

Thomas Dooms
Ward Gauderis
Kato Starckx
Lauren Van Nimmen

Inleiding tot Computacionele Biologie
Faculteit Wetenschappen
UAntwerpen
17/05/2021

Evolutionaire genetica

Genetisch polymorfisme

Inhoudsopgave

1	Inleiding	3
2	Model	5
3	Implementatie	8
4	Analyse vanuit de biologie	10
4.1	Cichliden	10
4.2	Evolutie van Cichliden	10
4.3	Genetisch polymorfisme via balancerende selectie	11
4.4	Simulaties	11
4.5	Toepassing simulaties	13
5	Analyse vanuit de wiskunde	14
5.1	Invloed selectiesterkte	15
5.2	Invloed generatieoverlap	15
5.3	Invloed aantal en grootte patches	15
5.4	Invloed ruimtelijke verschillen en migratiekans	15
5.5	Invloed tijdelijke correlatie	15
5.6	Invloed populatiegrootte	16
6	Conclusie	17
7	Taakverdeling	18
8	Referenties	19

1 Inleiding

Na de lezing van Prof. Hannes Svoldal over wat we uit onze genoomsequenties kunnen leren over de geschiedenis van de mens, was onze groep erg geïntrigeerd door het idee om onze genealogie te schetsen aan de hand van coalescentiesimulaties. Dit zijn simulaties waarin vanuit een steekproef van de huidige populatie vertrokken wordt, en een gemeenschappelijke voorouder bepaald wordt door terug in de tijd te gaan en als het ware de verschillende individuen te laten 'samensmelten'.

In deze lezing werd ook kort gehint naar het bestaan van een ander soort simulaties die gebruikt worden in de populatiegenetica, zogenaamde forward-time simulaties. Deze simulaties werken in de omgekeerde richting en modelleren natuurlijke selectie met de tijd mee. Forward-time simulaties zijn krachtiger en flexibeler, maar vereisen ook meer rekenkracht.

Vanuit onze achtergronden van biologie, wiskunde en computerwetenschappen, wilden we hier meer inzicht in krijgen. Vandaar onze onderzoeksvraag:

Hoe kunnen forward-time simulaties bijdragen tot het verklaren van genetisch polymorfisme binnen de evolutionaire genetica?

Het domein van de evolutionaire genetica bestudeert de evolutie van het genoom van een soort. Een grote uitdaging hierbij is het verklaren van de genetische variatie binnen een populatie.

Genetisch polymorfisme is een soort van genetische variatie. Men spreekt van genetisch polymorfisme wanneer twee of meer verschillende allelen tegelijkertijd aanwezig zijn op de locus van een gen binnen eenzelfde populatie. Dit zijn dus homologe allelen die worden overgeërfd van generatie op generatie, maar die wel verschillend van elkaar zijn, wat betekent dat er een mutatie moet zijn opgetreden in de lijn tussen de individuen met deze ongelijke allelen en hun gemeenschappelijke voorouder.

Om te spreken van genetisch polymorfisme, moet het allel voldoende verschillend (discreet) zijn, moet het voldoende lang aanwezig zijn binnen de populatie en moet de deelpopulatie met het verschillende allel een aanzienlijk deel van de totale populatie innemen. Wanneer deze meer dan 1% van de populatie inneemt, spreekt men van genetisch polymorfisme. Dit kan dan immers wijzen op een alternatief allel dat niet enkel behouden blijft door toevallige genetische mutaties, maar wel degelijk tegelijk langdurig naast een ander allel binnen een populatie aanwezig is.

Genetisch polymorfisme kan op verschillende schalen voorkomen, van een enkele nucleotide tot repeterende sequenties van basenparen. Dit verschijnsel vertaalt zich dan naar fenotypische diversiteit binnen de populatie waarbij geen van de fenotypes een duidelijk voordeel heeft qua natuurlijke selectie.

Bekende voorbeelden van genetisch polymorfisme zijn de patronen en kleur van de dekschilden van lieveheersbeestjes of huisjes van slakken, polymorfismen van het leverenzym CYP1A1 bij de mens wat kan leiden tot een hogere kans op hartklachten en hypertensie en polymorfismen in de menselijke genen verantwoordelijk voor sikkelcelanemie.

Genetisch polymorfisme wordt in stand gehouden door zogenaamde balancerende selectie, wat een verzamelnaam is voor de verschillende processen in de natuurlijke

selectie die het mogelijk maken dat meerdere allelen naast elkaar bestaan in een populatie. Het zijn deze selectieprocessen die we verder analyseren door middel van een forward-time simulatie.[1, 2]

Vanuit onze achtergronden hebben we ervoor gekozen om ons te verdiepen in de werking van forward-time simulaties en onze eigen forward-time simulatie te implementeren. Hiervoor baseren we ons voornamelijk op het onderzoek van Prof. Svardal zelf [3, 4]. Vervolgens bouwen we hierop verder door verschillende real-time visualisaties en parameters toe te voegen opdat het onderliggende model ook intuïtief kan worden gepresenteerd en uitgelegd aan buitenstaanders. Tot slot analyseren we met behulp van dit model en de bronnen de balancerende selectie. Dit doen we zowel vanuit het perspectief van de biologie met de use case van de cichliden, als vanuit het wiskundige perspectief waarbij we ons baseren op principes uit de evolutionaire speltheorie.

