

eDNA metabarcoding amfibieën.



Colofon

Titel	eDNA metabarcoding amfibieën.
Tekst, foto's en samenstelling	K. van Bochove
In opdracht van	Sweco
Naam opdrachtgever	N. de Nijs
Rapportnummer	RA2017273
Datum oplevering rapport	21 augustus 2017
Aantal pagina's	9
Wijze van citeren	van Bochove K. 2017. eDNA metabarcoding amfibieën. Rapport RA2017273, Datura, Wageningen
Laboratorium analist	J. Rook



Datura Molecular Solutions BV

Gevestigd te:
Johan Buziastraat 55
6708 NR Wageningen
Nederland

Postadres laboratorium:
t.a.v. Datura (NCB)
Sylviusweg 72
2333 BE, Leiden
Nederland

0031(0)629455328
www.datura.nl
keesvanbochove@datura.nl

Inhoudsopgave

1. Doelstelling	4
2. Methode	5
2.1 Sampling	5
2.2 Laboratoriumanalyse	5
2.3 Data-analyse	6
2.4 Kwaliteitswaarborging	7
2.4.1 Hoe vals positieve waarnemingen voorkomen worden	7
2.4.2 Hoe vals negatieve waarnemingen voorkomen worden	8
3. Resultaten	9

1. Doelstelling

Vaststellen van amfibieën eDNA in watersamples in opdracht van Sweco, met nadruk op aantonen of uitsluiten van eDNA van alpenwatersalamander.

2. Methode

2.1 Sampling

De bemonstering is uitgevoerd door Sweco.

2.2 Laboratoriumanalyse

De eDNA monsters zijn geanalyseerd op de aanwezigheid van eDNA van amfibieën. Het analyseren van een eDNA monster vindt plaats in een aantal stappen. Eerst wordt het DNA uit een monster geconcentreerd en gezuiverd. Vervolgens wordt eDNA geamplificeerd (vermeerderd) met behulp van PCR. De PCR fragmenten zijn gezuiverd en een DNA library is voorbereid. De library is gesequenced met behulp van Next Generation Sequencing (HiSeq).

1. Het monster is geconcentreerd in een pellet door 35 minuten te centrifugeren (7830g, 6 °C). Vervolgens is het DNA in de pellet gezuiverd met behulp van de Qiagen Dneasy Blood & Tissue Kit. Storende stoffen als humuszuren kunnen detectie van het eDNA inhiberen wat kan leiden tot vals negatief resultaat. Met deze stap zijn dergelijke inhiberende stoffen zo veel mogelijk verwijderd.
2. Er is een controle uitgevoerd om na te gaan of eDNA detectie in een sample geïnhibeerd wordt. Dit is gedaan door een bekende hoeveelheid van een fragment artificieel DNA toe te voegen. Vervolgens is de concentratie van dit fragment artificieel DNA gemeten. Dit is zowel gedaan in een reactie waar een hoeveelheid sample aan toegevoegd is, als in een reactie waar geen sample aan toegevoegd is. Als DNA detectie in een sample geïnhibeerd wordt, dan is de gemeten concentratie artificieel DNA in de reactie waarin sample toegevoegd wordt lager ten opzichte van de reactie waaraan geen sample aan toegevoegd is. Met name in zuur water, waarin veel organische deeltjes aanwezig zijn kan inhibitie optreden. In een dergelijk geval wordt een extra zuivering stap uitgevoerd of wordt het sample verdund. Daarna wordt opnieuw gekeken of de inhiberende stoffen voldoende verwijderd zijn.
3. Vervolgens is het eDNA van amfibieën sterk vermeerderd middels PCR. Hiervoor zijn primers gebruikt die een kort fragment (~100 bp) van het mitochondriaal DNA vermeerderen. Deze primers bevatte een unieke tag (7 nucleotiden). Gedurende de bioinformatica analyse kunnen de reads aan de hand van deze tags toegewezen worden aan het juiste sample. De PCR is uitgevoerd met behulp van de TaqMan® Environmental Mastermix 2.0 (Life Technologies®). De PCR is uitgevoerd met twaalf replica's.
4. Door middel van gelelektroforese is vastgesteld of de PCR geresulteerd heeft in PCR producten van de juiste lengte. Middels van een tweede PCR zijn Illumina Nextera XT adaptors aan de PCR producten gezet. Vervolgens zijn de PCR producten samengevoegd. De pool van PCR producten van verschillende samples is gezuiverd.
5. Door middel van digital droplet PCR en een bioanalyser run is de concentratie van de pool van PCR producten vastgesteld. De concentratie van de PCR producten is verdund, om optimale clustering op de flow cell van de sequencer te bewerkstelligen.
6. De PCR producten zijn gesequenced met behulp van Next Generation Sequencing (HiSeq4400

