Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Krzysztof Gogolewski
Nr albumu: 291538

Modelowanie proliferacji transpozonów indukowanej stresem środowiskowym

Praca licencjacka
na kierunku MATEMATYKA

Praca wykonana pod kierunkiem
dr hab. Anny Gambin

Wrzesie? 2014

Oświadczenie kierującego pracą

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data Podpis kierującego pracą

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data Podpis autora (autorów) pracy

Streszczenie

Niniejsza praca skupia się na analizie metod modelowania proliferacji transpozonów w populacjach aseksualnych. Po krótkim wprowadzeniu niezbędnych pojęć z pogranicza biologii molekularnej i genetyki, następuje opis modeli obliczeniowych i towarzyszących im metod formalizacji w postaci modeli analitycznych, pochodzących z dwóch różnych etapów rozwoju genetyki populacji. W pierwszej części modele zawarte w [Cha83] zostają poddane dogłębnej analizie i formalizacji, łącznie z przytoczeniem narzędzi i teorii genetyki populacji niezbędnych do uzasadnienia poprawności poszczególnych etapów modelowania. Następnie, przytoczony zostaje opis stochastycznego modelu obliczeniowego [Sta13], w którym stres środowiskowy został uwzględniony, jako jeden z potencjalnych czynników mogących mieć wpływ na aktywność transpozonów. Na zakończenie zaprezentowany jest jeden z możliwych sposobów formalizacji tego modelu, opierający się na konstrukcji operatora opisującego zmiany w populacji między dwoma kolejnymi pokoleniami [Sta14].

Słowa kluczowe

genetyka populacji, transpozony, modelowanie matematyczne, stres środowiskowy

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.1 Matematyka

Klasyfikacja tematyczna

92-xx Biology and other natural sciences
92Dxx Genetics and population dynamics
92D25 Population dynamics (general)

Tytuł pracy w języku angielskim

Modeling stress-induced transposons' proliferation

Spis treści

Rozdział 1

Wprowadzenie

Dokonując przeglądu współczesnej biologii obliczeniowej, biologii teoretycznej czy biomatematyki można zaobserwować istnienie wspólnej części tych trzech dziedzin, a mianowicie dynamicznie rozwijającej się dziedziny, jaką stanowi genetyka populacji. W obrębie tej dziedziny naukowcy starają się wyjaśniać zachowania różnych populacji pod wpływem działania mechanizmów ewolucyjnych. Przedmiotem zainteresowania jest dziedziczność pewnych cech i własności oraz sposób w jaki wspomniane mechanizmy ewolucyjne, takie jak selekcja naturalna, dryft genetyczny, czy mutacje, kontrolują przebieg samej ewolucji.

W niniejszej pracy przedstawiono i przedyskutowano dwa rodzaje modeli proliferacji transpozonów, sekwencji DNA zdolnych do przemieszczania się w obrębie genomu pojedynczej komórki, w populacjach aseksualnych, składające się zarówno z symulacyjnych modeli obliczeniowych jak i ich analitycznych formalizacji.

Pierwszy rozdział stanowi wprowadzenie do tematyki od strony biologicznej, przywołane zostały fundamentalne pojęcia z zakresu biologii molekularnej oraz genetyki, pozwalające na dalszą dyskusję nad proponowanymi modelami.

Zasadnicza część pracy bazuje na artykule [Cha83], opracowanym przez biologów Briana i Deborah Charlesworth, który opisuje aktywność transpozonów w kontekście regulowanej transpozycji oraz selekcji naturalnej. Wyniki w nim zawarte są cytowane także dziś, dlatego, po wprowadzeniu opisu modelu obliczeniowego, przeprowadzono szczegółową analizę kolejnych kroków konstrukcji modeli analitycznych z jednoczesnym przybliżeniem użytych narzędzi i teorii, głównie z zakresu genetyki populacji, niezbędnych do uzasadnienia ich poprawności.

Następnie, przedstawiona została alternatywna propozycja opisana w [Sta13] której autorzy, we wprowadzonym stochastycznym modelu obliczeniowym, proponują uwzględnianie stresu środowiskowego, jako dodatkowego czynnika mającego wpływ na aktywność transpozonów. Zaprezentowany został także szkic jednej z możliwych metod formalizacji modelu, poprzez wyprowadzenie operatora modelującego cykl życia populacji [Sta14]. Zakończenie pracy tworzy krótkie podsumowanie przedstawionych metod wraz z ich porównaniem.

Rozdział 2

Tło biologiczne

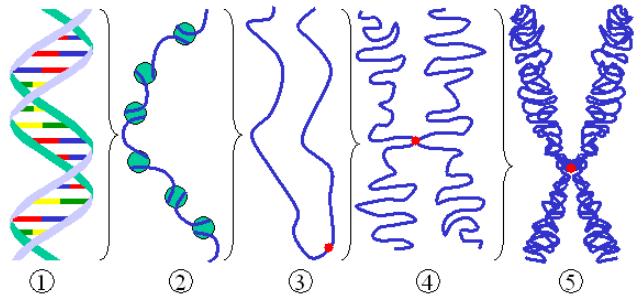
Przejście do zasadniczej części pracy wymaga przygotowania pewnego słownika podstawowych pojęć biologicznych, które pozwolą zrozumieć omawiane zagadnienia i prezentowane modele zarówno obliczeniowe jak i formalne. Następnie, po określeniu czym jest transpozon, zostanie przedstawiona motywacja jaka stoi za ich badaniem oraz po krótce opisany dotychczasowy stan wiedzy na ich temat.

2.1. Podstawowe pojęcia

Celem rozpoczęcia opisu modeli proliferacji transpozonów, niezbędne jest przytoczenie kilku definicji z zakresu biologii molekularnej i genetyki. Szczegółowość poniższych opisów została utrzymana na poziomie pozwalającym zrozumieć dalszą część pracy, zaś część z nich dodatkowo uzupełniono o przykłady ilustrujące konkretną definicję w kontekście genetyki człowieka. Chcąc odnaleźć dokładniejsze opisy struktur czy procesów molekularnych czytelnik może sięgnąć do przykładowej literatury wymienionej w bibliografii [Rid99, Weg12, Hig08].

DNA Kwas dezoksyrybonukleinowy, *DNA*, jest liniowym związkiem chemicznym, składającym się z dwóch nici nukleotydów. W pewnym uproszczeniu szkielet pojedynczej nici składa się z reszt deoksyrybozy połączonych przez grupy fosforanowe w liniowy łańcuch. Reszta deoksyrybozy składa się z pięciu atomów węgla, którym standardowo przyporządkowuje się numery od 1 do 5. Grupy fosforanowe łączą się z deoksyrybozą poprzez atom węgla 3 i 5, natomiast atom 1. jest miejscem wiązania z jedną z czterech *zasad azotowych*: Adenina, Cytozyna, Guanina i Tymina. Powstały związek reszty fosforanowej, deoksyrybozy i jendej zasady azotowej nazywamy nukleotydem. Pełny łańcuch DNA współtworzą najczęściej dwie ułożone antyrównolegle nici nukleotydów, powiązane wiązaniami wodorowymi pomiędzy parami zasad: Adenina - Tymina, Cytozyna - Guanina, formujące się w podwójną helisę. To właśnie sekwencje wspomnianych zasad azotowych opisują najczęściej materiał genetyczny organizmu (wyjątek stanowią, np wirusy).

Genom i chromosomy Wśród organizmów eukariontycznych całość dziedzicznej informacji genetycznej - *genom*, składa się z pojedynczego, haploidalnego zestawu chromosomów (np. zestaw 23 chromosomów u człowieka). Każdy *chromosom*, zbudowany jest z dwóch siostrznych chromatyd, połączonych w jednym punkcie nazywanym centromerem, które są zbudowane z ciągłego łańcucha DNA niejako „ponawijanego" na białka histonowe - histony (patrz rys. 2.1).



Rysunek 2.1: Budowa chromosomu. Kolejne struktury to: (1) helisa DNA, (2) łańcuch DNA nawinięty na histony, (3) chromatyda, (4) siostrzane chromatydy połączone w centromerze, (5) chromosom. źródło: [dostęp online] `https://commons.wikimedia.org/wiki/File:Chromatin_chromosom.png` (z dnia 12.04.2014r.)

