



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Scuola di
Scienze Matematiche
Fisiche e Naturali

Corso di Laurea in
Fisica e Astrofisica

Studio numerico di un modello di formazione di motivi per reazione-diffusione nello spazio genico

**Numerical study on a model for formation
of motives for reaction-diffusion in genetic
space**

Relatore
Franco Bagnoli

Candidato
Tommaso Tarchi

Indice

1	Introduzione	3
1.1	Speciazione simpatrica	6
2	Riproduzione asessuata	7
2.1	Modello macroscopico	7
2.1.1	Assenza di degenerazione	7
2.1.2	Degenerazione	8
2.2	Modello microscopico	9
2.3	Risultati	9
3	Riproduzione sessuata	13
3.1	Modello	13
3.2	Risultati	15
4	Considerazioni finali	17
A	Dettagli sul calcolo della degenerazione	18
A.1	Calcolo dei coefficienti di degenerazione	18
A.2	Metodo approssimato per calcolo della degenerazione	18
	Bibliografia	20

1 Introduzione

In questo elaborato di tesi esamineremo alcuni modelli teorici di una popolazione in evoluzione, e in particolare il processo di formazione di nuove specie in assenza di isolamento (*speciazione simpatica*).

Nel caratterizzare un organismo dal punto di vista evolutivo possiamo considerare due elementi fondamentali: il *genotipo* e il *fenotipo*. Il genotipo è il patrimonio genetico dell'individuo, ovvero l'insieme delle informazioni contenute nel suo DNA relative alla produzione delle proteine e degli enzimi per lo sviluppo dell'organismo stesso. Questo processo, insieme all'interazione con l'ambiente esterno, definisce i cosiddetti caratteri fenotipici, ovvero quelle manifestazioni esterne che determinano a loro volta l'interazione con l'habitat e con gli altri individui della stessa specie; in questo lavoro assumeremo il genotipo come unico fattore determinante del fenotipo (non si considera per esempio l'invecchiamento). Sono esempi di caratteri fenotipici l'altezza, il colore degli occhi, la presenza o assenza di coda e la sua forma, la forza fisica, ecc... ma anche caratteristiche meno evidenti come la personalità, le preferenze sessuali e i gusti alimentari. Come ha evidenziato Dawkins [1], la definizione di fenotipo è sufficientemente generale da poter includere anche effetti a distanza sull'ambiente esterno (si considerino ad esempio le possibili conseguenze che può avere il lavoro di una comunità di castori sulla morfologia di un'area geografica).

Il processo di formazione dei caratteri fenotipici è in realtà estremamente complesso, e in generale obbedisce al così detto *dogma centrale della biologia molecolare* (dogma al quale in realtà esistono numerose eccezioni, quali ad esempio i retrovirus), secondo il quale l'informazione genetica fluisce sempre dagli acidi nucleici (che compongono il DNA) verso le proteine; in altre parole il genotipo determina il fenotipo, e non viceversa. In senso lato possiamo considerare la selezione naturale come un flusso di informazioni nel senso opposto. Infatti le mutazioni e la ricombinazione genetica (di cui parleremo in seguito) producono nuovi genotipi a partire da quelli delle generazioni precedenti, e quelli tra di essi che determinano fenotipi più adatti alla sopravvivenza permettono ai loro possessori di raggiungere l'età della riproduzione. Essi vengono così trasmessi alla generazioni successive e diffusi nella popolazione, mentre altri scompaiono; si ha cioè una sorta di feedback dovuto all'interazione tra il fenotipo e l'ambiente.

Seguendo Williams [2, pagina 26], possiamo definire un *gene* come una qualche porzione di materiale cromosomico che potenzialmente si conserva per abbastanza generazioni per servire come unità di selezione. Possiamo quindi pensarlo come una certa quantità di informazioni genetiche, che cambia estensione a seconda del fenotipo che consideriamo; dopotutto anche la

definizione di un carattere fenotipico (ovvero della sua ‘estensione’) è del tutto arbitraria. Lo spazio che il gene occupa all’interno del *genoma* (termine che useremo come sinonimo di genotipo, anche se a rigore il primo conterrebbe anche le porzioni di DNA non codificanti) è detto *locus*, e le diverse ‘versioni’ del gene che possono occupare quel locus *alleli* (ad esempio geni per occhi verdi o azzurri).