2 Model

Om een eenvoudig interpreteerbaar en parametrizeerbaar model te ontwerpen, maken we verschillende assumpties.

We modelleren slechts een enkele erfelijke eigenschap binnen een populatie waarvoor geldt dat het domein waarin deze eigenschap kan worden uitgedrukt alle reële getallen omvat. We veronderstellen dat dit een genetische additieve eigenschap is, wat betekent dat het uiteindelijke fenotype deterministisch bepaald wordt door het gecombineerde effect van meerdere genen. De waarde van de eigenschap voor een individu is gelijk aan de som van de allelen op de verschillende relevante loci van deze genen. Volgens het model zijn deze loci gelijk verdeeld over lineaire chromosomen. Deze loci kunnen eveneens een oneindig aantal verschillende allelen bevatten, waarbij elk allel een reëel getal is (in de implementatie wordt wel gewerkt met een stapgrootte).

Vermits we additieve genetica veronderstellen, sluit dit uit dat bepaalde allelen dominant kunnen zijn over andere. Modellen die ook met dominantie rekening houden, kunnen enkel diploïde soorten simuleren omdat de aanwezigheid van dominante allelen steeds heterozygote over homozygote individuen begunstigt. Het model dat wij hanteren heeft hier het voordeel dat zowel clonaal reproducerende, haploïde, als seksueel reproducerende, diploïde, soorten kunnen worden gesimuleerd. Op deze manier kan het model dit zogenaamde heterozygootvoordeel, wat zelf een bekend balancerend selectieproces is, illustreren.

Op populatieniveau is de simulatie dan weer gebaseerd op het eilandenmodel. Dit model veronderstelt dat de populatie verspreid is over een eindig aantal (figuurlijke) eilanden, patches. De individuen binnen eenzelfde patch komen met elkaar in aanraking en kunnen dus onderling reproduceren (in het geval van seksuele voortplanting). De patches zelf zijn afgezonderd van elkaar, maar migratie van individuen tussen patches is mogelijk in een gecontroleerde mate. Elke patch heeft een aantal omgevingsvariabelen (temperatuur, vochtigheid, ...) die kunnen evolueren in de tijd en verschillen van patch tot patch. Om het eenvoudig voor te stellen, veronderstellen we dat al deze omgevingsfactoren samengevat worden door één reëel getal per patch per tijdseenheid. Hoe dichter het fenotype van een individu bij de omgevingsfactor van zijn patch ligt, hoe hoger zijn voortplantingssucces (fitness). Op deze manier kan ook de invloed van temporele en spatiale variatie in de omgeving op het polymorfisme binnen de populatie bestudeerd worden.

Verder veronderstelt het model dat de populatiegrootte binnen de patches constant blijft in de tijd, wat de implementatie efficiënter maakt. Dit is echter ook een redelijk realistische veronderstelling vermits populatiegroei altijd wel beïnvloed wordt door bepaalde constante limiterende factoren zoals ruimte of voedsel.

De simulatie zelf verloopt in de volgende stappen:

- De omgevingsfactoren per patch worden geüpdatet.

- De volwassen populatie plant zich voort binnen elke patch, clonaal of seksueel. We veronderstellen dat de genetische eigenschap die we bestuderen reeds vroeg aanwezig is in de nieuwe generatie, zodat de nakomelingen onmiddellijk aan de zogenaamde lokale levensvatbaarheidsselectie kunnen worden onderworpen. Op deze manier kunnen werken met het reproductiesucces van een ouder in een patch, wat het aantal nakomelingen van deze ouder na de levensvatbaarheidsselectie is. Hoe dichter het fenotype x van de ouder ligt bij de optimale omgevingfactor θ in de patch in die iteratie, hoe groter zijn reproductiesucces $r(x, \theta)$. Hier hanteren we Gaussische stabiliserende selectie, waarbij het reproductiesucces een Gaussische functie is.

$$r(x, \theta) = me^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

m staat hier voor het maximum aantal nakomelingen en σ voor de selectiesterkte (hoe groter σ hoe zwakker de kracht van de selectie). Deze functie zorgt ervoor dat wanneer er erg sterke plotse veranderingen optreden in de omgeving, dit nefast kan zijn voor het aantal nakomelingen met hetzelfde fenotype.