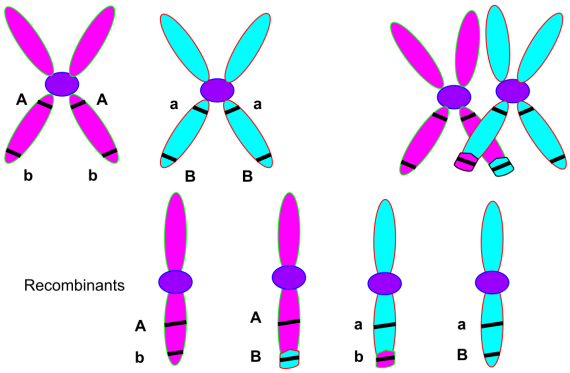
Gen i allele W obrębie każdego chromosomu występują zarówno sekwencje kodujące - geny, jak i sekwencje niekodujące. *Gen*, jest podstawową jednostką dziedziczenia, która jest opisana odcinkiem DNA określonej długości i znajduje się na określonej pozycji (*locus*, l.mn. *loci*) w chromosomie. Odcinek ten koduje jeden rodzaj białka (bądź RNA) wpływający na fenotyp organizmu. Jedną z wersji genu w określonym *locus* na chromosomie nazywamy *allelem*. Allele tego samego genu różni zazwyczaj kilka nukleotydów, przez co w przypadku organizmów diploidalnych mogą być one dominujące bądź recesywne.

Gamety i zygota *Gametą* nazywamy haploidalną komórkę rozrodczą powstałą na drodze mejozy, zawierającą całą informację genetyczną organizmu z którego pochodzi. W przypadku człowieka, prawidłowa gameta, składa się z 22 autosomalnych chromosomów i jednego chromosomu płci. Podczas reprodukcji, dwie gamety pochodzące od matki i ojca, łączą się tworząc jedną diploidalną komórkę - *zygotę*.

Crossing-over i linkage equilibrium Podczas podziału komórkowego, w wyniku którego powstają gamety, na etapie zwanym profazą I dochodzi do wymiany materiału genetycznego pomiędzy chromosomami homologicznymi (patrz rys. 2.2). Zjawisko to określane jako *crossing-over* w uproszczeniu polega na wymianie ciągłych odcinków pomiędzy chromosomami, dzięki czemu w populacji zachowana jest różnorodność genetyczna. Jeśli miejsce w którym dochodzi do wymiany materiału jest tak samo prawdopodobne dla każdej pozycji na chromosomie ma miejsce tzw. *linkage equilibrium*. Przypadek, w którym pewne geny są ze sobą mocno sprzężone i szansa na to, że nastąpi wymiana odcinków zawierających tylko jeden z nich jest niewielka jest określany jako *linkage disequilibrium*.

Tranpozon Ostatecznie może zostać zdefiniowane główne pojęcie biologiczne, będące tematem tej pracy. *Transpozon* jest sekwencją DNA mającą zdolność przemieszczania się w obrębie genomu, czyli w obrębie jednego chromosomu bądź też pomiędzy nimi.

W dalszej częsci uwaga skupiona zostanie na dokładniejszym poznaniu transpozonów, w szczególności zostanie dokonany ich podział ze względu na mechanizm odpowiadający za ich mobilność oraz umiejętność uruchomienia odpowiedniego mechanizmu. Rozdział zamknie krótka motywacja jaka stoi za badaniem transpozonów, która ma na celu przekonanie czytelnika o istotności omawiania głównego modelu rozważanego w kolejnym rozdziale.



Rysunek 2.2: produkcja gamet z pojedynczym crossing-over, w wyniku którego dwie gamety dokonały wymiany alleli jednego z genów ($B \leftrightarrow b$) źródło: [dostęp online] `http://commons.wikimedia.org/wiki/File:Chromosomal_Crossover.svg` (z dnia 12.04.2014r.)

2.2. Transpozony

Jak zostało ustalone, transpozonem nazwiemy fragment DNA, który potrafi zmieniać swoje położenie w obrębie genomu. Pierwszym znaczącym momentem w badaniach nad transpozonami było odkrycie dokonane przez Barbarę McClintock, która prowadząc prace nad genomem kukurydzy zaobserwowała, że mutacje, insercje i transpozycje spowodowane przez ruchome fragmenty DNA powodują zmiany koloru ziaren kukurydzy. Za to odkrycie McClintock została w 1983 roku uhonorowana Nagrodą Nobla w dziedzinie medycyny lub fizjologii i pokazała zupełnie nowe spojrzenie na ówczesną genetykę.

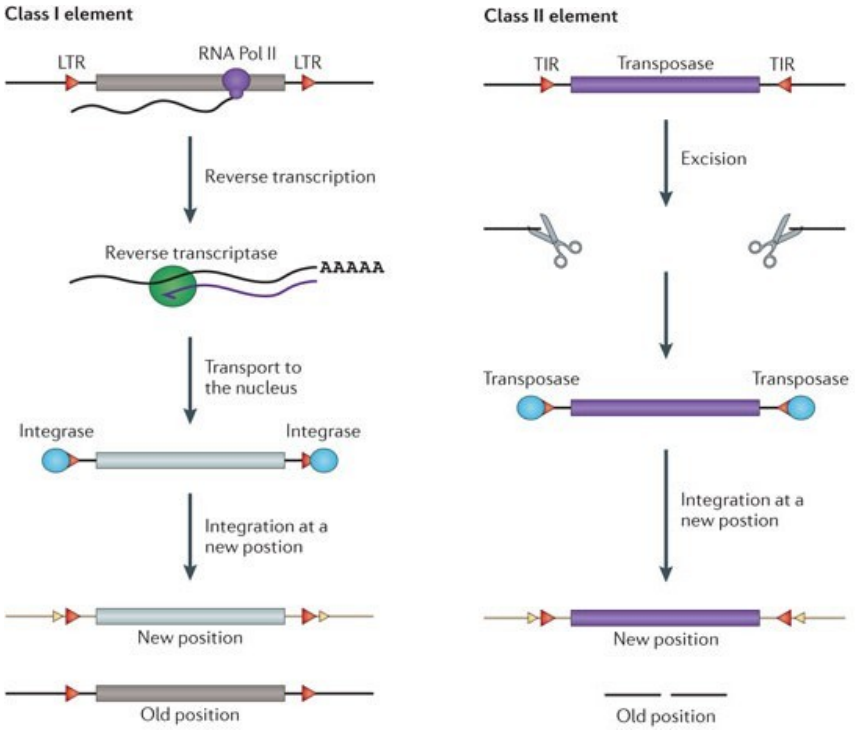
Od tamtego czasu badacze zdołali dokładniej poznać całe rodziny transpozonów i ich naturę oraz na tej podstawie sklasyfikować je ze względu na mechanizm odpowiedzialny za transpozycję [Wic07]. Poniżej znajduje się opis mechanizmów oraz ich schematyczna reprezentacja graficzna (patrz rys. 2.3).

Retrotranspozony Uznawane za tranpozony klasy pierwszej, retrotranspozony, dokonują transpozycji metodą *copy-paste*, składającej się z dwóch etapów. W pierwszej fazie następuje transkrypcja odcinka DNA kodującego transpozon na łańcuch RNA, następnie w obecności enzymu odwrotnej transkryptazy RNA jest ponownie przepisywane na DNA i integrowane w ustalonym miejscu genomu.

Transpozony DNA Transpozony klasy drugiej przemieszczają się metodą *cut-paste*. Opierają swój mechanizm transpozycji na obecności enzymów transpozazy, które odpowiadają za wycięcie fragmentu DNA kodującego dany transpozon i przeniesienie go w nowe miejsce w obrębie genomu.

Poza podziałem ze względu na rodzaj mechanizmu transpozycji, można podzielić transpozony ze względu na ich zdolność do samodzielnego przeprowadzenia procesu transpozycji [Pra09].

Transpozony autonomiczne Jest to grupa tych transpozonów (składająca się z przedstawicieli zarówno klasy pierwszej jak i drugiej), które są zdolne do samodzielnego przemieszczenia w obrębie genomu. Wynika to z faktu zakodowania odwrotnej transkryptazy, w przypadku retrotranspozonów czy ezymu transpozazy, w przypadku transpozonów DNA w obrębie odcinka kodującego dany transpozon.



Rysunek 2.3: Mechanizmy transpozycji, źródło: Damon Lisch (2013) *How important are transposons for plant evolution?* Nature Reviews Genetics 14, 49-61

Transpozony nieautonomiczne Do tej grupy zaliczane są wszystkie te transpozony, które nie są autonomiczne, a zatem w celu dokonania transpozycji wymagają obecności odpowiednich enzymów.

2.3. Motywacja

Przy rozważaniu motywacji należy podkreślić jeszcze jeden istotny fakt, wynikający z natury transpozonów, acz mający miejsce również wtedy gdy nie są one rozważane. Struktura genomu podlega ciągłym zmianom wynikającym z translokacji, mutacji czy delecji [Hua12]. Ściślej mówiąc, transpozycje nie tylko zmieniają strukturę genomu w kontekście kolejności par zasad, ale są także w stanie, np. dezaktywować bądź aktywować produkcje pewnego białka w wyniku zmian dokonanych w obrębie odcinka kodującego [Kaz04, Kid97]. Ponadto występowują punktowe delecje czy mutacje zasad na genomie, nie związane z transpozonami, które także mogą rzutować na wiele czynników, np. wspomnianą aktywność genów [Hal90].