Oltre alla selezione naturale, sono almeno due i processi che è necessario prendere in considerazione quando parliamo di evoluzione: la *riproduzione* e la *mutazione*.

Schematicamente possiamo indicare due tipi di riproduzione: asessuata e sessuata. Nel primo caso il singolo organismo si scinde producendo due discendenti (ma noi considereremo una semplice replicazione) identici al genitore a meno di mutazioni. Il secondo è un processo generalmente più lento, in cui due individui della stessa specie si accoppiano producendo discendenti il cui corredo genetico risulta dalla combinazione di quelli dei genitori. In generale la riproduzione asessuata è caratteristica degli organismi *aploidi*, i quali posseggono un solo set di cromosomi; quella sessuata al contrario è caratteristica degli organismi *diploidi*, che posseggono due set di cromosomi, uno ereditato dal padre e uno dalla madre. In questo lavoro studieremo sia riproduzione asessuata che sessuata supponendo specie aploidi.

Una mutazione consiste nella possibile variazione casuale del genotipo in uno o più punti della sequenza. Ce ne sono di molti tipi e possono essere immediatamente letali (ad esempio se impediscono il ripiegamento di proteine essenziali). Nel seguito considereremo tutte le mutazioni non immediatamente letali.

Abbiamo usato la parola *specie* dando per scontato il suo significato. In realtà la definizione di questo concetto è tutt’altro che univoca. Per praticità lo sostituiremo con quello, introdotto da Eigen, di *quasispecie*, ovvero una distribuzione di genomi piccata attorno ad una sequenza maestra localizzata in corrispondenza di un massimo della *fitness*.

Un altro concetto centrale è appunto quello di fitness. Ne esistono almeno cinque definizioni, non sempre sovrapponibili [1, capitolo 10]. Nella maggior parte delle applicazioni viene definita come il numero medio di figli che raggiunge l’età della riproduzione, ma ai fini del presente lavoro possiamo considerarla semplicemente come la probabilità di sopravvivenza dell’individuo. I fattori che la determinano sono sia ambientali (disponibilità di risorse, predatori, cambiamenti climatici, ecc...) che intraspecifici, ovvero dovuti all’interazione con gli altri individui della stessa specie; e possono essere sia di natura cooperativa che competitiva. Seguendo Bagnoli e Bezzi [4] abbiamo modellizzato la fitness tramite due contributi, uno statico relativo all’interazione con l’ambiente esterno e l’altro dinamico relativo alla sola competizione

tra gli individui. Si è quindi fatta l'ipotesi che le condizioni ambientali varino sensibilmente solo su tempi estremamente lunghi, se confrontati con quelli caratteristici di evoluzione della specie. Il contributo dinamico risulta essere tanto più intenso quanto più due individui sono fenotipicamente simili; questa assunzione si basa sul fatto che due organismi simili tenderanno a sfruttare risorse simili e ad occupare la stessa nicchia ecologica. Non si considera la cooperazione intraspecifica.

Poiché la selezione naturale agisce solo sul fenotipo, la fitness dipende dal genotipo solo indirettamente attraverso il fenotipo. Possiamo ora pensare a due *spazi* in cui lavorare: quello *fenotipico* e quello *genotipico*. Considerando un unico carattere fenotipico (ma potremmo chiaramente prenderne un numero arbitrario: cambierebbe solo la dimensione dello spazio) e stabilendo due valori estremi di tale carattere, possiamo definire lo spazio fenotipico come un segmento con valori, a fini computazionali, discreti. Ad esempio, per una certa specie di pesce questi due estremi potrebbero rappresentare squame bianche e squame nere e i valori intermedi tutte le possibili tonalità di grigio. In questo spazio la fitness può essere rappresentata semplicemente come una funzione reale a valori reali del fenotipo, detta *panorama adattivo* (*fitness landscape*).

A questo punto possiamo scegliere di rappresentare il genoma di un individuo, alternativamente a seconda degli scopi, come un numero reale compreso in un intervallo oppure come una stringa di valori. Sceglieremo la prima rappresentazione nel caso in cui si voglia stabilire una corrispondenza biunivoca tra genotipo e fenotipo. In generale si possono avere più genotipi corrispondenti ad uno stesso fenotipo, pertanto possiamo parlare, in analogia con la fisica, di *degenerazione* di un fenotipo come del numero di genotipi ad esso corrispondenti. In questa prima rappresentazione del genoma non si ha degenerazione, e lo spazio fenotipico combacia con quello genotipico.