- In het geval van seksuele voorplanting simuleren we ook het optreden van mogelijke recombinaties bij het vormen van de gameten. Hierbij berekenen we de kans op een crossing-over per locus vanuit de kans dat er een of meer recombinaties gebeuren. Vervolgens smelten de haploïde gameten twee per twee per patch samen tot diploïde zygoten.
- In de nieuwe individuen kunnen er eveneens mutaties in de allelen optreden met een bepaalde kans. De grootte van een mutatie wordt getrokken uit een normaalverdeling met instelbare standaardafwijking en is steeds een veelvoud van de eerder vernoemde stapgrootte.
- Na de reproductie zal in elke iteratie van de simulatie een fractie van de volwassen generatie sterven. De kans op sterven ligt vast voor elk individu en staat los van hun fenotype of de omgeving. Deze kans is gelijk aan $1 - \gamma$ waarbij γ de parameter is die staat voor generatieoverlap. Dit komt overeen met de veronderstelling dat ons model enkel zachte selectie simuleert, wat inhoudt dat er geen grotere kans op sterven bestaat wanneer een volwassen individu een slechtere fitness heeft.
- Vermits de populatiegrootte steeds constant blijft per patch, moet de jonge generatie vervolgens nog de dichtheidsregulatie per patch ondergaan. Per patch kunnen er slechts evenveel nakomelingen overleven als er ouderen gestorven zijn. Deze selectie gebeurt lokaal en onafhankelijk van fenotype.
- Vervolgens zal nog een percentage van de overlevende nakomelingen vertrekken uit hun patch en migreren naar een willekeurige patch. Hier wordt verondersteld dat de kans dat een bepaald individu aankomt in een bepaalde patch proportioneel is ten opzichte van de grootte van de (resterende) jonge gene-

ratie in die patch. Dit vereenvoudigt de implementatie, maar is ook redelijk, vermits dit eveneens proportioneel is met de hoeveel resterende levensmiddelen in die patch. Alle gemigreerde individuen die aankomen in een patch die zijn populatiecapaciteit reeds heeft bereikt, zullen ook sterven.

Al deze verschillende stappen worden per iteratie in de simulatie herhaald.

Vermits onze simulatie in realtime verloopt, hebben we dit model uitgebreid door verschillende parameters live aanpasbaar te maken. Op deze manier kan de invloed van veranderingen erin eenvoudig worden opgemerkt.

De parameters die per simulatie constant zijn: de ploïditeit, de populatiegrootte, het aantal patches, het aantal additieve loci (dubbel voor diploïde organismen) en de initiële populatieverdeling.

De parameters die realtime kunnen worden aangepast: de kans op een of meer recombinaties, de functie die de omgevingsfactoren bepaalt, de selectiesterkte, de generatieoverlap, de migratiekans, de mutatiekans, de mutatiegrootte (standaardafwijking) en de stapgrootte voor de allelen. [3, 4]

3 Implementatie

We hebben ervoor gekozen zelf een simulatie te maken en niet eenvoudigweg een bestaande tool te gebruiken. Dit wilden we doen, niet enkel om bij te leren, maar om iets te maken dat bruikbaar en nuttig is. Om dit te bereiken zijn er enkele vereisten waaraan ons programma moet voldoen. Bovenal willen we dat onze simulatie gebruiken en begrijpen, eenvoudig is. Daarom hebben we geopteerd om onze simulatie te presenteren als een website. Vervolgens willen we dat onze simulatie interactief is zodat de gebruiker direct de gevolgen kan zien van de parameters die hij aanpast. Om dit te bereiken moet de simulatie snel genoeg zijn.

Omdat we onze visualisatie tonen in een browser, hebben we de keuze over hoe we de simulatie runnen. We kunnen dit doen in de browser zelf of aan de hand van een client-server model. Omdat performantie essentieel is, hebben we gekozen om dit apart op een server te runnen en periodisch data door te sturen. Zo kan de browser zich bezighouden met het plotten van de grafieken en de server met het simuleren. Omdat we willen dat de server data kan versturen als het klaar is met een stap, kunnen we geen http gebruiken. Gelukkig is er een nieuwe technologie, websocket, die ons toelaat aan full-duplex communicatie te doen.

Met deze vereisten in het achterhoofd hebben we besloten de programmeertaal Rust te gebruiken. Rust is een taal die nog in zijn kinderschoenen staat en daarom nog niet bekend is. Toch heeft deze veel features die enorm nuttig zijn voor ons project. Ten eerste is Rust even snel als c of c++, dit laat toe om zonder veel moeite efficiënte code te schrijven. Ten tweede wordt Rust zeer veilig genoemd, dit is omdat het een zeer strikte taal is. Zo laat Rust bijvoorbeeld niet toe om pointers te gebruiken omdat deze mogelijks naar ongeldig geheugen wijzen. Het voordeel aan deze regels is dat de compiler bijna alle bugs kan detecteren op voorhand en veel interne optimalisatie kan uitvoeren. Daarnaast wilden we meer leren over de werking van Rust door dit toe te passen in een reële context.

Een volgend pluspunt aan Rust is dat het mogelijk is om deze taal ook voor de frontend te gebruiken door middel van webassembly. Webassembly is een binair instructieformaat dat beschrijft hoe code veilig kan runnen op eender welke machine. Momenteel wordt het vooral gebruikt in de context browsers. Door deze technologie te gebruiken, bestaat ons project uit dezelfde taal, dit vereenvoudigt de communicatie tussen de front en back-end omdat we dezelfde datastructuren kunnen hanteren. Voor de webpagina zelf maken we gebruik van Seed, dit is een framework gelijkend aan Vue waarbij je dynamisch componenten kan definiëren. Dit laat toe dat elke verandering, zoals een slider verschuiving, direct wordt doorgegeven aan de simulatie.