W literaturze można odnaleźć wiele hipotez na temat potencjalnej roli transpozonów oraz ich genezy. Część z nich traktuje transpozony jako pasożyty, żerujące na genomie gospodarza, bez szczególnej roli [Hic82]. Istnieją jednak także takie, które sugerują istnienie relacji pomiędzy stresem środowiskowym, a aktywnoscią transpozonów [Cap00, Wes96]. Brak wyraźnie określonego związku przyczynowo-skutkowego jest głównym powodem rozważania wielu możliwych hipotez, np. Czy wybuchy aktywności transpozonów są odpowiedzią organizmu na nagłe zmiany środowiskowe? Próba odpowiedzi na to pytanie pojawiła się w literaturze jako stochastyczny model obliczeniowy. Z przeprowadzonych symulacji wynika, że w istotnie zmieniających się warunkach środowiskowych transpozony potrafią zwiększać swoją aktywność celem ułatwienia organizmowi gospodarza adaptacji do zachodzących zmian [Sta13].

Inny kontekst badań polega na obserwowaniu czy asymptotycznie liczba transpozonów na genomie utrzymuje się na pewnym nietrywialnym poziomie (ang. *transposition-selection equilibrium*). Jedna z prób rozważania tego problemu, poprzez modelowanie obliczeniowe jak i analityczne, przedstawiona została w [Cha83]. Wyniki pracy autorów wskazują na istnienie takiego stanu, jednak przyjęte założenia mogą być podstawą do podważania ich znaczenia w kontekście czysto biologicznym i zostaną poddane dyskusji w niniejszej pracy.

Wydaje się zatem, że analiza proliferacji transpozonów może przynieść odpowiedź na wiele istotnych pytań, a zatem zasadne jest zarówno tworzenie modeli obliczeniowych, jak i poszukiwanie jak najlepszych modeli formalnych, które ją opisują. Pierwszym naturalnym krokiem w tym kierunku może byc próba formalizacji wspomnianego modelu obliczeniowego, którego opis przedstawiony jest w następnym rozdziale.

Rozdział 3

Obliczeniowe modele proliferacji tranpozonów

Poniżej znajduje się opis dwóch modeli symulujących proliferację transpozonów. Pierwszy z nich opisany w [Cha83], dla którego autorzy przedstawili kilka modeli formalnych, których przegląd nastąpi w dalszej części pracy. Drugi model [Sta13] jest niejako rozszerzeniem swojego poprzednika, a jedną z głównych różnic jest rozważanie stresu środowiskowego jako czynnika wpływającego na proliferację transpozonów.

3.1. Model B. Charlesworth, D. Charlesworth

Autorzy artykułu, na którym oparta jest niniejsza praca, w części poświęconej modelowi obliczeniowemu nie nadają oznaczeń dla konkretnych parametrów swojego modelu, a w to miejsce podają tylko przykładową wartość, której najczęściej używali. Celem poprawienia czytelności wprowadzone zostaną własne oznaczenia na poszczególne parametry opisywanego modelu. Autorzy rozważają model populacji rozmnażającej się w sposób płciowy, w której organizm jest reprezentowany przez dwie pary homologicznych chromosomów (seksualna populacja diploidalna), zawierających T *loci*, na których może pojawić się transpozon (tj. $\frac{T}{2}$ *loci* na jednym zestawie chromosomów). Początkowa populacja składa się z N organizmów.

Inicjalizacja populacji Dla każdego organizmu, w każdym jego *locus* z prawdopodobieństwem będącym stosunkiem oczekiwanej liczby transpozonów na organizmie t do liczby wszystkich dostępnych *loci* T umieszczono transpozon. Tak skonstruowana populacja, była punktem startowym dla przeprowadzonej symulacji.

3.1.1. Cykl życia modelu

Przyjęto założenie, że początkowa populacja składa się z dorosłych organizmów zdolnych do wydania potomstwa. Stąd cykl życia w populacji rozpoczyna się od etapu reprodukcji, która przebiega w następujący sposób.

Wydanie potomstwa i selekcja

Pierwszy krok w cyklu polega na wygenerowaniu nowego pokolenia. W tym celu wybieramy dwa dowolne organizmy, a następnie z każdego generujemy jedną gametę. Autorzy zakładają, że produkcja gamety odbywa się w zgodzie ze ściśle określonym przebiegiem mechanizmu *crossing-over*.

Crossing-over Liczba miejsc, w których dochodzi do zajścia zdarzenia crossing-over jest losowana z rozkładu Poissona z parametrem λ, który autorzy opisują jako iloczyn liczby możliwych miejsc wystąpienia rekombinacji $T-1$ i częstości jej występowania między dwoma sąsiadującymi *loci* ξ. Miejsca występowania zdarzeń są losowane jednostajnie ze wszystkich możliwych (pomiędzy sąsiednimi *loci*) w obrębie chromosomu. Po wybraniu miejsc zajścia crossing-over, generowane są dwa nowe chromosomy i losowo wybierany ten, który będzie tworzył gametę.

Po wygenerowaniu przez rodziców dwóch gamet są one łączone celem stworzenia potomnej zygoty. Decyzja o tym czy organizm będzie należał do następnej generacji zostaje podjęta na etapie selekcji.

Selekcja Opis kolejnego kroku symulacji wymaga przytoczenia poniższej definicji [Orr09]

Definicja: (Dopasowanie (ang. *fitness*)) *W teorii ewolucji, dopasowanie jest abstrakcyjną miarą przystosowania organizmu do życia w ustalonych warunkach środowiskowych i może być traktowane jako opis zdolności zarówno do przeżycia, jak i reprodukcji.*

Autorzy przyjmują, że dopasowanie nowo powstałego organizmu jest funkcją liczby transpozonów w genomie n, zadaną wzorem:

$$\omega(n) = 1 - s \cdot n^t$$

gdzie s, t są parametrami modelu. Rozważane wartości, które zostały zaprezentowane należały odpowiednio do przedziału $(0, 0.0015)$ oraz $(1.2, 2)$. Zygota przechodzi do następnej generacji z prawdopodobieństwem $\omega(n)$. Procedurę generowania zygot powtarzano, do momentu, aż dokładnie N z nich przejdzie proces selekcji.

Transpozycje i delecje

Po wygenerowaniu nowego pokolenia, następuje realizacja transpozycji (typu *copy-paste*) oraz delecji transpozonów we wszystkich organizmach. Prawdopodobieństwo wystąpienia transpozycji danego elementu jest malejącą funkcją $u(n)$ liczby transpozonów na całym genomie, zaś delecji stałą wartością v. Liczba zdarzeń transpozycji i delecji w obrębie całego genomu jest zmienną losową o rozkładzie Poissona z parametrem odpowiednio $n \cdot u(n)$ oraz $n \cdot v$. *Locus*, na którym nastąpiła zarówno delecja, jak i transpozycja wybierany jest w sposób jednostajny z *loci* zajętych i wolnych odpowiednio. Warto podkreślić, że autorzy poszukując wolnego miejsca posługiwali się dość nieefektywną metodą: wybierz dowolny *locus* na genomie, jeśli jest wolny wstaw transpozon, w przeciwnym wypadku powtórz losowanie.

Po zajściu transpozycji i delecji na wszystkich organizmach, zygoty uznane zostają za zdolne do reprodukcji i naturalnie cykl życia trafia ponownie do początkowego stanu.

Już w tym miejscu czytelnik może zwrócić uwagę na wątpliwe założenia pojawiające się w powyższym modelu. Autorzy chcąc udowodnić istnienie stanu równowagi ze względu na liczbę transpozonów w organizmie, poczynili bardzo silne ograniczenia dotyczące proliferacji transpozonów, jednak ich dokładniejsza analiza i potencjalne wady przeanalizowane zostaną w dalszej części pracy. W szczególności widzimy, że funkcja dopasowania zależy explicite od liczby transpozonów w genomie, czyli nadmierna proliferacja jest karana i tłumiona tym założeniem.