L'altra schematizzazione che useremo sarà quella in cui il genoma corrisponde ad una stringa di valori booleani (cioè due alleli per locus) di lunghezza L . In questo caso lo spazio genotipico può essere rappresentato come un ipercubo di dimensione 2^L (vedi figura 1), in cui ogni vertice rappresenta una sequenza genica possibile e ogni spigolo il collegamento tra due sequenze tramite *mutazione puntuale*, ovvero mutazione di un singolo bit della stringa; nel seguito tutte le mutazioni saranno sottintese di questo tipo. Il fenotipo sarà poi determinato dalla somma dei bit della stringa, il che schematizza un'interazione additiva (*non epistatica*) dei geni, nell'ipotesi in cui ogni gene interagisca con lo stesso peso. In questa rappresentazione lo spazio fenotipico è rappresentato da un segmento lungo $L+1$, e la degenerazione risulta uguale a $\binom{L}{f}$, con $f = f(x)$ fenotipo dipendente dal genotipo x .

Potremmo chiaramente considerare altri tipi di relazione tra il fenotipo e

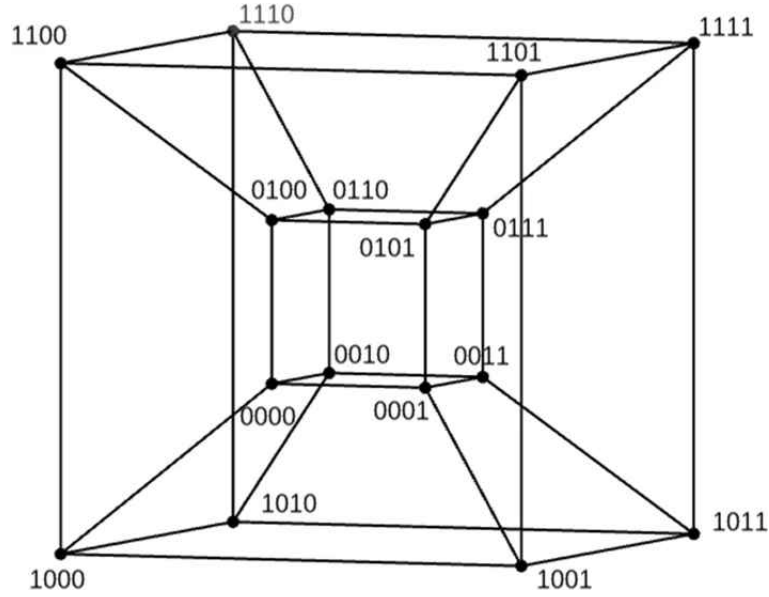


Figura 1: rappresentazione dello spazio genotipico per $L = 4$.

il genotipo, e la degenerazione (dove presente) dipenderebbe dalla scelta.

In ogni caso il tempo è rappresentato da una variabile discreta che simula il succedersi delle generazioni. Ogni generazione sarà divisa, a seconda del modello, in più fasi (ad es. riproduzione, mutazione, competizione).

1.1 Speciazione simpatrica

Il fenomeno della speciazione (ovvero la formazione di nuove specie a partire da una preesistente) è stato per anni considerato possibile solo in condizioni di isolamento fisico (*speciazione allopatrica*), ovvero in presenza di barriere naturali che separino fisicamente una porzione della popolazione, a quel punto libera di evolversi autonomamente. In seguito molti studi (ad esempio [3]) si sono concentrati sulla possibilità che tale fenomeno si verifichi anche in assenza di barriere fisiche. Questo particolare processo è detto speciazione simpatrica, e può essere studiato sia in un contesto sessuato che asessuato. Lo scopo principale di questo lavoro è quello di riprodurre e se possibile espandere risultati già ottenuti in entrambe le condizioni.

Vedremo che per specie asessuate sarà possibile osservare il fenomeno solo in presenza di una sufficiente intensità di competizione intraspecifica, e che la distribuzione finale della popolazione dipenderà dalla forma della distribuzione delle risorse. Per specie sessuate, invece, il fattore determinante

sarà l'*assortatività*, ovvero la preferenza nella scelta del partner dipendente da qualche tratto genetico (no *random mating*).

