**Beschrijving onderzoeksplan**

Voor het onderzoek gebruiken we de sequentie data die ons is aangeleverd door het HAN Biocentre. Deze data is verkregen doormiddel van een metagenomics onderzoek naar de verschillende micro-organismen in de compost die gebruikt wordt om champignons te telen. Voor het analyseren van de sequentie data gebruiken we de tool Blastx (Altschul S. G.-4., sd) van de NCBI database. De keuze voor Blastx (Altschul S. G.-4., sd) komt omdat dit naar verwachting de meest nauwkeurige resultaten geeft binnen de beperkte tijd die het onderzoek duurt. Mochten de resultaten van Blastx (Altschul S. G.-4., sd) der mate onnauwkeurig zijn. Dan stappen we over op tBlastx (Altschul S. G.-4., sd) dit neemt meer tijd in beslag maar is nauwkeuriger. Om de nauwkeurigheid van de sequenties te verifiëren gebruiken we een scorematrix. Hierbij gaan we er vanuit dat de meeste sequenties verwant zijn, maar we gaan er ook vanuit dat er micro-organismen tussen zitten die geen verwantschap delen. Omdat deze matrix voor de verwachte data de beste resultaten geeft, gebruiken we de BLOSUM62 matrix. Bij Blastx (Altschul S. G.-4., sd) gebruiken we als database de non-redundant GenBank, eukaryoten worden uitgesloten van de resultaten. Dit omdat ons onderzoek zich alleen richt op de prokaryoten en de mogelijk interessante eiwitten die zij produceren en daarmee van invloed kunnen zijn op de champignongroei. Voor elke gap in de sequentie rekenen wij 11 minpunten voor het openen van een gap en 1 minpunt voor elke positie in de gap. Het opslaan van de gevonden matches doen we met een “cut-off” op een E-value van 1\*10-10. Alle gevonden matches met een lagere E-value worden als te onnauwkeurig beschouwt en daarom niet meegenomen in het verdere onderzoek. Andere “cut-off” criteria zijn een minimale identity van 75% (bij eiwitten 25%), een positive percentage hoger dan 70% en een coverage hoger dan 80%. Alleen de eerste 10 gevonden matches die aan deze criteria voldoen worden opgeslagen in een door ons gemaakte Oracle database. Doormiddel van een programma geschreven in de programmeertaal Python (Rossum, 1995) worden de resultaten in de database verwerkt. Deze data is vervolgens uit te lezen via een website die weer met behulp van Python (Rossum, 1995) verbonden is met de database. De opzet van deze website word geschreven in HTML. De website laat de gebruiker de verschillende micro-organismen opvragen en de functies/ genen van de gevonden eiwitten. Ook laat het zien welke domeinen de eiwitten bevatten en wat de locatie van de eiwitten is in de cel. Om de applicatie te testen worden er 10 forward en 10 reverse “reads” als test ingevoerd. Doordat deze “reads” al bij ons bekend zijn kunnen wij hier aan zien of de applicatie goed functioneert.