3.2. Model Startek, Le Rouzic i in.

Autorzy kolejnego modelu obliczeniowego [Sta13] proponują odejście od bezpośredniego związku liczby transpozonów z dopasowaniem organizmu do życia. W to miejsce wprowadzone zostaje pojęcie fenotypu reprezentowane jako wektor liczby rzeczywistych, który zmienia się pod wpływem mutacji zachodzących w obrębie genomu. Fenotyp stanowi zatem pewną abstrakcyjną miarę opisującą organizm w kontekście dopasowania do warunków środowiskowych w których żyje. Ponadto zakłada się, że w środowisku, w którym żyje cała populacja istnieje tzw. optymalny fenotyp $\hat{\pi} \in \mathbb{R}^n$, czyli potencjalnie najkorzystniejsza (ze względu na naturalną selekcję) wartość fenotypu przy zadanych warunkach środowiskowych. Dzięki takiemu podejściu aktywność transpozonów, a co ważne ich liczba, wpływa na dopasowanie tylko pośrednio. Pełna symulacja stochastyczna oparta jest na następującym cyklu życia populacji (patrz rys. 3.1 (A)).

3.2.1. Cykl życia modelu

Początkowa populacja

Cykl życia rozpoczyna się od populacji rozmiaru m, rozmnażającej się przez klonowanie. Każdy z organizmów jest reprezentowany przez parę $o = (\pi, t) \in \mathbb{R}^n \times \mathbb{N}$, gdzie π jest wektorem opisującym fenotyp organizmu, zaś t liczbą tranpsozonów na genomie. Na początku pierwszego cyklu każdy z organizmów ma fenotyp równy optymalnemu $\pi = \hat{\pi}$ oraz posiada jeden transpozon autonomiczny.

Proliferacja transpozonów

W tak skonstruowanej populacji pierwszy etap stanowi proliferacja tranpozonów. Z odpowiednimi częstościami zachodzą zdarzenia transpozycji i delecji, a ponadto transpozony mogą utracić swoją autonomiczność, tj. zdolność produkcji mechanizmu odpowiedzialnego za przeprowadzenie transpozycji, zgodnie z rys. 3.1 (B)

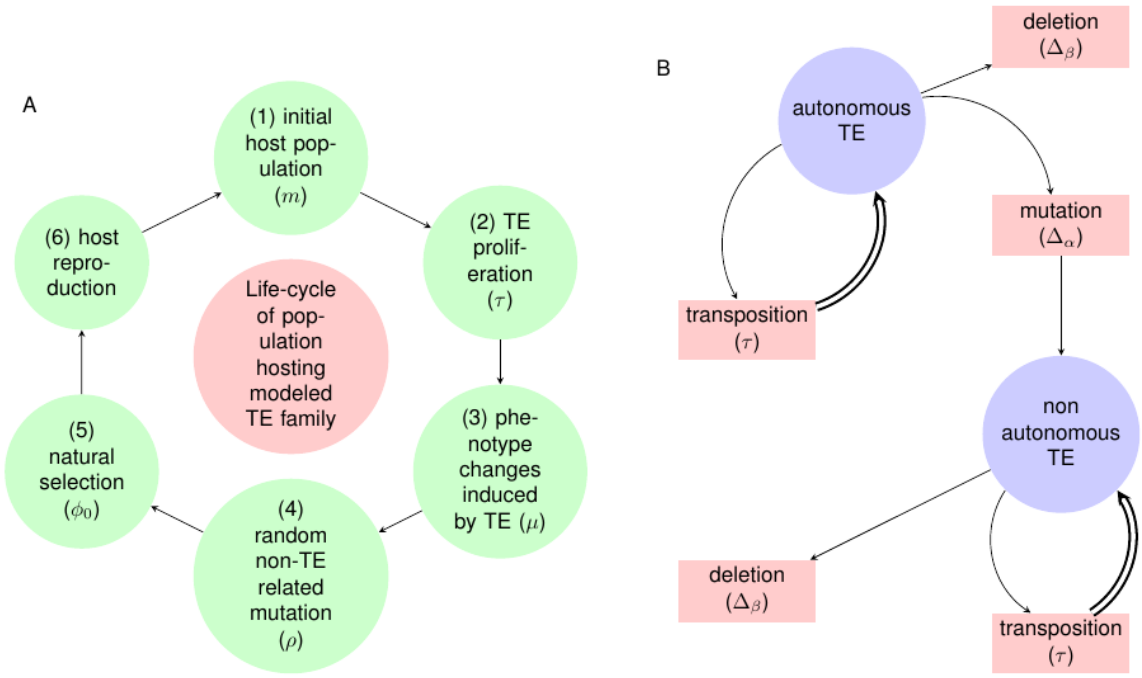
Zmiany wartości fenotypu

Każde zdarzenie translokacji czy delecji implikuje pewną zmianę w sekwencji genomu, przez co wnosi pewną zmianę do fenotypu. Dla każdego zdarzenia zmiana ta jest losowana ze scentrowanego rozkładu normalnego $\mathcal{N}(0, \mu)$ (gdzie μ jest parametrem modelu) i wpływa na jedną współrzędną wektora fenotypu.
Dodatkowo na genomie dochodzi do punktowych mutacji nie związanych z transpozonami. Każda mutacja modyfikuje jedną współrzędną wektora opisującego fenotyp o wartość próbkowaną z rozkładu $\mathcal{N}(0, \rho)$ (gdzie ρ jest parametrem modelu).

Selekcja i reprodukcja

Po wszelkich zmianach fenotypu następuje etap naturalnej selekcji zgodnie z funkcją dopasowania
$F : \mathbb{R}^n \times \mathbb{N} \to \mathbb{R}$ zadaną jako:

$$F(o) = \exp\left(-\|\pi(o) - \hat{\pi}\|\right) = \exp\left(-\sum_{i=1}^{n} \left(\pi_i(o) - \hat{\pi}_i\right)^2\right)$$



Rysunek 3.1: (A) Cykl życia populacji; (B) Proliferacja transpozonów na genomie hosta; Model obliczeniowy Startek, Le Rouzic et al.; źródło: [Sta13]

gdzie $\pi(o)$ oznacza fenotyp organizmu o. Jak zostało wcześniej wspomniane, funkcja dopasowania nie zależy od ilości transpozonów znajdujących się na genomie gospodarza, dzięki czemu nie wymusza w sztuczny sposób *transposition-selection equilibrium*.

Organizmy, które przetrwały etap naturalnej selekcji wydają potomstwo. Liczba potomków organizmu o jest losowana z rozkładu Poissona z paramterem proporcjonalnym do jego dopasowania $F(o)$. Współczynnik proporcjonalności jest ustalany dla każdej generacji tak, aby jej oczekiwany rozmiar wynosił m.

Stres środowiskowy

W symulacji wprowadzony jest także stres środowiskowy modelowany, jako zmiana optymalnego fenotypu w każdej generacji. Innymi słowy optymalny fenotyp traktować należy jako funkcję dyskretnego czasu $\hat{\pi} : \mathbb{N} \to \mathbb{R}$. Autorzy rozważają dwa scenariusze:

- stała zmiana w każdej generacji: $\hat{\pi}(n) = \hat{\pi}_0 + \alpha \cdot n$
- skokowe zmiany występujące co k generacji: $\hat{\pi}(n) = \hat{\pi}_0 + \beta \cdot \lfloor \frac{n}{k} \rfloor$

gdzie α, β są parametrami opisującymi szybkość zmian optimum, zaś $\hat{\pi}_0$ jest parametrem modelu opisującym fenotyp optymalny na starcie symulacji. Zmiana optymalnego fenotypu zamyka cykl życia opisywanego modelu obliczeniowego. Zaproponowany model może być uznany za wielowymiarowy geometryczny model Fishera ze zmiennym optimum.

3.2.2. Wyniki

Wyniki symulacji wskazują, że w obu scenariuszach *transposition-selection equilibrium* jest osiągane, a ponadto aktywność transpozonów w organizmach wzrasta na skutek pojawiających się zmian środowiskowych. Silniejsze zmiany fenotypowe generowane dodatkowymi

mutacjami indukowanymi przez transpozony wydają się ułatwiać organizmom *dążenie* do optimum fenotypowego. Takie rezultaty stanowią mocną przesłankę do wyprowadzenia analitycznych modeli, które pozwoliłyby potwierdzić wyniki symulacji. Przykładowa konstrukcja zostanie skrótowo przedstawiona w dalszej części pracy.

Rozdział 4

Metody formalizacji modelu

W tym rozdziale uwaga zostanie poświęcona wyrażeniu przedstawionych modeli obliczeniowych przy pomocy formalego aparatu matematycznego. W pierwszej części przedstawione i poddane dyskusji zostaną wybrane modele analityczne zaproponowane w [Cha83], których celem jest znalezienie wartości n liczby transpozonów na genomie w stanie równowagi. Ponadto, jak zostało wcześniej wspomniane, przeanalizowana zostanie zasadność pewnych założeń przyjętych przez autorów w trakcie tworzenia modelu obliczeniowego. Podobnie jak poprzednio dokonane zostanie sformalizowanie notacji, np. w miejsce częstości x_i pojawi się zmienna losowa X_i przyjmująca wartość 1 z prawdopodobieństwem x_i oraz 0 w przeciwnym przypadku.