2 Riproduzione asessuata

In questa sezione cerchiamo di riprodurre i risultati ottenuti in Bagnoli e Bezzi [4] sia nel caso non degenerare che in quello degenerare e indagando lo stesso fenomeno anche dal punto di vista ‘microscopico’, ovvero considerando l’interazione tra i singoli individui. Proveremo anche a modificare la distribuzione delle risorse.

2.1 Modello macroscopico

Il seguente modello consiste in un’estensione al caso degenerare di quello di Bagnoli e Bezzi, che di seguito riassumiamo. In ogni caso sarà caratterizzato da una popolazione costante e da un tasso di mutazione anch’esso costante. Inoltre il processo evolutivo potrà sempre essere suddiviso in due fasi, una di riproduzione-mutazione e l’altra di competizione-sopravvivenza, e il risultato sarà indipendente dall’ordine in cui si implementano.

2.1.1 Assenza di degenerazione

Lo spazio genotipico è lineare e monodimensionale (ma il fenomeno di speciazione qui esposto è indipendente dalla dimensione) ed ogni individuo è descritto da un valore reale compreso tra 0 e L (genoma).

Nella fase di riproduzione-mutazione di ogni generazione gli individui appartenenti ad una determinata sequenza mutano genotipo da x a $x - 1$ e da x a $x + 1$ con probabilità μ .

La fitness $A(x)$ è definita come l’esponenziale di una funzione che chiamiamo *fitness totale* H , data dalla somma del contributo statico H_0 e di quello dinamico H_1 . Come fitness statica si considera una funzione simmetrica linearmente decrescente eccetto in prossimità di $x = \frac{L}{2}$, dove ha un massimo quadratico:

$$H_0(x) = b \left(1 - \frac{|x - \frac{L}{2}|}{r} - \frac{1}{1 + \frac{|x - \frac{L}{2}|}{r}} \right),$$

con b ed r che regolano intensità e pendenza della curva. Il contributo dinamico è invece dato da:

$$H_1(x) = -J \sum_y K \left(\frac{x - y}{R} \right) p(y, t),$$

con nucleo:

$$K(z) = \exp\left(-\frac{|z|^\alpha}{\alpha}\right),$$

dove J e α controllano l'intensità e la pendenza, e $p(y, t)$ rappresenta la distribuzione di popolazione nello spazio genotipico al tempo t .

Per la fitness statica abbiamo usato anche una forma alternativa, data dalla gaussiana:

$$H_0(x) = \frac{1}{\sqrt{2\pi}q} \exp\left(-\frac{|x - \frac{L}{2}|^2}{2q^2}\right).$$

In generale la distribuzione della degenerazione può essere approssimata ad una gaussiana, pertanto se lavoriamo vicino al suo massimo vale l'approssimazione di assenza di degenerazione.

2.1.2 Degenerazione

Consideriamo prima di tutto il caso in cui il genoma è rappresentato da una stringa di valori booleani di lunghezza L e μ è la probabilità di avere una mutazione puntuale (indipendentemente dalla posizione del gene). Un generico individuo con fenotipo x si trova in un vertice dell'ipercubo, o se preferiamo in un nodo del grafo, collegato ad altri L vertici che differiscono da esso per il valore di un solo bit. Pertanto, essendo x il numero di 1 nella sequenza, la probabilità che una mutazione porti l'individuo (o meglio suo figlio) in un nodo con fenotipo associato $x - 1$ è di $\mu \frac{x}{L}$; analogamente la probabilità di essere portato in un fenotipo $x + 1$ è pari a $\mu \frac{L-x}{L}$.

Il modello è in questo caso identico a quello senza degenerazione, eccezion fatta per il processo di riproduzione-mutazione in cui le probabilità di mutazione cambiano.

Come abbiamo visto la mutazione è più probabile verso fenotipi più degeneri. Ricordando ciò possiamo cercare di adattare il modello precedentemente descritto a tipi di degenerazione diversi da quello relativo a stringhe binarie. A tal fine abbiamo deciso di costruire una funzione che, a partire da una distribuzione di degenerazione supposta nella forma:

$$d(x) = D \exp\left(-\frac{\left|\frac{x - \frac{L}{2}}{k}\right|^h}{h}\right),$$

calcoli in modo approssimato l'evoluzione del sistema. In questo modello L è la lunghezza dello spazio fenotipico e μ la probabilità totale di avere mutazioni (uguale per tutti gli individui, come nel caso della stringa binaria).