Een ander probleem dat webassembly oplost, is onze grafieken real-time maken. We willen complexe plots tonen die computationeel zeer duur zijn zoals bijvoorbeeld de fenotypes van de individuen. Dit is omdat we veel datapunten moeten tekenen doorheen de tijd en dit verschillende keren per seconde. Hiervoor zijn wel nog

steeds enkele essentiële optimalisaties nodig geweest, zoals enkel het tonen van een steekproef van de volledige populatie. Deze steekproef wordt elke iteratie random gekozen zodat er toch een compleet beeld geschetst wordt. Bijkomend geven we de optie om oude data niet meer te tonen zodat de grafieken overzichtelijker zijn.

De laatste stap die we hebben genomen om onze simulatie gebruiksvriendelijker te maken is het voorzien van duidelijke foutmeldingen. Zo wordt de gebruiker niet verrast indien hij een foute waarde ingeeft.

4 Analyse vanuit de biologie

Om het biologische aspect te bekijken van genetisch polymorfisme in forward-time simulaties, worden de simulaties vergeleken met cichliden en worden er voorspellingen gedaan in welke omstandigheden de cichliden zouden ontstaan kunnen zijn.

4.1 Cichliden

Cichliden behoren tot de Persiformes (baarsachtigen) en komen voor in het Indische subcontinent, het Midden-Oosten, Zuid- en Centraal-Amerika, Madagascar en Afrika [5]. In Oost-Afrika, waar deze paper op toegespitst wordt, komen ze onder andere voor in de 3 grootste meren van Oost-Afrika, namelijk in Lake Victoria, Lake Tanganyika en Lake Malawi [6], verder komen ze voor in meer dan 30 andere meren in Afrika [7]. Het is een enorm diverse soortengroep die diversiteit vertoont in morfologie, kleur en gedrag. Dit maakt hun modelorganismen voor het bestuderen van speciatie en adaptieve evolutie [5].

Het ontstaan van cichliden wordt gelinkt aan grote geologische gebeurtenissen, maar dit vraagstuk is tot nu toe nog niet opgelost [5]. De eerste fossielen van cichliden dateren van het Eoceen, *Mahengechromis* uit Tanzania wordt gedateerd op een leeftijd van 46 miljoen jaar [8] en *Proterocara* uit Argentinië op een leeftijd van 33.9-55.8 miljoen jaar [9]. Fylogenetische reconstructies wijzen echter uit dat de eerste cichliden tussen 121-165 miljoen jaar ontstaan staan zijn [5].

4.2 Evolutie van Cichliden

Cichliden zijn echter wel geen monofyletische groep [10], iets wat vroeger gedacht werd [5]. Het genetisch materiaal bevat genen van verschillende rivierpopulaties, er zijn dus meerdere invasies en hybridisatiestappen voorgekomen [10]. Dit in tegenstelling met wat er in vroegere papers staat, dat cichliden van biogeografische regio's gesplitst zijn in een chronologische volgorde [5]. Deze ecologische speciatie zal, theoretisch, het meest waarschijnlijk zijn wanneer populaties naar nieuwe adaptieve zones met ondergebruikte niches evolueren. Dit zal dan leiden tot een sterke toename van soorten, iets wat kenmerkend is voor adaptieve radiatie [7]. Het bestaan van ecologische mogelijkheden wordt belangrijk geacht in het proces van adaptieve radiatie omdat de meeste radiatie optreedt na de kolonisatie van nieuwe meren. Maar dat speciatie bij het koloniseren van nieuwe gebieden een snel proces is, bewijst niet dat radiatie aangedreven wordt door het bestaan van ecologische mogelijkheden [7]. De snelle evolutie van cichliden kan verder ook verklaard worden door drift en ecologisch neutrale divergerende seksuele selectie tussen geïsoleerde populaties [11].

Een andere hypothese is dat speciatie ook kan plaatsvinden zonder geografische isolatie. De sensory drive hypothesis stelt dat adaptatie aan lokale omgevingen in sensorische en signaleringssystemen een isolatie kan veroorzaken bij secundair contact van populaties [12, 13]. Met andere woorden: doordat lokale omgevingen

kunnen verschillen van elkaar, zal selectie op signalen en het ontvangen van signalen verschillen tussen deze omgeving en dit leidt dan tot divergentie van de individuen in de omgevingen. Doordat selectie inwerkt op de signalen bij de vorming van paringsbanden, zogenaamde mating signals, zal er divergentie optreden in mating traits of paringskenmerken. Door de divergentie van mating traits zal speciatie versneld kunnen worden [12]. Dit is ook van toepassing op de cichliden waar sensory drive heeft geleid tot de evolutie van kleurpolymorfisme en speciatie. In Lake Victoria bestaat een lichtgradiënt waardoor er een verschillende selectie is op het visueel systeem dat heeft geleid tot de radiatie van verschillende soorten [13].