Następnie podjęta zostanie próba przedstawienia jednej z możliwych dróg dokonania formalizacji modelu Startek, Le Rouzic i in. naświetlone zostaną jej wady jak i zalety.

4.1. Model B. Charlesworth, D. Charlesworth

4.1.1. Regulowana transpozycja w nieskończonej populacji

Pierwszy model analityczny proponowany przez autorów rozważa nieskończoną populację, w której nie występuje selekcja, natomiast prawdopodobieństwo wystąpienia transpozycji i delecji, zgodnie z modelem obliczeniowym, jest odpowiednio malejącą funkcją $u(n)$ liczby transpozonów n na genomie organizmu i stałą v.

Zauważmy na wstępie, że n jest zmienną losową przyjmującą wartości naturalne. Rozważamy ciąg X_i zmiennych losowych określających czy i-ty *locus* jest zajęty taki, że

$$\mathbb{P}(X_i = 1) = x_i = 1 - \mathbb{P}(X_i = 0)$$

Oznaczając przez $\overline{n}$ średnią liczbę transpozonów na pojedynczym organizmie, mamy

$$\overline{n} = \mathbb{E}\left(2\sum_{i=1}^{T/2} X_i\right) = 2\sum_{i=1}^{T/2} \mathbb{E}(X_i) = 2\sum_{i=1}^{T/2} 1 \cdot x_i = 2\sum_{i=1}^{T/2} x_i$$

Celem znalezienia stanu stacjonarnego, rozważana jest średnia zmiana wartości $\overline{n}$ między dwiema kolejnym generacjami, oznaczana jako $\Delta\overline{n}$. Taka zmiana jest określona jako różnica pomiędzy liczbą nowopowstałych transpozonów, a liczbą tych, które uległy delecji

$$\Delta\overline{n} = \mathbb{E}(n \cdot u(n) - n \cdot v) = \mathbb{E}(n \cdot u(n)) - \overline{n} \cdot v$$

Celem wyznaczenia wartości oczekiwanej z $u(n) \cdot n$, autorzy rozwijają funkcję wokół $\overline{n}$ i otrzymują przybliżenie:

$$\Delta\overline{n} \approx \overline{n}(u_{\overline{n}} - v) + \frac{\mathrm{Var}(n)}{2}\left(2\frac{\partial u_{\overline{n}}}{\partial \overline{n}} + \overline{n}\frac{\partial^2 u_{\overline{n}}}{\partial \overline{n}^2}\right) \tag{4.1}$$

Chcąc powtórzyć ten wynik, można się spodziewać, że należy korzystać ze wzoru Taylora

Definicja: (Wzór Taylora) *Niech $f : \mathbb{R} \to \mathbb{R}$, będzie funkcją n-krotnie różniczkowalną w punkcie $a \in \mathbb{R}$. Wówczas istnieje funkcja $h_{n+1} : \mathbb{R} \to \mathbb{R}$ taka, że:*

$$\sum_{k=0}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) + h_{n+1}(x)(x-a)^{(n+1)}$$

gdzie $\lim_{x\to a} h_{n+1}(x) = 0$.

Funkcja $u(n) = 1 - s \cdot n^t$ dla $n > 0$ spełnia założenia, zatem korzystając ze wzoru Taylora możemy zapisać:

$$u(n) = u(\overline{n}) + u'(\overline{n})(n - \overline{n}) + u''(\overline{n})\frac{(n-\overline{n})^2}{2} + \hat{h}(n)$$

gdzie $\hat{h}(n) \overset{n\to\overline{n}}{\longrightarrow} 0$. Mnożąc obustronnie przez n i kładąc wartość oczekiwaną mamy:

$$\mathbb{E}\left(n \cdot u(n)\right) = \mathbb{E}\left(n \cdot u(\overline{n}) + n \cdot u'(\overline{n})(n - \overline{n}) + n \cdot u''(\overline{n})\frac{(n-\overline{n})^2}{2} + n \cdot \hat{h}(n)\right)$$

następnie korzystając z liniowości wartości oczekiwanej oraz tożsamości $\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$ otrzymujemy:

$$\mathbb{E}\left(n \cdot u(n)\right) = \overline{n} \cdot u(\overline{n}) + u'(\overline{n})\mathrm{Var}(n) + \frac{u''(\overline{n})}{2}\mathbb{E}\left(n \cdot (n-\overline{n})^2\right) + \mathbb{E}\left(n \cdot \hat{h}(n)\right) \tag{4.2}$$

Okazuje się zatem, że autorzy korzystając ze wzoru Taylora stosują przybliżenie nie tylko ze względu na resztę, ale także na błąd związany z założeniem o niezależności zmiennych n i $(n-\overline{n})^2$. Przy tym założeniu mamy bowiem $\mathbb{E}\left(n \cdot (n-\overline{n})^2\right) = \mathbb{E}(n) \cdot \mathrm{Var}(n)$ i otrzymujemy (4.1) z dokładnością do oznaczeń na pochodne, tj.

$$\Delta\overline{n} \approx \overline{n}(u_{\overline{n}} - v) + \frac{\mathrm{Var}(n)}{2}\left(2u'(\overline{n}) + \overline{n}u''(\overline{n})\right) \tag{4.3}$$

Następny krok polega na oszacowaniu wariancji $\mathrm{Var}(n)$ na podstawie metod istniejących w literaturze, które są rozważane w przypadku wariancji genetycznej mediowanej przez obecność genów na skończonej liczbie *loci* na genomie.

Wariancja genetyczna

Bulmer (1985) wprowadza pojęcie wartości genotypowej (ang. *genotypic value*), które jest abstrakcyjną wartością przypisywaną genomowi na podstawie obecności na nim poszczególnych genów. W szczególnośći przyjmujemy, że każdy pojedynczy gen ma także swoją wartość genotypową, wówczas wartość genotypowa organizmu, to suma wartości genotypowych wszystkich genów należących do genomu rozważanego organizmu, tj.

$$G = X_{1m} + X_{1p} + X_{2m} + X_{2p} + \ldots + X_{nm} + X_{np}$$

gdzie G jest wartością genotypową organizmu, zaś X_{im}, X_{ip} to wartości genotypowe genów znajdujących się na i-tym *locus* gamety odpowiednio matczynej i ojcowskiej. Przyjmując oznaczenie C_{ij} na kowariancję wartości genotypowej genów i oraz j oraz zakładając, że zachodzi:

$$C_{ij} = Cov(X_{ip}, X_{jp}) = Cov(X_{im}, X_{jm})$$
$$C'_{ij} = Cov(X_{ip}, X_{jm}) = Cov(X_{im}, X_{jp})$$

otrzymujemy wyrażenie na wariancję genetyczną:

$$\mathrm{Var}(G) = \mathrm{Var}\left(\sum_{i=1}^{n}(X_{im} + X_{ip})\right) = 2\sum_{i=1}^{n} C_{ii} + 4\sum_{i<j} C_{ij} + 4\sum_{i<j} C'_{ij}$$

Dokładając jednak pewne biologiczne założenia dotyczące populacji, w której żyją organizmy (mowa tu o równowadze Hardy-Weinberga w populacji, jednak jej dokładny opis wykracza poza potrzeby tej pracy), zachodzi $4\sum\limits_{i<j} C'_{ij} = 0$, czyli ostatecznie mamy:

$$\mathrm{Var}(G) = \mathrm{Var}\left(\sum_{i=1}^{n}(X_{im} + X_{ip})\right) = 2\sum_{i=1}^{n} C_{ii} + 4\sum_{i<j} C_{ij}$$

Przywołaną definicję wariancji genetycznej autorzy wykorzystują do wyznaczenia wariancji liczby transpozonów, otrzymując:

$$\mathrm{Var}(n) = 2\sum_{i=1}^{T/2} \mathrm{Var}X_i + 4\sum_{i<j} C_{ij} \overset{\star}{=} \overline{n}\left(1 - \frac{\overline{n}}{T}\right) - T\sigma_x^2 + 4\sum_{i<j} C_{ij} \tag{4.4}$$

gdzie C_{ij} jest współczynnikiem opisującym *linkage disequlibrium* pomiędzy *loci* o numerach i oraz j.

