel 2. Ter vergelijking zijn dezelfde statistieken weergegeven voor het recent gepubliceerde genoom van de kleine wintervlinder, *Operophtera brumata*, en dat van het best bestudeerde insect, de fruitvlieg *Drosophila melanogaster* Meigen (getallen van NCBI-website: <http://www.ncbi.nlm.nih.gov/genome>). Duidelijk blijkt dat met de gegevens die we hadden van de Franse meidoornspanner slechts een heel schetsmatig genoom kan worden bepaald. In vergelijking met de kleine wintervlinder en fruitvlieg bestond onze assembly uit heel veel scaffolds en was dus nog erg gefragmenteerd. De N50 (de helft van de totale assembly werd gevangen in scaffolds langer dan deze waarde) was kort, wat eveneens duidt op een gefragmenteerde assembly. De voornaamste reden hiervoor was het ontbreken van informatie over veel langere DNA fragmenten dan de 600 tot 700 nucleotiden die wij ter beschikking hadden. Zulke gegevens zijn duur en bewerkelijk om te verkrijgen. Slechts voor 29% van het DNA kregen we scaffolds groter dan 2.000 letters. Voor onze doeleinden was dit echter ruim voldoende.

De reads van de vier meidoornspanners uit verschillende populaties werden vergeleken met de scaffolds van de Franse populatie (figuur 4). Dit resulteerde in een lijst van 898.716 één-letterverschillen, waarvan er na filteren 128.203 overbleven. De verwantschappen tussen de vier meidoornspanners zijn weergegeven in figuur 5. De populaties uit Walcheren en Frankrijk blijken dichter aan elkaar verwant dan elk van deze aan die uit Utrecht of Engeland. Dit verschil is statistisch significant: de lengte tussen de twee splitsingen in het midden van de boom is groter dan nul (betrouwbaarheidsinterval= 0.0066-0.0240, $p < 0.01$). Daar staat tegenover dat de taklengte van elke

monster / sample	aantal paired-end reads / number of pair end reads	aantal baseparen ($\times 1.000.000.000$) / number of base pairs ($\times 1.000.000.000$)
Dishoek, Walcheren	72.752.835	18,2
Oxfordshire, Engeland	51.765.910	12,9
Machiennes, Frankrijk	89.059.275	22,3
Grebbeberg, Utrecht	69.849.028	17,4

Tabel 1. Sequentiedata voor de vier meidoornspanners gebruikt in deze studie.

Table 1. Sequence data obtained for the four early moths used in this study.

Tabel 2. Resultaten van de de novo genoom assembly van de Franse meidoornspanner, *Theria primaria*. Ter vergelijking worden ook de getallen voor de genomen van de kleine wintervlinder, *Operophtera brumata*, en de fruitvlieg *Drosophila melanogaster* gegeven.

Table 2. Summary statistics for the de novo genome assembly of the French *Theria primaria* sample. The same statistics for the genomes of the winter moth, *Operophtera brumata*, and the fruitfly *Drosophila melanogaster* are provided for comparison.

waarde / variable	meidoornspanner, <i>Theria primaria</i>	meidoornspanner (alleen contigs langer dan 2.000 baseparen) / <i>Theria primaria</i> (only contigs longer than 2,000 base pairs)	kleine wintervlinder, <i>Operophtera brumata</i>	<i>Drosophila melanogaster</i>
Totale hoeveelheid sequentie ($\times 1.000.000$ baseparen)	537	156	638	144
Aantal scaffolds	2.352.258	45.701	25.801	1.870
N50 (DNA letters)	919	3.431	65.630	25.485.538



4. Voorbeeld van een 'alignment'. In het bovenste paneel staat een contig van de Franse meidoornspanner met daaronder de reads van de Walcherse meidoornspanner die dezelfde sequentie hebben. In het onderste paneel is ingezoomd op een deel van de contig. De lichte verticale baan markeert een één-letterverschil.

4. Example of an alignment. In the top panel, a contig from the genome of the French early moth runs along the top of the screen. Below that are sequence reads from the early moth from Walcheren that match the same genome region. The bottom panel shows a small part of this contig, with the light-coloured bar marking a single-nucleotide difference between the two.

populatie tot dit berekende middelpunt fors groter is: tussen de 0.33 (Frankrijk) en 0.60 (Utrecht) relatieve lengte-eenheden. Dit betekent dat de genetische verschillen tussen alle vier de meidoornspanners onderling groot zijn.