In praktijk zullen zowel neutrale (niet-adaptieve) als adaptieve processen beide bijdragen aan de soortvorming bij cichliden. Hoe deze processen een invloed hebben op soortvorming is echter wel verschillend en hun relatieve bijdrage is doorheen de tijd veranderd. Neutrale processen lijken cruciaal te zijn voor het bekomen van de genomische variatie in de cichlidenpopulatie, terwijl de adaptieve processen (selectie) die variatie sorteert. Het is de interactie van beide processen die waarschijnlijk nodig geweest is om in een zeer korte tijd de vele diverse soorten te genereren.

4.3 Genetisch polymorfisme via balancerende selectie

Doorheen de speciatie van de cichliden zijn er, zoals eerder besproken, genen van verschillende rivierpopulaties binnengeslopen [10]. Hierdoor zijn er tijdens de speciatie polymorfismen ontstaan, die vervolgens in verschillende soorten differentieel gefixeerd werden. Deze polymorfismen zijn verspreid over verschillende soorten en worden in stand gehouden door balancing selection [14]. De balancing selection hypothese stelt dat vele genen een genetisch polymorfisme bevatten en dat deze varianten in de populatie gehouden worden door positieve selectie van de heterozygoten [15]. Met andere woorden zullen heterozygote individuen een voordeel ondervinden van het heterozygoot zijn en hun genen doorgeven. Homozygoten aan de andere kant zullen hun genen niet of minder kunnen doorgeven omdat het homozygoot zijn een nadeel in overleving is. Doordat de heterozygoten een voordeel ondervinden blijft het gen in de populatie aanwezig in een bepaald evenwicht [15].

Deze balancing selection heeft ook een rol gespeeld in de evolutie van cichliden. Het heeft een belangrijke rol omdat balancing selection ervoor zorgt dat polymorfismen worden behouden en dat er niet een bepaalde morfologie gefixeerd wordt door directionele selectie of genetische drift [16].

4.4 Simulaties

Vanuit de simulaties kan worden afgeleid in welke omgevingsomstandigheden de cichliden mogelijk geëvolueerd zijn. In het model zijn meerdere soorten omgevingsvariëaties ingebouwd die in wat volgt worden overlopen met hun biologische relevantie. De andere parameters staan ingesteld op hun standard instelling. De optimale waarde van de omgeving, in dit model -0.5 en 0.5, kan worden beschouwd als twee niches die kunnen worden ingenomen door een soort cichliden. Een niche is

de specialisatie van een soort in een rol of positie van het ecosysteem. [17]

- Random: Hierbij fluctueert de omgeving in een patch willekeurig waardoor het fenotype niet naar een optimaal fenotype kan evolueren. Er zit geen richting in de evolutie waardoor het fenotype vaak rond een niet optimale waarde blijft.
- Alternating with 50% chance: Bij deze fluctueert de omgeving alternerend tussen twee optimale waarden of niches waarbij beide waarden evenveel kans hebben om als optimale waarde beschouwd te worden en te blijven bestaan. Het fenotype zal hierdoor steeds evolueren naar de twee niches, als er dan een fenotype uitsterft zal het andere fenotype zo evolueren dat ook deze optimale waarde terug wordt ingenomen.
- Alternating with 67% chance: De omgeving fluctueert ook hier tussen twee optimale waarden, hierbij hebben echter beide niches niet evenveel kans om te blijven bestaan. Op deze manier zal de ene niche meer kans hebben om uit te sterven. Dit is dan ook waar te nemen in het model.
- Slow/Medium/Fast sinusoid with patch offset: Omdat de omgeving hier fluctueert in een sinus, zal het fenotype wel evolueren in de richting van de optimale waarden maar het zal stoppen op een niet-optimale waarde. De optimale waarde wordt niet bereikt omdat de omgeving zich voor een te korte tijd op een bepaalde waarde bevindt.
- Random walk: Hierbij fluctueert de omgeving via het principe van random walk waardoor het fenotype niet naar een optimale waarde kan evolueren.
- Constant: Bij dit soort omgeving blijft de omgeving constant waardoor het fenotype evolueert naar de optimale waarde en dan niet verder evolueert. Het is echter al optimaal voor dat soort omgeving, een mutatie zal daardoor niet meer optimaal zijn waardoor de mutatie zal worden weggeselecteerd.
- Constant with jumps: De omgeving zal bij deze voor het grootste deel van de tijd constant zijn met af en toe een verandering naar een andere soort omgeving. Na deze sprongen zal het fenotype zich aanpassen aan de nieuwe optimale conditie.

In het model is ook een functie ingebouwd zodat de initiële populatie kan worden aangepast aan de hand van hoe de loci verdeeld zijn. Dit heeft een invloed vanaf welke waarde van het fenotype wordt gestart, maar op de verdere evolutie heeft dit weinig invloed.

Een andere functie die kan worden aangepast in het model is de populatiegrootte. De populatiegrootte heeft een effect op de snelheid van evolutie. Namelijk hoe groter de populatie, hoe duidelijker de evolutie naar de optimale waarde gebeurt. Dit is ook wat de theorie voorspelt, omdat er in grotere populaties meer mutaties voorkomen dan bij kleinere populaties [18].