Równość oznaczoną przez $\star$ w równaniu 4.4 przyjmujemy korzystając z definicji wariancji dla próby n-elementowej oraz dokonując kilku równoważnych przekształceń

Definicja: (Wariancja próbkowa) *Wariancją n-elementowej próby $(x_1, x_2, \ldots, x_n)$ nazwiemy wartość:*

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

gdzie m jest standardową średnią z próby, tj. $m = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$

Chcąc wykazać równość $\star$ wyznaczamy

$$T \cdot \sigma_x^2 = T \cdot \frac{2\sum\limits_{i=1}^{T/2}(x_i - m)^2}{T} = 2\sum_{i=1}^{T/2}\left(x_i - \frac{2}{T}\sum_{i=1}^{T/2} x_i\right)^2 = 2\sum_{i=1}^{T/2}\left(x_i - \frac{\overline{n}}{T}\right)^2 =$$
$$= 2\sum_{i=1}^{T/2} x_i^2 - \frac{4\overline{n}}{T}\cdot\sum_{i=1}^{T/2} x_i + 2\sum_{i=1}^{T/2}\left(\frac{\overline{n}}{T}\right)^2 = 2\sum_{i=1}^{T/2} x_i^2 - \frac{2\overline{n}^2}{T} + \frac{\overline{n}^2}{T} = 2\sum_{i=1}^{T/2} x_i^2 - \frac{\overline{n}^2}{T}$$

Kładąc wyznaczoną wartość $T \cdot \sigma_x^2$ do prawej strony równości $\star$ otrzymujemy

$$\overline{n}\left(1 - \frac{\overline{n}}{T}\right) - 2\sum_{i=1}^{T/2} x_i^2 + \frac{\overline{n}^2}{T} = \overline{n} - 2\sum_{i=1}^{T/2} x_i^2 = 2\sum_{i=1}^{T/2} x_i - 2\sum_{i=1}^{T/2} x_i^2 = 2\sum_{i=1}^{T/2} x_i(1 - x_i) = 2\sum_{i=1}^{T/2}\mathrm{Var}(X_i)$$

co ostatecznie dowodzi $\star$.

W tym miejscu wyprowadzania modelu wprowadzone są kolejne założenia dotyczące aktywności transpozonów. Mianowicie przyjmujemy, że efekty związane z *linkage disequlibrium* są na tyle małe, że zaniedbywalne. Ponadto należy zauważyć,że w rozważanej, nieskończonej, populacji zdarzenia transpozycji i delecji z czasem wyrównają prawdopodobieństwa x_i pojawienia się transpozonu pomiędzy wszystkimi *loci*.

Z powyższego wynika, że w modelu przyjmujemy następujące założenia: $4\sum_{i<j} C_{ij} = 0$ oraz $\sigma_x^2 \to 0$ co implikuje oszacowanie na wariancję liczby transpozonów postaci $\mathrm{Var}(n) = \overline{n}\left(1 - \frac{\overline{n}}{T}\right)$ przykładając do (4.3) mamy:

$$\Delta\overline{n} \approx \overline{n}(u_{\overline{n}} - v) + \frac{\overline{n}}{2}\left(1 - \frac{\overline{n}}{T}\right)\left(2u'(\overline{n}) + \overline{n}u''(\overline{n})\right) \tag{4.5}$$

W tym miejscu uwaga autorów skupia się na czynniku $2u'(\overline{n}) + \overline{n}u''(\overline{n})$, twierdzą oni, że zakładając brak silnej regulacji transpozycji, tzn. $u(n)$ jest dostatecznie dobrze przybliżana przez funkcję stałą, możemy uznać, że wartość $2u'(\overline{n})+\overline{n}u''(\overline{n})$ jest równa 0. Skąd otrzymujemy $\Delta\overline{n} \approx \overline{n}\,(u(\overline{n}) - v)$ i możemy wnioskować, że nietrywialnym stanem stacjonarnym $\hat{n}$ liczby transpozonów jest stan zadany wyrażeniem:

$$u(\hat{n}) = v$$

Dodatkowo przykładając funkcję u określającą aktywność transpozycji zadaną wzorem $u(n) = \frac{u_0}{1+kn}$ autorzy wyznaczają jawny wzór na liczbę transpozonów w stanie stacjonarnym, tj.:

$$\hat{n} = \frac{u_0 - v}{kv}$$

podkreślając równocześnie, że powyższe przybliżenia są dostatecznie dobre przy założeniu, że $k \ll 1$.

Dygresja

Ze względu na otrzymany wynik zasadnym wydaje się zadanie następującego pytania: czy wszystkie powyższe obliczenia były konieczne? Choć weryfikacja ich poprawności nie wykazała szczególnych nieprawidłowości, to wydaje się, że w momencie w którym autorzy oznajmiają, że regulacja transpozycji jest na tyle słaba, że $2u'(\overline{n}) + \overline{n}u''(\overline{n}) \approx 0$, to niejako wykluczają wszystko co mogłoby powstrzymać czytelnika przed następującymi wnioskami.

Poszukujemy liczby transpozonów, dla której zdarzenia transpozycji będą równoważyć zdarzenia delecji. Dodatkowo wiemy, że zdarzenia mają stałe częstości występowania odpowiednio $u(n) \approx const$ dla urodzin i v dla śmierci. Wówczas zakładając istnienie nietrywialnego stanu stacjonarnego $\hat{n}$, wiemy, że musi on spełniać równość:

$$\mathbb{E}\left(\hat{n} \cdot u(\hat{n}) - \hat{n} \cdot v\right) = 0$$

skąd natychmiast otrzymujemy końcowy wniosek autorów, że liczba transpozonów w stanie stacjonarnym zadana jest wyrażeniem $u(\hat{n}) = v$.

4.1.2. Transpozycje i selekcja

W kolejnym modelu autorzy rozważają model, w którym selekcja kontroluje powielanie się tranpozonów w nieskończonej populacji. Przyjęto założenie, że prawdopodobieństwa transpozycji i delecji są stałymi, odpowiednio u i v, niezależnymi od ilości transpozonów na genomie.

Dopasowanie organizmu zawierającego n kopii transpozonów jest z założenia malejącą funkcją n: $\omega(n)$. Pozostała notacja pozostaje niezmieniona względem pierwszego modelu.

W celu dalszego modelowania zjawiska selekcji naturalnej należy przytoczyć fragment teorii, na której opierają się autorzy, autorstwa Sewalla Wrighta zaliczanego do grona twórców genetyki populacji.

Twierdzenie: (Formuła Wrighta) *Rozważmy parę alleli a, A, które pojawiają się na ustalonym locus z prawdopodobieństwami odpowiednio x i $1-x$, w diploidalnej populacji rozmnażającej się w sposób płciowy z losowym kojarzeniem w pary. Ponadto niech ω będzie funkcją wyznaczającą dopasowanie organizmu z rozważanej populacji.*
Wówczas zmiana prawdopodobieństwa występowania allelu a na l-tym locus mediowana przez selekcję wynosi:

$$\Delta_s x = \frac{x(1-x)}{2\overline{\omega}} \frac{d\overline{\omega}}{dx}$$

gdzie $\overline{\omega}$ oznacza średnie dopasowanie w całej populacji.

Dowód: *Przyjmujemy następujące oznaczenia, jak w treści twierdzenia gdzie s_1, s_2, s_3 są*

allel/genotyp	a	A	aa	aA	AA
p-stwo wystąpienia	x	1-x	x^2	$2x(1-x)$	$(1-x)^2$
dopasowanie $\omega(\cdot)$	-	-	s_1	s_2	s_3

pewnymi stałymi opisującymi dopasowanie odpowiednio genotypów aa, aA oraz AA. Wówczas, średnie dopasowanie w populacji wynosi:

$$\overline{\omega} = x^2 s_1 + 2x(1-x)s_2 + (1-x)^2 s_3$$

natomiast zmiana prawdopodobieństwa występowania allelu a wynikająca ze zdarzenia selekcji może zostać wyrażona jako różnica prawdopodobieństwa po i przed jej zajściem, gdzie prawdopodobieństwo wystąpienia po selekcji jest stosunkiem liczby alleli typu a po selekcji, do liczby wszystkich alleli, które przetrwały selekcję, czyli średniego dopasowania populacji. Przekształcając otrzymujemy:

$$\Delta_s x = \frac{x^2 s_1 + x(1-x)s_2}{\overline{\omega}} - x = \frac{x(1-x)}{2\overline{\omega}} \left(\frac{2xs_1}{1-x} + 2s_2 - \frac{2\overline{\omega}}{1-x} \right) \tag{4.6}$$

Zauważmy ponadto, że zachodzi:

$$\begin{aligned}
\frac{d\overline{\omega}}{dx} =& \frac{d}{dx}\left(x^2 s_1 + 2x(1-x)s_2 + (1-x)^2 s_3 \right) = 2xs_1 + 2s_2(1-2x) - 2(1-x)s_3 = \\
&= \frac{2x(1-x)s_1 + 2(1-x)(1-2x)s_2 - 2(1-x)^2 s_3}{1-x} = \\
&= \frac{2xs_1 + 2(1-x)s_2 - 2x^2 s_1 - 4(1-x)xs_2 - 2(1-x)^2 s_3}{1-x} = \\
&= \frac{2xs_1 + 2(1-x)s_2 - 2\overline{\omega}}{1-x}
\end{aligned}$$

Podstawiając wyznaczoną wartość pod (4.6) *otrzymujemy tezę twierdzenia.*

□

Pierwszy krok polega na wyznaczeniu zmiany prawdopodobieństwa występowania transpozonu na konkretnym *locus* w populacji, po zajściu transpozycji, tj. $\Delta_t x_i$ dla każdego i. W tym celu zauważamy, że prawdopodobieństwo pojawienia się nowego transpozonu na pustym *locus* jest równe $\frac{u\overline{n}}{T-\overline{n}}$, czyli stosunkowi oczekiwanej liczby nowych transpozonów do liczby wolnych *loci*.