Dopo aver calcolato la degenerazione di ogni fenotipo basandosi sulla distribuzione inserita, vengono calcolati $2L$ coefficienti di degenerazione di valore compreso tra 0 e 1. Ogni fenotipo x è associato a due coefficienti (per calcolo vedi appendice), i quali moltiplicati per il tasso di mutazione danno la probabilità che si abbia mutazione verso $x - 1$ e verso $x + 1$; in altre parole i due coefficienti regolano il ‘flusso’ di popolazione verso i fenotipi vicini.

2.2 Modello microscopico

Prendendo spunto da Bagnoli e Guardiani [5] abbiamo costruito un modello che prendesse come ‘unità fondamentale’ non più il genoma ma il singolo individuo. Anche il questo caso supponiamo popolazione N e tasso di mutazione μ costanti. Ogni individuo è caratterizzato da un genoma ‘binario’ di lunghezza L .

In ogni fase di riproduzione, e per ogni individuo, la mutazione avviene con probabilità μ su un locus scelto casualmente. In seguito ogni nuovo nato è sottoposto ad una fase di competizione-sopravvivenza nella quale sopravvive con una probabilità data dalla condizione:

$$r < \frac{A(x)}{\bar{A}},$$

con \bar{A} fitness media e r numero casuale compreso tra 0 e 1.

2.3 Risultati

Abbiamo posto $L = 100$.

Per prima cosa abbiamo controllato che in assenza di degenerazione si ottenessero gli stessi risultati di Bagnoli e Bezzi. In assenza di competizione si osserva, come atteso, una distribuzione piccata in corrispondenza del massimo della fitness statica, dove l’altezza e la larghezza del picco (correlate a causa della popolazione costante) dipendono dalla scelta dei parametri μ , b ed r : aumentando il tasso di mutazione la distribuzione si allarga (e quindi si abbassa) a causa della maggior probabilità di popolare zone a più bassa fitness per effetto delle mutazioni; al contrario l’aumento dell’intensità della fitness statica e la diminuzione di r aumentano la pressione di selezione sui fenotipi più esterni, favorendo il popolamento di quelli più vicini alla sequenza maestra. Il risultato è indipendente dalla distribuzione iniziale, e in generale sono sufficienti un migliaio di generazioni per raggiungere la distribuzione asintotica. Nel caso limite in cui $r = \infty$ o, alternativamente, $b = 0$, ovvero di fitness statica piatta, la popolazione tende ad una distribuzione uniforme (partendo da una distribuzione iniziale a delta, una campana che piano piano

si allarga). In assenza di competizione il tempo necessario per l'equilibrio è molto grande e dipende dal tasso di mutazione (ad esempio circa 50,000 per $\mu = 0.1$ e circa 100,000 per $\mu = 0.01$). Introducendo una competizione (anche piccola), e quindi una pressione selettiva che stimoli lo sparpagliarsi della popolazione (oltre alla sola diffusione dovuta alle mutazioni), si osserva una notevole accelerazione del processo (ad esempio per $\mu = 0.1$, $J = 0.1$ si osserva una distribuzione generalmente uniforme già dopo 1,000 generazioni), con la presenza però di accumuli preceduti da avvallamenti agli estremi; ciò è dovuto al fatto che i fenotipi estremi hanno meno fenotipi vicini con cui competere, mentre quelli immediatamente adiacenti risentono della forte competizione degli estremi; questo effetto è tanto più evidente quanto più piccolo è il tasso di mutazione, essendo la larghezza (e quindi l'altezza) dei cluster determinata da μ , e, ovviamente, quanto più intensa è la competizione. Ci si avvicina alla distribuzione uniforme diminuendo il raggio di competizione; aumentandolo, invece, si osserva un graduale ripopolamento delle zone centrali dello spazio fenotipico, fino alla formazione per $R = 30$ (e $J > 0.2$) di tre quasispecie nettamente distinte. Per $R = 40$ la competizione provoca nuovamente lo spopolamento dei fenotipi centrali, e la seguente formazione di due sole quasispecie agli estremi.

Tornando al caso più generale di fitness statica non piatta, e introducendo la competizione, si osserva asintoticamente (circa 1,000 generazioni, indipendentemente dalla distribuzione iniziale) il fenomeno di speciazione riportato da Bagnoli e Bezzi, in cui due quasispecie periferiche si formano ai lati della quasispecie principale, ogni picco in corrispondenza di un massimo della fitness sempre con lo stesso valore (vedi figura 2; per dettagli vedi [4]). Lasciando invariati gli altri parametri, il fenomeno si verifica qualitativamente uguale per ogni valore del tasso di mutazione, purché si rimanga al di sotto di $\mu=0.1$ e si attenda un numero sufficiente di generazioni; ciò che cambia è, come sopra, la forma (altezza e larghezza) dei picchi.