Discussie

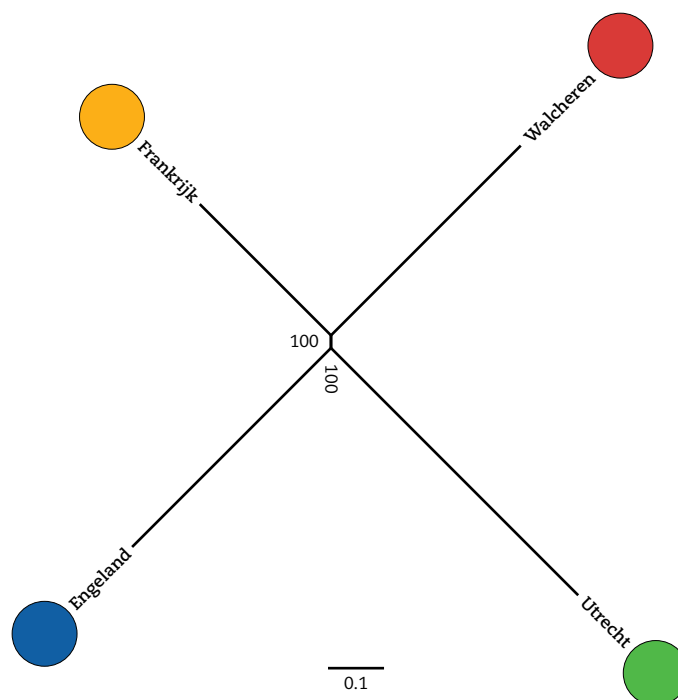
Het DNA van de meidoornspanner van Walcheren bleek bijna net zoveel verschillen te hebben met het DNA van de andere drie meidoornspanners als deze drie onderling. Dat maakt het onwaarschijnlijk dat de populatie in Walcheren recentelijk is aan komen waaien vanuit één van de andere populaties. Toch lijkt het DNA van de Walcherse meidoornspanner net iets meer op dat van die uit Frankrijk dan dat van de andere twee. Op basis hiervan is het meest waarschijnlijke scenario dat de Walcherse populatie een relict is van een populatie met een veel groter verspreidingsgebied, dat het nauwst in contact stond met de Franse populatie. Dat deze populatie pas recent ontdekt werd, komt waarschijnlijk omdat nachtvlinderonderzoekers zelden in de winter actief zijn. Aangezien het een substantiële populatie in een redelijk omvangrijk natuurgebied betreft zijn aanvullende beheers- en beschermingsmaatregelen niet nodig.

Uit elk van de vier populaties is slechts het DNA van één individu bekeken. Een belangrijke volgende stap zou kunnen zijn om te bepalen hoe de variatie binnen de Walcherse populatie zich verhoudt tot de variatie binnen en tussen de andere populaties. Daarvoor moet dan DNA uit meerdere individuen uit elk van de populaties worden bekeken. Verder zou het nuttig zijn om het DNA van meidoornspanners uit een veel groter gebied af te lezen. Mogelijk zijn de Walcherse meidoornspanners recent of historisch afkomstig uit een veel verder weg gelegen populatie.

In zo'n enorme hoeveelheid gegevens zoals we hier hebben verzameld zit natuurlijk veel meer interessante informatie dan we hier hebben beschreven. Hopelijk vinden we de komende jaren tijd om daar meer mee te doen. Vooral de vergelijking met het DNA van de kleine wintervlinder (Derks *et al.* 2015) biedt mogelijkheden. Zo zagen we dat de meidoornspanner, in tegenstelling tot de kleine wintervlinder, niet geïnfecteerd is met de veelvoorkomende bacterie *Wolbachia*: geen enkele scaffold van de meidoornspanner vertoonde gelijkenis met het DNA van de *Wolbachia* die in de kleine wintervlinder aanwezig is. *Wolbachia* manipuleert in veel insecten de reproductie door bijvoorbeeld mannen te doden of te veranderen in vrouwen. Of de *Wolbachia* in de kleine wintervlinder ook een dergelijk effect heeft, is vooralsnog niet bekend.