platform, 150 bp paired-end). Hierbij worden miljoenen stukjes (zogenaamde reads) van het DNA uitgelezen. In deze stap wordt het fysieke DNA in het buisje dus vertaald in digitale reads.

2.3 Data-analyse

Eerst wordt een standaard verwerking van Illumina paired-end data uitgevoerd. Deze omvat de volgende stappen:

1. FASTQ sequence files zijn gegenereerd met behulp van de Illumina Casava pipeline.
2. Een eerste kwaliteitscheck is uitgevoerd door middel van Illumina Chastity filtering.
3. Vervolgens zijn reads welke PhiX controle bevatte verwijderd.
4. (Restanten van) de sequencing adapters zijn uit de reads geknipt.
5. De kwaliteit van de overgebleven reads is getest met de FastQC tool.

Vervolgens worden de sequenties geanalyseerd met behulp van het software package Obitools. Deze pipeline resulteert uiteindelijk in een tabel waarin voor elk sample aangegeven is hoeveel reads er van elke soort gedetecteerd zijn. Omdat er behoorlijke rekenkracht nodig is voor het verwerken van de sequencing data wordt een workstation gebruikt welke beschikt over 2 six core processoren met hyper-threading en 48 Gb Ram-geheugen. De volgende stappen zijn doorlopen:

1. **Illuminapairedend**: Genereren van een consensus sequentie op basis van de forward en reverse read.
2. **Obigrep**: sequenties die slecht aligned werden zijn verwijderd.
3. **Ngfilter**: Op basis van de gebruikte primers, en de tags die toegevoegd zijn in de eerste en tweede PCR zijn alle sequenties toegewezen aan het corresponderende sample.
4. **Obisplit**: Met deze stap worden de reads van de verschillende samples van elkaar gescheiden. Elk sample wordt weggeschreven naar een eigen tekstbestand.
5. **Obiuniq**: Om de dataset die nu nog bestaat uit miljoenen reads hanteerbaarder te maken zijn alle sequenties samengevoegd.
6. **Obigrep**: Sequenties die binnen een sample minder vaak dan 1500 keer voorkomen zijn verwijderd.
7. **Obiclean en obigrep**: Waarschijnlijke sequencing fouten worden uit de dataset gefilterd.
8. **Obitag**: De obitag tool gebruikt om de sequenties te matchen met de referentie database. Deze database bevat uitsluitend gevalideerde sequenties. Alle in Nederland voorkomende amfibieën zijn hierin opgenomen. Deze database is opgebouwd uit referentie materiaal dat verzameld is door Datura, aangevuld met sequenties afkomstig van onder andere vouchers uit museum collecties. Uitsluitend reads die 98% of meer overeen komen met één van de sequenties in de referentie database zijn beschouwd als hit. Sequenties die minder dan 98% overeenkomen met de gevalideerde referentie database zijn gematched met referentie sequenties die afkomstig

zijn van NCBI Genbank. Deze referentie database is opgebouwd door de ecoPCR te runnen op alle mitochondriale sequenties van vertebraten, welke gedownload zijn van NCBI Genbank. Zodoende kunnen soms ook niet verwachte soorten (bijvoorbeeld een onbekende exoot) gedetecteerd worden.

9. **Obitab:** Tenslotte worden de resultaten geëxporteerd naar een .tab file. Deze file kan geopend worden in Excel.