Variando anche i parametri relativi alla fitness abbiamo cercato di far comparire cinque quasispecie. Si osserva che mantenendo b non troppo elevato ed r non troppo piccolo (in modo da non avere un'eccessiva pressione verso il fenotipo centrale), per μ piccoli e R dell'ordine (ma più grandi) di r , e con valori non troppo alti né troppo bassi di J , si ottiene una distribuzione della popolazione con cinque picchi non perfettamente separati (vedi figura 3). Possiamo spiegare queste osservazioni considerando il significato dei parametri R e μ : come abbiamo visto un tasso di mutazione minore corrisponde ad una minore larghezza della quasispecie, pertanto, considerando che la porzione di spazio fenotipico con fitness statica sufficientemente alta da permettere la coesistenza di più quasispecie rimane invariata (avendo lasciato invariati i parametri relativi ad H_0), il restringimento dei picchi do-

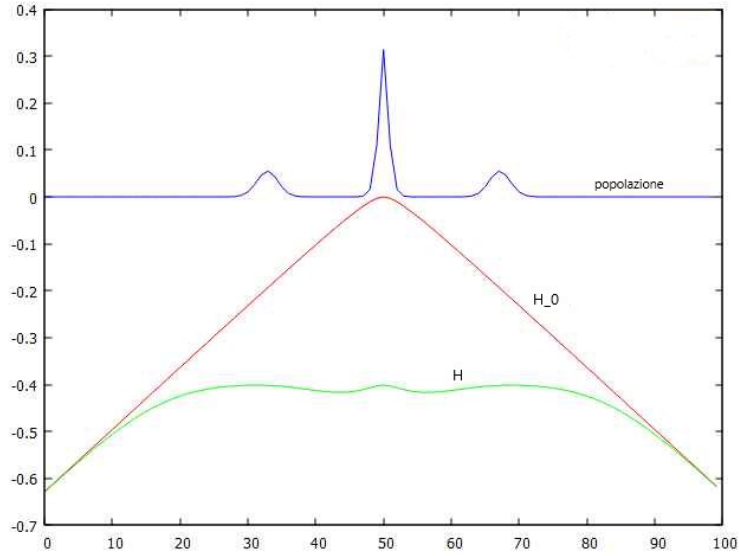


Figura 2: speciazione nel caso non degenerare con fitness statica linearmente decrescente e parametri $b = 0.04$, $r = 3$, $J = 0.6$, $R = 10$, $\mu = 0.01$, dopo $T = 1000$ generazioni.

vuto all'abbassamento del tasso di mutazione fa, per così dire, spazio alla formazione di nuove quasispecie, quasi permessa (evidentemente lo spazio non è sufficiente) dal piccolo raggio di competizione. Come ci si aspettava seguendo questa linea di ragionamento, aumentando l'intensità o diminuendo il raggio della fitness statica si osserva la diminuzione, fino alla scomparsa completa, dei picchi periferici; allo stesso modo l'incremento dell'intensità di competizione impedisce la formazione dei cluster tra le tre quasispecie.

In seguito abbiamo provato ad utilizzare la forma gaussiana della fitness statica nello stesso modello privo di degenerazione. A differenza della fitness usata in precedenza, che decresceva quasi linearmente lontano dal massimo, in questo caso la fitness decresce rapidamente fino a diventare praticamente piatta a distanza di circa $3r$ dal centro. Con questa forma della fitness, se r non è troppo grande, gli individui che si trovano a distanza sufficiente dalla sequenza maestra non risentono della pressione di selezione dovuta al contributo statico; di conseguenza tenderanno sempre a formarsi, per effetto della competizione, due quasispecie in corrispondenza dei fenotipi estremi, e in particolari condizioni (ad esempio per $r = 3$, $J = 0.6$, $R = 15$ e $\mu = 0.01$) osserveremo anche la formazione di due cluster intermedi non perfettamente separati da quelli estremi. Si osserva che l'unico modo per eliminare i picchi periferici pur mantenendo la competizione è quello di estendere il raggio di