Ook kan het aantal patches worden aangepast, een hoger aantal patches leidt tot een stabielere fenotype. Dit is te verklaren omdat als een fenotype in meerdere patches kan voorkomen, er minder kans van uitsterven is van het fenotype. Als een fenotype maar in één patch voorkomt en hier dan ook uitsterft, is het hele fenotype uitgestorven. Terwijl als het fenotype in vier patches voorkomt en het dan in één patch uitsterft, overleeft het fenotype nog in de andere drie fenotypes en kan het van hieruit de uitgestorven patch terug koloniseren.

Verder kan ook het aantal loci aangepast worden. Het aantal plaatsen waar het gen voorkomt kan dus aangepast worden. Dit maakt het model complexer omdat het fenotype bepaald wordt door meerdere genen, die elk apart hun effect uitoefenen op het fenotype. Het effect van elke locus apart kan verschillend zijn en dat dan ook te zien in de grafiek voor de aparte loci.

Als laatste biologisch interessante functie kan er gekozen worden voor haploïde of diploïde organismen. Bij het kiezen voor diploïde organismen wordt er een derde niet optimale fenotype waargenomen dat tussen de twee optimale in ligt. Dit kan verklaard worden doordat er nu drie soorten individuen zijn, namelijk twee soorten homozygote en één soort heterozygote individuen en deze zijn anders aangepast aan hun omgeving. Het derde niet optimale fenotype kan zowel een homozygoot als heterozygoot fenotype zijn, dit is niet te onderscheiden in het model. Als het heterozygote fenotype optimaal is en dus één van de twee homozygote fenotypes niet optimaal, spreekt men van de balancing selection hypothese.

4.5 Toepassing simulaties

Vanuit de simulaties kan worden afgeleid welke omstandigheden het meest gunstig zijn voor soortvorming en in deze toepassing de soortvorming van cichliden. De meest optimale omstandigheid voor soortvorming is een omgeving die vrij constant is, maar af en toe eens fluctueert. Een mooi voorbeeld in het model hiervan is de environment function “constant with jumps”, deze functie is een goede toepassing op de cichliden die in meren leven. Bij de kolonisatie van nieuwe meren verandert de omgeving plotseling om daarna vrij constant te blijven. Deze vrij constante periode kan beschouwd worden als een random walk van de waarde van de omgeving, in deze omgeving is dan ook geen evolutie zoals het model aangeeft.

Er zijn heel wat endemische cichliden soorten die enkel in een bepaald meer voorkomen [10]. Deze soorten hebben dus een laag aantal patches, namelijk enkel dat ene meer. Deze soorten hebben hierdoor dus geen erg stabiel fenotype en kan bij een verandering in het meer sneller uitsterven dan andere soorten die verspreid over meerdere meren voorkomen.

Verder zijn cichliden diploïde organismen waardoor er ook individuen met een niet optimaal aangepast fenotype voorkomen. Deze individuen zullen de populatie ervan weerhouden om te evolueren naar een optimaal fenotype, zodat polymorfismen worden behouden, zoals de balancing selection hypothese stelt [16].

5 Analyse vanuit de wiskunde

Om het wiskundig aspect te bestuderen van genetisch polymorfisme in forward-time simulaties, gaan we kijken naar een aantal belangrijke functies en wat het effect is als we hierin een parameter van waarde laten veranderen. Hierbij baseren we ons op de paper van Prof. Svardal [3].

Hiervoor bekijken we de volgende 2 ongelijkheden die de boven- en ondergrenzen voor de gemiddelde kromming van de locale selectie bepalen.

$$E_T[E_S[\partial^2 s]] < 0 \quad (1)$$

$$\sigma_{crit}^2 = E_T[Var_S[\theta]] + \gamma Var_T[E_S[\theta]] + 2\frac{1-m}{m} Var_S[E_T[\theta]] + C[\theta] > \sigma^2 \quad (2)$$

Bespreking termen:

- Ruimtelijke ($E_S[a]$) en temporele ($E_T[b]$) gemiddelden:

$$E_S[a] = \sum_{i=1}^n c_i a_i \quad (3)$$

$$E_T[b] = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T b_t \quad (4)$$

Hierbij is c_i de draagkracht van jongeren, a_i en b_t willekeurige functies van de omgevingsomstandigheden en T het aantal gesimuleerde tijdstappen.

- Afwijkingen:

$$Var_S[\sigma] = E_S[\sigma^2] - E_S[\sigma]^2 \quad (5)$$

$Var_T[\sigma]$ wordt op gelijke manier gedefinieerd.

- $C[a]$ geeft het effect van de temporele autocorrelatie weer.

$$C[a] = 2(1-\gamma)(1-m) \sum_{\Delta t=1}^{\infty} (1-(1-\gamma)m)^{\Delta t-1} \times (E_S[Cov_T[a_t, a_{t+\Delta t}]] - Cov_T[E_S[a_t], E_S[a_{t+\Delta t}]]) \quad (6)$$

- Selectiesterkte σ

Een eigenschapswaarde x^* is een evolutionaire attractor als aan vergelijking 1 wordt voldaan. Als aan de tweede vergelijking wordt voldaan, zullen er vertakkingen kunnen plaatsvinden.