Het meidoornspannerproject illustreert wat er vandaag de dag met beperkte middelen mogelijk is met DNA-onderzoek. Het aflezen van het complete genoom lijkt wellicht overdreven. Al sinds de jaren 60 van de vorige eeuw worden verwantschap-

pen immers bepaald met behulp van een klein aantal variabele stukjes op het DNA, zogenaamde 'merkers'. Tegenwoordig is de zogenaamde DNA-barcoding in zwang, waarbij men gebruik maakt van een klein stukje DNA (bij dieren meestal afkomstig uit de mitochondriën) dat variabel is tussen soorten. Tussen populaties van dezelfde soort verschilt het echter meestal niet of nauwelijks. Bovendien is dit stukje DNA zo kort dat het niet genoeg informatie bevat om betrouwbare verwantschappen te bepalen. Het voordeel van de methode die we hier gebruikt hebben, is dat het volledig met standaardtechnieken wordt uitgevoerd. Voor DNA-isolatie en next generation sequencing maakt het niet uit of het DNA afkomstig is van een meidoornspanner, een aardbei of een zeeolifant – het labwerk blijft hetzelfde. Het gaat ook snel. DNA-isolatie kost een dag, het voorbereiden van het DNA voor aflezen twee dagen en het eigenlijke aflezen minder dan een week, maar dat laatste gebeurt volledig automatisch. Bij een studie met behulp van DNA-merkers kost het ontwikkelen en testen van geschikte



5. Fylogenetische verwantschap tussen de vier meidoornspanners, *Theria primaria*, gebaseerd op 128.203 één-letterverschillen in hun DNA. De lengte van een tak geeft het relatieve aantal één-letterverschillen weer.

5. Phylogenetic relationships between the four early moths, *Theria primaria*, based on 128,203 single-nucleotide polymorphisms in their DNA. Branch lengths are proportional to the number of SNPs.

merkers al snel maanden laboratoriumwerk. Daardoor valt het al snel duurder uit dan het botweg aflezen van al het DNA in een meidoornspanner.

Dankwoord

Martin Corley (Oxfordshire) en Jurriën van Deijk (Wageningen) leverden meidoornspanners uit respectievelijk Engeland en

Utrecht. Janine Marien (Vrije Universiteit Amsterdam) isoleerde het DNA. Henk Buermans en Emile de Meijer (Leiden Genome Technology Center, Leids Universitair Medisch Centrum) voerden de DNA sequencing uit. Bo Blanckenburg (Hogeschool Leiden) begeleidde de studenten bij hun project. Generade (m.m.v. Helma Kaptein en Tamara van Mólken) verzorgde de financiering en coördinatie van het project.

Literatuur

Derks MFL, Smit S, Salis L, Schijlen E, Bossers A, Mateman C, Pijl AS, de Ridder D, Groenen MAM, Visser ME & Megens HJ 2015. The genome of winter moth (*Operophtera brumata*) provides a genomic perspective on sexual dimorphism and phenology. *Genome biology and evolution* 7: 2321-2332.

Heppner JB 1991. Brachyptery and aptery in lepidoptera. *Tropical Lepidoptera* 2: 11-40.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T & Itoh T 2014.

Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research* 24: 1384-1395.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L & Wilson RK 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 22: 568-576.

Kraaijeveld K 2012. DIY-Genomics. *Entomologische Berichten* 72: 269.

Langmead B & Salzberg SL 2012 Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357-9.

Lee T-H, Guo H, Wang X, Kim C & Paterson AH 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC genomics* 15: 162.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G & Durbin R 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079.

Roff D 1990. The Evolution of flightlessness in insects. *Ecological Monographs* 60: 389-421.

Geaccepteerd: 14 december 2016

Summary

Identifying the origin of an isolated population of the early moth using state-of-the-art DNA technology

An isolated population of the early moth *Theria primaria* was discovered on the island of Walcheren in the South-west of the Netherlands. This occurrence was remarkable, given that female moths cannot fly. A project was set up in collaboration with the *LeveDNA!* foundation to trace the origin of this population using next generation DNA sequencing. The entire genomes of four early moths (from France, England, Utrecht and Walcheren) was sequenced. Based on nearly 130,000 single-nucleotide differences in their DNA, the early moth from Walcheren turned out to be most closely related to the one from France. However, the differences between all four early moths were of comparable magnitude. It is therefore unlikely that the population on Walcheren arrived recently from one of the other three populations. It seems more likely that this population is a relic of a much larger population that reached into France.



Ken Kraaijeveld & Mark Lammers
Vrije Universiteit Amsterdam
Afdeling Ecologische Wetenschappen
De Boelelaan 1085
1081 HV Amsterdam
ken@kenkraaijeveld.nl

Jesse Kerkvliet, Jolanda Essens, Dwin Grashof & Jens Zwart
Studenten Bioinformatica
Hogeschool Leiden
Zernikedreef 11
2333 CK Leiden

Johan den Dunnen
Stichting LeveDNA!
Plesmanlaan 1D
2333BZ Leiden

Jan Goedbloed
Sardijngeullaan 12
4371 PS Koudekerke