

BCCM GEN-ERA

BCCM collections in the genomic era

Contract - B2/191/P2/BCCM GEN-ERA

SAMENVATTING

Context

De Belgische Gecoördineerde Verzamelingen van Micro-organismen (BCCM) is een uniek initiatief van de Belgische federale regering en haar Wetenschapsbeleid (Belspo). BCCM is een Biologisch Resource Centrum (BRC) dat microbiële en genetische hulpbronnen bewaart en ter beschikking stelt ter ondersteuning van de biowetenschappen en de biotechnologiesector in fundamenteel en toegepast onderzoek.

Prokaryote en eukaryote micro-organismen vertegenwoordigen het grootste deel van de op aarde aanwezige diversiteit en zijn te vinden in elke omgeving die leven kan ondersteunen. Ze spelen een belangrijke rol in een groot aantal functies, van ecosystemen tot de mens, en dragen bij tot talloze toepassingen. Toch onthult de huidige kennis in de microbiologie nauwelijks het topje van de ijsberg en rest er nog een aanzienlijke taak om de ongekende diversiteit verder te ontdekken. Cultuurcollecties spelen hierbij een essentiële rol door het isoleren, kweken, identificeren, bewaren en distribueren van de kweekbare diversiteit.

Sinds de oprichting in 1983 heeft BCCM zich verankert als een toonaangevende cultuurcollectie onder de Europese BRC's. De sterkte van BCCM ligt onder andere in het aanbod van een aantal belangrijke voorzieningen, waaronder een ISO 9001-certificering, een website met online catalogus, een systeem voor het beheer van laboratoriuminformatie en de erkenning als internationale depositaris. Verder volgt BCCM de meest recente ontwikkeling in de microbiologie op de voet, dit om microbiële stammen op een efficiënte manier te kunnen analyseren. Tegenwoordig zijn er grote stappen vooruitgezet in het microbiële onderzoek door nieuwe technieken in de genomica, waaronder bv. sequenceren van volledige genomen. Zo kan de volledige genetische informatie van een organisme achterhaald worden, wat steeds vaker gevraagd wordt in verschillende disciplines. Het verwerven van know-how in de genomica is dus cruciaal voor BCCM om een belangrijke BRC te blijven, voor toekomstige nationale en internationale samenwerking en om toekomstige onderzoeksvragen te beantwoorden.

Doelstellingen

De eerste doelstelling van het BCCM GEN-ERA-project was het implementeren van expertise in genomica in de BCCM-collecties. De uitdaging hierbij was het omgaan met en analyseren van genomische big data. Dit vereiste de installatie van specifieke bio-informaticastructuren (hardware) en software waarvoor de BCCM-wetenschappers moesten worden opgeleid om het gebruik ervan op lange termijn te garanderen. De nadruk lag vooral op de sequenceren van hele genomen, omdat de bepaling hiervan de expertise van de collecties op het gebied van taxonomie en fylogenie ondersteunt en tegelijkertijd potentiële functionele analyses mogelijk maakt. Bovendien is het aanbieden van stammen waarvan het volledige genoom is gesequeneerd noodzakelijk om te voldoen aan de behoeften van de gebruikers van de BRC's en vormt het dus een toegevoegde waarde voor de zichtbaarheid en aantrekkelijkheid van BCCM.

Het BCCM GEN-ERA-project had ook tot doel specifieke onderzoeksvragen te beantwoorden over de microbiële diversiteit van de BCCM-collecties (i.e. bacteriën, mycobacteriën, cyanobacteriën, gisten en schimmels) en in het bijzonder micro-organismen die een maatschappelijke impact hebben (bv. micro-organismen die geassocieerd zijn met bestuivende insecten, die ziektes veroorzaken of die bioactieve verbindingen produceren). Vijf van de

zeven BCCM-collecties waren betrokken in het project en werkten samen met de onderzoeksgroep Eukaryotic Phylogenomics van de Universiteit van Luik, die de nodige expertise in bio-informatica en (fylo)genomica inbracht.