5.1 Invloed selectiesterkte

We kunnen uit vergelijking 2 duidelijk afleiden dat een kleine σ de vertakking zal bevorderen. Dit betekent dus dat een sterke stabiliserende selectie de vertakking zal aandrijven. Dit zien we dan ook terug in onze simulatie. Als we hier de selectiesterkte aanpassen tijdens een simulatie zien we al heel snel dat er meer of minder vertakkingen zullen plaatsvinden.

5.2 Invloed generatieoverlap

Bij de tweede term aan de linkerkant van vergelijking 2 is γ (generatieoverlap) een belangrijke factor. We zien duidelijk dat als we de generatieoverlap gaan vergroten, de tweede term vergroot en dus ook heel het eerste stuk van deze vergelijking. Dit wil dus zeggen dat een grotere generatieoverlap de vertakkingen zal belemmeren. Ook dit zien we al snel terug in onze simulatie.

5.3 Invloed aantal en grootte patches

Om hier duidelijk de invloed van te onderzoeken, gaan we er even vanuit dat er geen ruimtelijke of tijdelijke correlaties zijn. Dan kunnen we onze vertakkingsvoorwaarde gaan vereenvoudigen tot:

$$\frac{2-m}{m}Var_S[E_T[\theta]] + (1 - \frac{1-\gamma}{n})E_S[Var_T[\theta]] - (1-\gamma)Cov_S[c, Var_T[\theta]] > \sigma^2 \quad (7)$$

In de tweede term van deze vergelijking zien we dat we gaan delen door het aantal patches. Dit wil dus zeggen dat de tweede term verkleint als we meerdere patches hebben. Een toenemend aantal patches zorgt dus voor minder vertakkingen.

De derde term zal nul zijn als de patches even groot zijn of als de omgevingsschommelingen in alle patches even sterk zijn. Deze term toont aan dat een bepaalde hoeveelheid temporele fluctuatie veroorzaakt door relatief sterke schommelingen in kleinere patches, de vertakkingen zullen bevorderen. Als deze worden veroorzaakt in grotere patches zal dit minder bijdragen tot de vertakking.

5.4 Invloed ruimtelijke verschillen en migratiekans

De invloed van ruimtelijke verschillen zien we terug in de derde term van vergelijking 2. Als de omgevingsomstandigheden in alle patches gelijk zijn of als alle nakomelingen zich verspreiden ($m=1$), zal deze term nul zijn. Deze term zal toenemen als de migratiekans afneemt. Dit wil dus zeggen dat er meer vertakkingen zullen plaatsvinden als we een kleine migratiekans hebben.

5.5 Invloed tijdelijke correlatie

Nu kijken we naar de laatste term in vergelijking 2. Deze is nul als er geen tijdcorrelaties zijn. Als deze autocorrelaties lokaal zijn maar niet wereldwijd, zal vertakking

worden bevorderd door positieve autocorrelatie en belemmerd door negatieve autocorrelatie. Als deze autocorrelatie wereldwijd is, zal dit geen effect hebben op onze vertakking.

5.6 Invloed populatiegrootte

De populatiegrootte heeft een effect op de selectiesterkte. Voor kleine populaties, zal σ ook kleiner zijn, waardoor vertakking makkelijker wordt. Heel kleine populaties kunnen er dus ook voor zorgen dat onze resultaten onnauwkeurig worden.

6 Conclusie

We kunnen besluiten dat forward-time simulaties niet erg geschikt zijn om reeds bestaande genetische polymorfismen te verklaren, maar wel zeer interessant zijn om inzicht te krijgen in de natuurlijke processen achter het fenomeen. Mechanismen zoals het heterozygootvoordeel, bet-hedging en de invloed van spatiale en temporele variatie in omgevingsfactoren kunnen met zeer concrete voorbeelden verklaard en geïllustreerd worden.

Balancerende selectie kan veroorzaakt worden door nog vele andere (vaak soort-specifieke) factoren, die nooit allemaal gevat kunnen worden in één model. Bij het bestuderen van het model en ontwikkelen van onze simulatie kwamen dan ook meerdere ideeën voor mogelijk uitbreidingen naar boven, zoals het opheffen van de assumptie van een constante populatiegrootte of het toevoegen van controleerbare migratie. Door hiermee te experimenteren merkten we echter dat zulke uitbreidingen, hoewel interessant op zich, de bruikbaarheid en interpreteerbaarheid van de simulatie sterk beïnvloedden.

Vanwege de nadruk op heldere wetenschappelijke communicatie in dit vak, hebben we ons daarom gefocust op het ontwikkelen van een intuïtieve en praktisch nuttige applicatie die, door middel van real time parameters en visualisaties, onszelf en anderen, die minderen gekend zijn met het vakgebied, inzicht kan geven in de basisprincipes.

Vanuit biologisch standpunt kan dit model, met betrekking tot de use case van de cichliden, beamen dat de soortenrijkdom van de cichliden door een combinatie van verschillende omstandigheden is kunnen ontstaan. Veranderingen in de omgeving kunnen hier hun steentje hebben bijgedragen. Verder kan worden afgeleid dat vele soorten geen zeer stabiele soorten zijn omdat een deel van hen enkel voorkomt in één bepaald meer wat hen vatbaarder maakt voor existentie. Er kan ook besloten worden dat de balancing selection hypothese een rol kan hebben gespeeld in de soortvorming van de cichliden, doordat er verschillende polymorfismen optreden met elk een ander fenotype dat al dan niet optimaal is voor de omgeving.