In Excel wordt een handmatige check uitgevoerd of er sequenties aanwezig zijn die voor 90% of meer overeenkomen met een soort in de referentiedatabase, maar niet terug komen in de resultaten. Dat kan erop wijzen dat er een soort/ genvariant gedetecteerd welke niet in de referentiedatabase aanwezig is.

2.4 Kwaliteitswaarborging

2.4.1 Hoe vals positieve waarnemingen voorkomen worden

Het optreden van zowel vals positieve als vals negatieve waarnemingen wordt tot het minimum beperkt. Vals positieve waarnemingen kunnen op twee manieren ontstaan:

- Er vindt contaminatie plaats in het laboratorium;
- Er vindt contaminatie plaats in het veld.

Hieronder wordt aangegeven hoe vals positieve waarnemingen voorkomen worden. Omdat de kans op vals positieve waarnemingen zeer klein is, kunnen we niet exact kwantificeren hoe groot de kans daadwerkelijk is. Datura kan daarom niet 100 % zeker garanderen dat vals positieve waarnemingen nooit optreden.

Om vals positieve waarnemingen te voorkomen werkt Datura in een specifiek voor (e)DNA ingericht laboratorium omgeving en worden strikte procedures gevolgd:

1. Verschillende onderdelen van de analyse workflow worden uitgevoerd in fysiek gescheiden laboratorium ruimtes. Het samenstellen van de eDNA sample kits en het voorbereiden van de PCR reagentia vindt plaats in een **DNA clean room**. Dit is een ruimte waarin geen DNA samples aanwezig zijn. Zodoende kunnen we garanderen dat er geen DNA aanwezig is in de eDNA sample kits en de reagentia die later gebruikt worden in de eDNA analyses. Het extraheren van de eDNA samples gebeurt in een **pre-PCR laboratorium**. Dit is een ruimte waarin uitsluitend lage concentraties DNA aanwezig zijn. Vervolgens worden hier de eDNA samples samen met de PCR reagentia in een 96-well plaat gepipetteerd. Deze plaat wordt luchtdicht afgesloten en naar een **post-PCR laboratorium** gebracht. De PCR vindt plaats in het **post-PCR laboratorium**. In dit laboratorium wordt het eDNA vermeerderd en hier zijn dus hoge concentraties DNA aanwezig.
2. Er wordt een **unidirectionele workflow** gehanteerd om contaminatie van de DNA clean room en het pre-PCR laboratorium te voorkomen. Dit houdt in dat materialen die eenmaal in het post-PCR laboratorium geweest zijn niet meer terug mogen naar de DNA clean room en het pre-PCR laboratorium. Ook mogen medewerkers van Datura niet dezelfde dag van een post-PCR laboratorium terug naar de DNA clean room en het pre-PCR laboratorium.
3. In iedere analyse worden **controle analyses** uitgevoerd. Zo worden er samples geëxtraheerd waaraan geen slootwater wordt toegevoegd (zogenaamde extractie controles). In de PCR worden naast de extractie controles ook negatieve PCR controles meegenomen. Zodoende kan heel nauwkeurig gemonitord worden of er inderdaad geen contaminatie optreedt.
4. Er wordt geen **positieve controle** van de doelsoort meegenomen in de PCR detectie. In de

praktijk is in alle samples eDNA van amfibieën of andere vertebraten aanwezig. Het is dus niet nodig om een positieve controle mee te analyseren. Om het risico op contaminatie te minimaliseren wordt er daarom geen positieve controle meegenomen.

Om contaminatie in het veld te voorkomen worden de volgende maatregelen genomen:

1. Het **sampling protocol** van Datura wordt gevolgd. Dit protocol schrijft een specifieke werkwijze voor. In de praktijk is gebleken dat er geen contaminatie plaats vindt als dit protocol gevolgd wordt;
2. Er dient rekening gehouden te worden met **waterverplaatsingen**. De sampling wordt daarom uitgevoerd op een moment dat er weinig stroming is. Zo worden eDNA samples niet verzameld direct na (hevige) regenval. Ook wordt er rekening gehouden met kunstmatig opgewekte stroming, bijvoorbeeld bij wisseling van zomer- naar winterpeil.