Wówczas na mocy tw. o prawdopodobieństwie całkowitym mamy:

$$\begin{aligned}\mathbb{P}\left(X_i'=1\right) =& \mathbb{P}\left(X_i'=1\middle|\, X_i=1\right)\mathbb{P}\left(X_i=1\right)+\mathbb{P}\left(X_i'=1\middle|\, X_i=0\right)\mathbb{P}\left(X_i=0\right)= \\ &= (1-v)x_i+\frac{u\overline{n}}{T-\overline{n}}(1-x_i)\end{aligned}$$

dzięki czemu wyznaczamy $\Delta_t x_i$, jako:

$$\Delta_t x_i = \mathbb{P}\left(X_i'=1\right)-\mathbb{P}\left(X_i=1\right) = (1-v)x_i+\frac{u\overline{n}}{T-\overline{n}}(1-x_i)-x_i = \frac{u\overline{n}}{T-\overline{n}}(1-x_i)-vx_i$$

Modelowanie selekcji polega na zastosowaniu formuły Wrighta, gdzie dwa allele są reprezentowane, jako występowanie bądź nie transpozonu na danym *locus*, z prawdopodobieństwami odpowiednio $x_i, 1-x_i$, skąd:

$$\Delta_s x_i = \frac{x_i(1-x_i)}{2\overline{\omega}}\frac{d\overline{\omega}}{dx}$$

i całkowita zmiana prawdopodobieństwa wystąpienia transpozonu na i-tej pozycji Δx_i wynosi:

$$\Delta x_i = \Delta_s x_i + \Delta_t x_i = \frac{x(1-x)}{2\overline{\omega}}\frac{d\overline{\omega}}{dx}+\frac{u\overline{n}}{T-\overline{n}}(1-x_i)-vx_i \tag{4.7}$$

Przybliżając średnie dopasowanie w populacji $\overline{\omega}$ przez $\omega(\overline{n})$ wyznaczamy $\frac{1}{2\overline{\omega}}\frac{d\overline{\omega}}{dx_i}$, jako:

$$\frac{1}{2\overline{\omega}}\frac{d\overline{\omega}}{dx_i} = \frac{1}{2\omega(\overline{n})}\frac{d\omega(\overline{n})}{dx_i} = \left(\frac{1}{\omega(\overline{n})}\frac{d\omega(\overline{n})}{d\overline{n}}\right)\left(\frac{1}{2}\frac{d\overline{n}}{dx_i}\right) = \frac{d\ln\omega(\overline{n})}{d\,\overline{n}} \tag{4.8}$$

gdyż $\frac{1}{2}\frac{d\overline{n}}{dx_i} = \frac{1}{2}\frac{d}{dx_i}\left(2\sum x_i\right) = 1$ i przykładając do (4.7) otrzymujemy, że:

$$\Delta x_i = x_i(1-x_i)\frac{d\ln\omega(\overline{n})}{d\,\overline{n}}+\frac{u\overline{n}}{T-\overline{n}}(1-x_i)-vx_i \tag{4.9}$$

Podobnie jak w poprzednim modelu autorzy zauważają, że w nieskończonej populacji, ostatecznie, prawdopodobieństwa występowania transpozonów wyrównają się pomiędzy wszystkimi *loci*, a wówczas zachodzić będzie $\overline{n} = Tx$, gdzie x jest wspomnianym prawdopodobieństwem. W tym stanie równowagi, równanie (4.9) przybiera postać:

$$\Delta x = x(1-x)\frac{d\ln\omega(\overline{n})}{d\,\overline{n}}+\frac{u\overline{n}}{T-Tx}(1-x)-vx = x(1-x)\frac{d\ln\omega(\overline{n})}{d\,\overline{n}}+x(u-v)$$

z której po obustronnym wymnożeniu przez T otrzymujemy zmianę średniej liczby traspozonów na pojedynczym organizmie w populacji:

$$\Delta\overline{n} = \overline{n}(1-\frac{\overline{n}}{T})\frac{d\ln\omega(\overline{n})}{d\,\overline{n}}+\overline{n}(u-v) = \overline{n}\left[(1-\frac{\overline{n}}{T})\frac{d\ln\omega(\overline{n})}{d\,\overline{n}}+(u-v)\right] \tag{4.10}$$

Podsumowując, założenie o malejącej funkcji dopasowania zależnej od liczby transpozonów w genomie pozwala na uzyskanie wyniku analogicznego jak w poprzednim podejściu. Niemniej, jak już wcześniej zaznaczono, tak silne założenie nałożone na funkcję ω, wiąże się z zaszyciem w wyprowadzanej konstrukcji odpowiedzi na stawiane pytanie o istnienie *transposition-selection equilibrium*.

4.2. Model Startek, Le Rouzic i in.

4.2.1. Operator populacyjny

Ostanią część pracy stanowi zaprezentowanie alternatywnej metody wyprowadzenia analitycznego modelu proliferacji transpozonów, uwzględniającej stres środowiskowy [Sta14]. W tym celu populację opisujemy gęstością ρ na $\mathbb{R}^n \times \mathbb{N}$ (patrz rys. 4.1), zaś cykl życia opisujemy operatorem Φ, który przyłożony do gęstości ρ opisuje populację $\Phi(\rho)$ w następnym pokoleniu.

4.2.2. Założenia

Przyjmujemy następujący zestaw założeń:

- Delecje i transpozycje występują równocześnie i nie mają na siebie wpływu, tj. prawdopodobieństwa zajścia tych zdarzeń są od siebie niezależne.
- $\theta = (\phi, n) \in \mathbb{R}^n \times \mathbb{N}$ jest reprezentacją organizmu rozważanej populacji.
- Wystąpienie d delecji w pojedynczym organizmie θ modelujemy rozkładem dwumianowym:
$$\binom{n}{d}\delta^d(1-\delta)^{n-d}$$
gdzie δ jest częstością występowania delecji.
- Wystąpienie k transpozycji w pojedynczym organizmie θ modelujemy rozkładem Poissona:
$$\frac{(\mu n)^k}{k!}e^{-\mu n}$$
gdzie μ jest częstością występowania transpozycji.
- Mutacje zmieniają rozkład fenotypu w populacji poprzez splot z gęstością rozkładu normalnego $\mathcal{N}(0, \sigma^2 + \sigma_k^2)$, gdzie σ^2, σ_k^2 opisują natężenie mutacji odpowiednio nie związanych i związanych z transpozonami (zależnych od ilości zdarzeń tranpozycji).
- Istnieje optymalny fenotyp środowiska równy 0, do którego dążą wszystkie organizmy. Stres środowiskowy modelowany jest jako zmiana fenotypu wszystkich organizmów w populacji o stałą wartość η.
- Selekcja naturalna wybiera najlepiej dopasowane organizmy zgodnie ze scentrowanym rozkładem normalnym $\mathcal{N}(0, \xi^2)$, gdzie ξ^2 jest zasięgiem selekcji, zaś 0 opisuje optymalny fenotyp.

4.2.3. Definicja operatora

Przy takim zestawie przyjętych założeń i przyjmując dla ułatwienia zapisu, że wymiar wektora fenotypu $n = 1$ pełny operator populacji $\Phi : \mathbb{R} \times \mathbb{N} \to \mathbb{R} \times \mathbb{N}$ przybiera postać

$$\Phi(\rho)(n, \phi) = \frac{\hat{\Phi}(\rho)(n, \phi)}{\sum_{n=0}^{\infty} \int_{\mathbb{R}^n} \rho(n, \cdot)}$$

gdzie $\hat{\Phi} : \mathbb{R} \times \mathbb{N} \to \mathbb{R} \times \mathbb{N}$ jest zdefiniowany jako

$$\hat{\Phi}(\rho)(n, \phi) = \nu(0, \xi^2) \cdot \sum_{d=0}^{\infty} \sum_{k=0}^{\infty} \left(\binom{n}{d}\delta^d(1-\delta)^{n-d} \cdot \frac{(\mu n)^k}{k!}e^{-\mu n} \cdot \rho(n, \cdot) \star \nu(0, \sigma^2 + \sigma_k^2) \right)(\phi - \eta)$$

W powyższym zapisie $\nu(0, \sigma^2)$ należy rozumieć, jako gęstość rozkładu normalnego o średniej 0 i wariancji σ^2, zaś $\star$ opisuje standardowy splot dwóch funkcji.