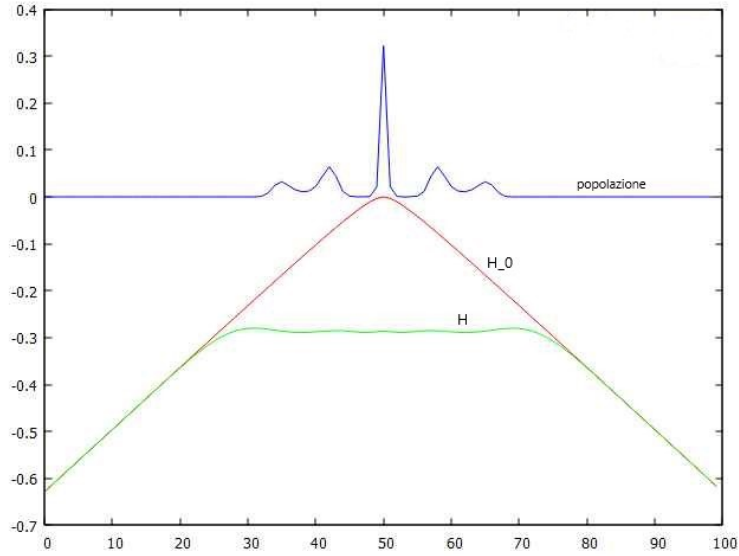


Figura 3: speciazione nel caso non degenerare con fitness statica linearmente decrescente e parametri $b = 0.04$, $r = 3$, $J = 0.6$, $R = 5$, $\mu = 0.0001$, dopo $T = 1000$ generazioni.

quest'ultima fino ad almeno $R = 70$, in modo tale da farli competere con la quasispecie centrale (allargare H_0 non è molto di aiuto in questo senso, dato che si va così verso un panorama quasi piatto). Naturalmente, non essendo la fitness statica lontano dal suo massimo perfettamente piatta, se si elimina la competizione avremo asintoticamente la formazione di un unico picco centrale. Si osserva infine che la distribuzione asintotica è indipendente da quella iniziale, pertanto anche una competizione poco intensa è sufficiente a popolare (dopo un numero sufficiente di generazioni) i fenotipi più lontani.

L'effetto della degenerazione è quello di aggiungere una pressione dovuta alle mutazioni verso i fenotipi centrali (a più alta degenerazione); questo effetto è particolarmente forte sui fenotipi estremi, dove lo 'sbilanciamento' del flusso di popolazione verso l'interno è maggiore. Finché manteniamo la forma linearmente decrescente della fitness i risultati non differiscono apprezzabilmente dal caso precedente. Anche con la fitness gaussiana i risultati sono molto simili al caso non degenerare: l'unica differenza evidente si osserva ai bordi dello spazio fenotipico, dove i picchi (se presenti) vengono spostati verso l'interno (vedi figura 4). In generale si osservano differenze qualitative (cioè nel numero di quasispecie) con il caso non degenerare solo per J molto basse. La differenza tra i due casi diventa palese con fitness statica piatta e assenza di competizione: in questo caso la distribuzione asintotica con

degenerazione è una campana centrata nella sequenza maestra.

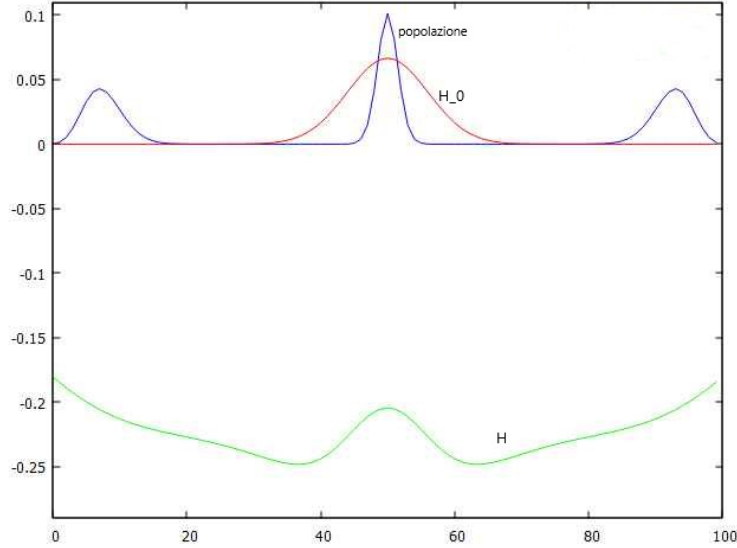


Figura 4: speciazione nel caso degenerare con fitness gaussiana e parametri $b = 0.04$, $r = 6$, $J = 0.6$, $R = 20$, $\mu = 0.01$, dopo $T = 1000$ generazioni.