2.4.2 Hoe vals negatieve waarnemingen voorkomen worden

Naast vals positieve waarnemingen kunnen ook vals negatieve waarnemingen optreden. Uit diverse **validatie studies** blijkt dat de eDNA detectiekans voor amfibieën en vissen in stilstaande en zwak stromende wateren 90-99% is. Er is dus altijd een kleine kans dat eDNA niet gedetecteerd wordt, ook al is de doelsoort wel aanwezig. Maatregelen die genomen worden om vals negatieve waarnemingen te voorkomen:

1. Per sample worden **20 subsamples** verzameld. Hiermee wordt de kans vergroot dat eDNA in het sample terecht komt.
2. Een zeer gevoelige **PCR detectie** wordt uitgevoerd met behulp van **12 replica's**. Wanneer minder replica's uitgevoerd worden kan er minder gevoelig gedetecteerd worden. Meer dan 12 PCR replica's leidt echter niet tot gevoeliger detectie;
3. Gebruik van een **zeer korte merkers** van maximaal 110 basepaar.
4. In ieder sample wordt **vastgesteld of de qPCR detectie geïnhibeerd** wordt door storende stoffen. Indien dit het geval is wordt er een **extra zuiveringstap** uitgevoerd. Vervolgens wordt nogmaals getest of de inhiberende stoffen er inderdaad geen inhibitie meer optreedt (zie methode voor een uitgebreidere beschrijving).
5. Van ieder sample worden **minimaal 500.000 reads** geanalyseerd. Zodoende kunnen ook soorten gedetecteerd worden die slechts in lage dichtheden aanwezig zijn.

3. Resultaten

3.1 Controles

Er is geen eDNA gedetecteerd van amfibieën in de negatieve extractie controles en de negatieve PCR controle reacties waar geen sample aan toe gevoegd is. Dit geeft aan dat er hoogst waarschijnlijk geen contaminatie heeft plaatsgevonden gedurende het laboratorium analyseproces.

Humuszuren kunnen een PCR reactie inhiberen wat kan leiden tot vals negatief resultaat. Daarom wordt altijd een qPCR met een interne controle geanalyseerd om vast te kunnen stellen of er sprake is van inhibitie. Er werd geen significante afwijking gevonden in de Cq-waarde van de interne controles waar een sample aan toegevoegd is ten opzichte van de reacties waar geen sample aan toegevoegd is. Dat geeft aan dat er geen sprake was van inhibitie.

3.2 Sequencing resultaten

In alle samples werd het minimum van 500.000 reads gehaald (zie tabel 1). Er zijn twee soorten aangetroffen (zie tabel 2). Er is geen eDNA van alpenwatersalamander aangetroffen.

Tabel 1. Overzicht van de reads die geanalyseerd zijn. Het aantal reads amfibieën geeft het aantal reads weer van amfibieën dat gedetecteerd is. Deze reads komen met 100% overeen met de referentie database. De overige reads betreffen reads afkomstig van andere soorten (vissen/vogels/zoogdieren waaronder humaan DNA) die gedetecteerd zijn. Daarnaast omvat dit ook reads van amfibieën die minder dan 98% overeenkomen met een sequentie in de referentie database. Deze sequenties bevatten een PCR/ sequencing fout.

	Sample 675
Reads amfibieën	26.267
Overige reads	2.366.340
totaal	2.392.607

Tabel 2. In deze tabel zijn de percentages reads weergegeven die afkomstig zijn van iedere gedetecteerde soort, ten opzichte van de totale hoeveelheid eDNA van amfibieën.

Wetenschappelijke naam	Nederlandse naam	Sample 675; # reads	Sample 675; % reads
Pelophylax kl. Esculentus / ridibundus	Bastaardkikker/meerkikker	21302	81,1%
Bufo bufo	Gewone pad	4965	18,9%
Totaal		26.267	