Conclusie

Twee verschillende bio-informatica platformen werden overwogen voor de analyse van de sequencing data, namelijk Galaxy en Nextflow. Beide werden getest en vergeleken op hun prestaties, geschiktheid, gebruiksvriendelijkheid en naleving van de FAIR-principes (i.e. Findable, Accessible, Interoperable, Reusable).

Galaxy is een bio-informatica-webplatform dat tot doel heeft genomische analyses op een laagdrempelige manier toegankelijk te maken voor onderzoekers. Dit wordt gerealiseerd via een gebruiksvriendelijke grafische interface. De installatie van een eigen "BCCM" Galaxy werd getest, maar er waren verschillende hiaten in de beveiliging die de installatie bemoeilijkten. Een systeembeheerder zou daarom nodig zijn om de infrastructuur te onderhouden. Deze moeilijkheden beperkten ook de interoperabiliteit en het hergebruik van de bio-informatica-workflows. Bovendien waren niet alle programma's beschikbaar die nodig waren voor de analyses. Om onder andere die redenen werd de voorkeur gegeven aan Nextflow. Dit platform werkt aan de hand van commandoregels in plaats van een user-interface. In totaal werden 14 Nextflow-workflows geïmplementeerd, ondersteund door 11 Singularity-containers. Deze dekken de algemene genomics-behoefte voor microbiële taxonomie en metabole modellering van de BCCM collecties. De workflows zijn namelijk geschikt voor de analyse van prokaryoten en kleine eukaryoten en dit op een volledig reproduceerbare manier. De workflows worden aan de eindgebruikers geleverd als een programma dat met een enkele commandoregel kan worden uitgevoerd. Deze "GEN-ERA toolbox" werd vrij beschikbaar gesteld via de GitHub repository <https://github.com/Lcornet/GENERA> die ook een uitgebreide documentatie voor de gebruikers bevat. Nextflow voldeed dus aan de meeste criteria voor een langdurig en FAIR gebruik van de bio-informatica-infrastructuur bij BCCM. Het enige nadeel ten opzichte van Galaxy is de gebruiksvriendelijkheid. Het werken met commandoregels is minder intuïtief en vereiste specifieke trainingen, maar kon evenwel tot een minimum worden beperkt dankzij de Singularity-containers.

De implementatie van de Nextflow infrastructuur gebeurde in samenwerking met de Universiteit van Luik en vergemakkelijkt genomische analyses bij BCCM, waardoor prangende onderzoeksvragen konden beantwoord worden. Zo werden onder andere schimmels die huidinfecties veroorzaken, niet-tuberculose mycobacteriën, bacteriën en gisten uit de darmen van bestuivende insecten (bijen en hommels) en cyanobacteriën die biologische activiteiten vertonen onderzocht. Deze analyses zorgden voor doorbraken in hun respectieve gebieden en openden nieuwe perspectieven voor toekomstig onderzoek.

Het BCCM GEN-ERA project heeft ertoe geleid dat er een expertise is opgebouwd binnen het BCCM-consortium op het gebied van genomica. Dit gebeurde door het creëren van een bio-informaticakader, door genomica-instrumenten ter beschikking te stellen en door analyse-vaardigheden aan te leren bij de BCCM-wetenschappers.. Deze investering werd dus gedaan met het oog op de lange termijn waarbij genomica een permanente activiteit binnen het BCCM-consortium kan worden. Een belangrijk hulpmiddel hierbij is de vrij toegankelijke BCCM GEN-ERA GitHub repository. Het is een webportaal dat is ontworpen om online en gratis toegang te bieden tot bio-informatica-tools en hun handleiding voor de analyse van genomische gegevens van prokaryote en eukaryote micro-organismen. Het centraliseert software voor onder andere genoomassemblage, genoomannotatie, orthologie-inferentie, fylogenomica en metabole modellering van genomen. Het werkt als een gemeenschappelijk platform dat kan worden gebruikt door alle BCCM-collecties en wetenschappers, en ook door andere faciliteiten (bv. BRC's, microbiologische laboratoria) die in dezelfde onderwerpen geïnteresseerd zijn.

Trefwoorden

Genomica; BCCM; micro-organismen; genomen; cultuurcollectie; bio-informatica; taxonomie; fylogenie; moleculaire evolutie; biodiversiteit.