Vanuit wiskundig standpunt kan er uit dit model afgeleid worden dat er veel parameters zijn die een effect zullen hebben op de vertakking. Bij een ideaal model laten we dus geen generatieoverlap toe, zullen er een groot aantal patches zijn, de omgevingsschommelingen zullen vooral plaatsvinden in kleine patches en de migratiekans zal klein zijn. Hierbij zullen er dus weinig vertakkingen plaatsvinden.

Opmerkelijk voor ons was dat de resultaten uit de biologische en wiskundige analyse mooi met elkaar stroken. Het was leerrijk om de voorkennis uit onze verschillende achtergronden te combineren, om ons te verdiepen in onderwerpen waarnaar actief wetenschappelijk onderzoek wordt gedaan. We hopen dat onze simulatie ook echt voor anderen van pas kan komen om inzicht te verwerven in het vakgebied.

7 Taakverdeling

De simulatie werd geïmplementeerd door Thomas Dooms en Ward Gauderis (3de bachelor Informatica), waarbij Ward zich focuste op de simulatie zelf en Thomas op de visualisatie en interactie. De analyse vanuit de biologie werd uitgevoerd door Lauren Van Nimmen (3de bachelor Biologie). De wiskundige analyse werd gerealiseerd door Kato Starckx (2de bachelor Wiskunde).

De hoofdstukken in dit verslag werden door de corresponderende studenten geschreven.

8 Referenties

- [1] Michel Veuille. Genetic polymorphism and variation, October 2016.
- [2] Theresa Phillips. Genetic Polymorphism and How It Lasts Over Generations.
- [3] Hannes Svoldal, Claus Rueffler, and Joachim Hermisson. A general condition for adaptive genetic polymorphism in temporally and spatially heterogeneous environments. *Theoretical population biology*, 99, 11 2014.
- [4] Hannes Svoldal, Claus Rueffler, and Joachim Hermisson. Comparing environmental and genetic variance as adaptive response to fluctuating selection. *Evolution*, 65(9):2492–2513, 2011.
- [5] Martin Genner, Ole Seehausen, Dave Lunt, Domino Joyce, Paul Shaw, Gary Carvalho, and George Turner. Age of cichlids: New dates for ancient lake fish radiations. *Molecular biology and evolution*, 24:1269–82, 06 2007.
- [6] Axel Meyer. Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology & Evolution*, 8(8):279–284, August 1993.
- [7] Ole Seehausen. African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597):1987–1998, August 2006.
- [8] Alison M. Murray. The oldest fossil cichlids (Teleostei: Perciformes): indication of a 45 million-year-old species flock. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1468):679–684, April 2001.
- [9] Maria C. Malabarba, Oscar Zuleta, and Cecilia Del Papa. *Proterocara argentina*, a new fossil cichlid from the Lumbrera Formation, Eocene of Argentina. *Journal of Vertebrate Paleontology*, 26(2):267–275, June 2006.
- [10] Domino A. Joyce, David H. Lunt, Martin J. Genner, George F. Turner, Roger Bills, and Ole Seehausen. Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology*, 21(3):R108–R109, February 2011.
- [11] A.J. Ribbink, B.A. Marsh, A.C. Marsh, A.C. Ribbink, and B.J. Sharp. A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi. *South African Journal of Zoology*, 18(3):149–310, January 1983.
- [12] Janette Wenrick Boughman. How sensory drive can promote speciation. *Trends in Ecology & Evolution*, 17(12):571–577, 2002.
- [13] Ole Seehausen, Yohey Terai, Isabel S. Magalhaes, Karen L. Carleton, Hillary D. J. Mrosso, Ryutaro Miyagi, Inke van der Sluijs, Maria V. Schneider, Martine E. Maan, Hidenori Tachida, Hiroo Imai, and Norihiro Okada. Speciation

through sensory drive in cichlid fish. *Nature*, 455(7213):620–626, October 2008.

- [14] S. Nagl, H. Tichy, W. E. Mayer, N. Takahata, and J. Klein. Persistence of neutral polymorphisms in Lake Victoria cichlid fish. *Proceedings of the National Academy of Sciences*, 95(24):14238–14243, November 1998.
- [15] Philip W. Hedrick. Balancing selection. *Current Biology*, 17(7):R230–R231, April 2007.
- [16] Alexandra M. Tyers, Gavan M. Cooke, and George F. Turner. Rare morph Lake Malawi mbuna cichlids benefit from reduced aggression from con- and hetero-specifics. preprint, *Evolutionary Biology*, April 2021.
- [17] R. H. Whittaker, S. A. Levin, and R. B. Root. Niche, Habitat, and Ecotope. *The American Naturalist*, 107(955):321–338, May 1973.
- [18] Toni Gossmann, Peter Keightley, and Adam Eyre-Walker. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome biology and evolution*, 4:658–67, 03 2012.