

Code : BFVH3TBIO	Tentamen: Proeftentamen Theorie van Bioinformatica 2 NIET VERTROUWELIJK	
Datum: 29-10-2018	Tijd: 10:30-12:30	School: ILST
Lokaal:	Klas: BFV3	Duur: 1 1/2 uur
Docent : Martijn Herber		Aantal pagina's: 2
Het secretariaat is tijdens het tentamen te bereiken onder nummer: 050 – 595 45 69.		
Hulpmiddelen: Kladpapier		Overig hulpmiddelen: Geen
Opgave inleveren: Ja		
Kladpapier inleveren: Ja		
Bijzonderheden: Bijzonderheden <hr/>		
Naam student: Klas:		Studentnummer:

Bij het beantwoorden van de vragen, schrijf duidelijk, houd het kort, **maar wees altijd volledig**.

- 1) Leg uit wat de “molecular clock hypothesis is”, hoe zich dit relateert aan afstanden uit een alignment en waarom we hierop nog een correctie moeten uitvoeren om een “eerlijk” beeld te krijgen van de verstreken tijd tussen twee sequenties die van een gemeenschappelijke voorouder afstammen. (10pt)
- 2) Beschrijf het UPGMA algoritme om fylogenetische bomen te construeren. Gebruik tenminste de concepten *clade*, *OTU*, *branch length*, en *rooted/unrooted*. (12pt)
- 3) Leg uit wanneer je een *unrooted phylogenetic tree* zou gebruiken, en wanneer een *rooted*. Welke aanname maak je bij het gebruik van een *rooted tree*? (10pt)
- 4) a) Leg uit waarom je bij de *Bruijn graph assembly* altijd meerder *kmer-sizes* moet uitproberen, om de “beste” *assembly* te verkrijgen. (5pt)
b) Wat is de meest gebruikte maat om de kwaliteit van *assembly* te meten, en wat betekent deze? (5pt)
- 5) Wat wordt opgeslagen in het *Variant Call Format*? Hoe heet de stap in de pipeline die deze files produceert? Waarom is het belangrijk voor de stap in je analysepijplijn die deze files produceert om genoeg *coverage* van het genoom te hebben? (6pt)
- 6) Beschrijf de algemene onderdelen van een *graaf* of wiskundig *netwerk*. Hoe worden deze toegepast in *assembly* bij een *Overlap Layout Consensus methode*? (10pt)
- 7) Beschrijf hoe *paired-end sequencing* je kan helpen om je *assembly* van losse *contigs* verder af te maken. Beschrijf twee lab-methoden waarmee je uiteindelijk kunt proberen om het genoom helemaal af te maken. (12pt)
- 8) Stel, je hebt *RNAseq data* van een organisme waar nog geen volledig genoom van bekend is. Hoe ga je te werk om toch iets over de *genexpressie* te zeggen? (10pt)
- 9) Wat wil de term “*Eulerian path door een netwerk*” zeggen en hoe wordt zo'n *netwerk* gebruikt om bij *assembly* “*contigs* uit te lezen” ? (8pt)
- 10) Wat is het verschil tussen “*positieve*” en “*negatieve*” selectie en hoe kun je dit bepalen aan de hand van de “*synonymous*” en “*non-synonymous*” *substitution rates*? (12pt)