4.2.4. Wyniki

Badanie powyższego operatora polega na wykazaniu istnienia punktu stałego. Dotychczas udowodniono, że prostszy operator,

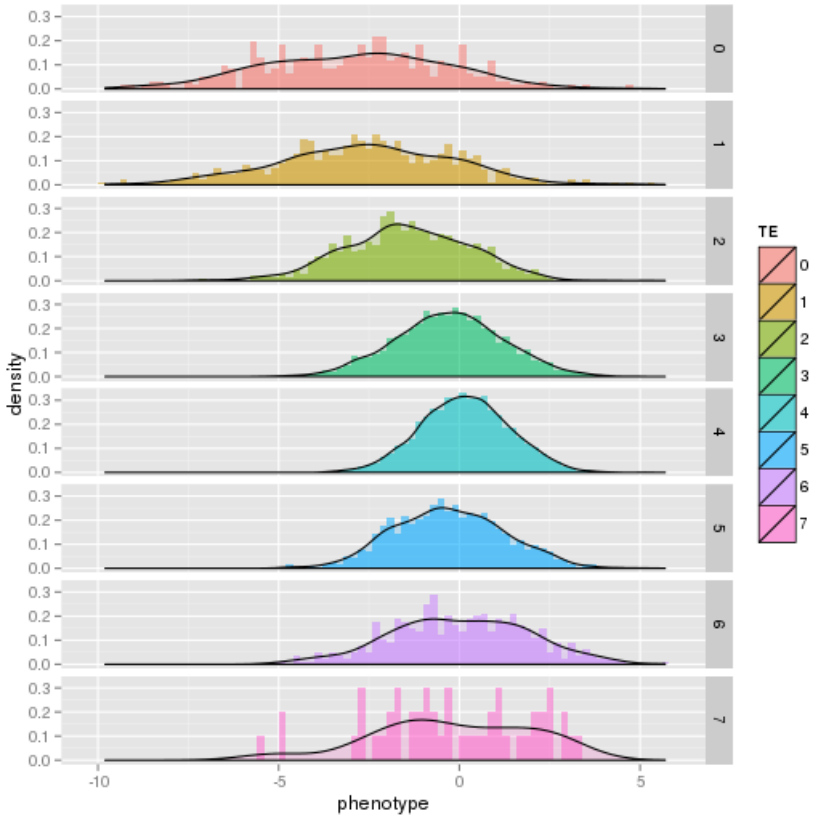
$$\widetilde{\Phi}(\rho)(\phi) = (\rho \star \nu(0, \sigma^2))(\phi - \eta) \cdot \nu(0, \xi^2)(\phi)$$

który modeluje losowe mutacje jest zbieżny i posiada nietrywialny punkt stały, będący gęstością ρ_0 rozkładu normalnego

$$\mathcal{N}\left(\frac{\eta\sqrt{4\xi^2+\sigma^2} - \eta \cdot \sigma}{2\sigma}, \frac{\sqrt{2\sigma\left(\sqrt{4\xi^2+\sigma^2} - \sigma\right)}}{2}\right)$$

którego wariancja nie zależy od natężenia stresu środowiskowego η [Sta14].

Ponownie należy podkreślić, że powyższy model analitycznym nie zawiera związku między liczbą transpozonów a dopasowaniem organizmu do życia. Mimo to istnienie punktu stałego operatora $\widetilde{\Phi}$ pokazuje, że może on istnieć także dla operatora Φ, co potwierdzałoby przypuszczenia wynikające z modelu obliczeniowego o istnieniu *transposition-selection equilibrium* w populacjach narażonych na stres środowiskowy.



Rysunek 4.1: Przykładowa reprezentacja populacji. Kolejne wykresy opisują rozkład fenotypu w podpopulacjach posiadających dokładnie $n \in \{0, \ldots, 7\}$ transpozonów.

Rozdział 5

Podsumowanie

W pracy zostały omówione analityczne, jak i obliczeniowe modele proliferacji transpozonów, zarówno z uwzględnieniem, jak i zaniedbaniem oddziaływania stresu środowiskowego. Wskazano pewne wątpliwe założenia w modelach [Cha83], których konstrukcja zawiera zaszytą odpowiedź na stawiane przez autorów pytania.

Dodatkowo przedstawiono alternatywne podejście do modelowania, które proponuje utożsamienie populacji z gęstością prawdopodobieństwa na nośniku $\mathbb{R}^n \times \mathbb{N}$. Dla takiej populacji można zdefiniować operator $\Phi : \mathbb{R}^n \times \mathbb{N} \to \mathbb{R}^n \times \mathbb{N}$ opisujący jeden cykl życia. Częściowe wyniki wskazują, że uproszczony operator $\widetilde{\Phi}$ posiada nietrywialny punkt stały, który stanowi przsłankę istnienia *transposition-selection equilibrium* w ogólnym przypadku.

Chcąc okazać, że stres środowiskowy może być jednym z regulatorów aktwności transpozonów dalsze prace nad opisanymi zagadnieniami można prowadzić dwojako. Jedna z metod polega na przeformułowaniu założeń przedstawionych w [Cha83], a konkretnie ich osłabieniu. Inna metoda polega na wykazaniu zbieżności i istnienia punktu stałego uogólnionego operatora Φ przedstawionego w [Sta14].

5.1. Podziękowania

Praca ta była wspierana przez Narodowe Centrum Nauki, grant: *Modelowanie aktywności transpozonów indukowanej stresem środowiskowym*, numer: 2012/06/M/ST6/00438.

Bibliografia

[Cap00] Capy P., et al. (2007) *Stress and transposable elements: coevolution or useful parasites?*, Heredity 85, 101-106

[Cha83] Charlesworth B., Charlesworth D. (1983) *The population dynamics of transposable elements*, Genetical Research, 42: 1-27.

[Hal90] Hall B. G. (1990) *Spontaneous point mutations that occur more often when advantageous than when neutral*, Genetics 126 (1): 5-16.

[Hic82] Hickey, D. A. (1982) *Selfish DNA: a sexually-transmitted nuclear parasite.*, Genetics 101, 519-553

[Hig08] Higgs, P. G., Attwood T. K. (2005) *Bioinformatyka i Ewolucja Molekularna*, Wydawnictwo Naukowe PWN, wyd. 1.: 22-52.

[Hua12] Huang C. R., Burns K. H., Boeke J. D. (2012) *Active transposition in genomes.*, Annual Review of Genetics 46: 651-675.

[Kaz04] Kazazian, H. H. (2004) *Mobile elements: Drivers of genome evolution.*, Science 303: 1626-1632.

[Kid97] Kidwell, M. G., Lisch D. (1997) *Transposable elements as sources of variation in animals and plants.*, Proceedings of the National Academy of Sciences of the USA vol. 94 no.15 :7704-7711.

[Orr09] Orr, H. A. (2009) *Fitness and its role in evolutionary genetics.*, Nature Reviews Genetics 10: 531-539.

[Pra09] Pray, L. (2008) *Transposons: The jumping genes.*, Nature Education 1 (1): 204

[Rid99] Ridley, M. (1999) *Genome*, str. 3-11, New York: HarperCollins Publishers, ISBN 0-06-019497-9.

[Sta13] Startek M, Le Rouzic A, Capy P, Grzebelus D, Gambin A. (2013) *Genomic parasites or symbionts? Modeling the effects of environmental pressure on transposition activity in asexual populations.*, Theoretical Population Biology 90: 145-51.

[Sta14] Startek, M. P. (2014) *On the existence and stability of equilibrium probability measures in Gaussian mutator models with environmental stress within Fisher's geometric framework.*, praca zaprezentowana w ramach ECMTB 2014, Göteborg, Szwecja.

[Weg12] Węgielski P., Bębas P. et al. (2012) *Genetyka Molekularna*, rozdz. I-III, Warszawa: PWN, ISBN: 978-83-01-14744-0.

[Wes96] Wesler S. R. (1996) *Plant retrotransposons: turned on by stress.*, Current Biology 6, 959-961

[Wic07] Wicker, T et al. (2007) *A unified classification system for eukaryotic transposable elements*, Nature Reviews Genetics 8: 973-982.