Infine abbiamo esaminato i casi principali appena esposti utilizzando sia il modello macroscopico che quello microscopico, ponendo in entrambi $L = 20$ per semplicità computazionale. I risultati sono decisamente compatibili, anche se nel modello microscopico è in generale necessario un numero maggiore di generazioni per raggiungere l'equilibrio.

3 Riproduzione sessuata

In questa parte ci proponiamo di riprodurre i risultati ottenuti da Dieckmann e Doebeli [6] per speciazione simpatica nel caso di riproduzione sessuata con aploidia e in presenza di assortatività.

3.1 Modello

I modelli usati sono di tipo microscopico e a numero di individui N costante. Il genoma è rappresentato, a seconda del caso, da una, due o tre stringhe di valori booleani di lunghezza L_i , ognuna relativa a un diverso tratto fenotipico. Per il calcolo del fenotipo useremo la stessa regola additiva del caso asessuato. Chiameremo carattere *ecologico* il fenotipo da cui dipendono la disponibilità

di risorse e la competizione con gli altri membri della specie, in altre parole la fitness.

Non facciamo distinzioni di sesso, cioè ogni individuo può (nei limiti dell'assortatività) accoppiarsi con qualunque altro. Questa condizione è detta *ermafroditismo*, e può essere riscontrata in molte specie (come ad esempio le chioccioline) nelle quali il singolo organismo è dotato sia dell'apparato riproduttore maschile che di quello femminile. All'inizio di ogni fase di riproduzione uno dei due partner, preso casualmente, sceglie con chi accoppiarsi secondo la sua preferenza, data da quello che chiameremo carattere *assortativo*.

La riproduzione sarà schematizzata come una semplice ricombinazione casuale dei due genomi parentali: per ogni locus si sceglie un bit da uno dei due genomi con uguale probabilità. Anche in questo caso introduciamo un tasso di mutazione puntuale μ , probabilità per ciascun fenotipo che durante la riproduzione un gene scelto casualmente venga mutato. La fase di sopravvivenza è identica a quella del modello microscopico per la riproduzione asessuata.

Per la funzione di competizione consideriamo la stessa forma del modello per speciazione asessuata, mentre per la fitness statica seguiremo Dieckmann e Doebeli:

$$H_0(x) = b \exp \left(-\frac{|x - \frac{L}{2}|^2}{2r^2} \right),$$

con x fenotipo ecologico e b ed r che controllano intensità e pendenza.

Inizialmente consideriamo assenza di assortatività; in questo caso si ha un solo fenotipo (quello ecologico) e la scelta del partner è totalmente casuale. In seguito aggiungiamo al fenotipo ecologico quello assortativo, il quale determina in chi lo possiede la scelta del partner in base ad un 'punteggio' (*score*) dato da:

$$s_i^j = \left(f_A^i - \frac{L_A}{2} \right) |f_E^i - f_E^j|,$$

dove i indica l'individuo che sceglie, j il candidato alla scelta e f_A , f_E rispettivamente il fenotipo assortativo e quello ecologico, e L_A la lunghezza del genoma assortativo (si noti che sono presenti individui che sono attratti dal fenotipo ecologico opposto, individui che si accoppiano casualmente e individui che preferiscono il proprio fenotipo ecologico). Infine introduciamo un terzo carattere, che chiameremo *marker*, non correlato con quello ecologico. L'assortatività (sempre regolata dal carattere assortativo) dipende stavolta da un punteggio calcolato come nel caso precedente ma usando il carattere *marker* al posto di quello ecologico.

Abbiamo inizializzato i genomi in modo casuale ottenendo nello spazio fenotipico curve a campana centrate nella sequenza maestra. Variando la pro-

babilità abbiamo poi spostato la distribuzione iniziale, osservando gli effetti su quella asintotica.

3.2 Risultati

Come atteso, nel caso di accoppiamento casuale non si riesce ad osservare speciazione neanche per valori elevati della competizione e fitness statica quasi piatta (al massimo si ottiene una distribuzione bimodale). Ciò è dovuto al fatto che la ricombinazione casuale dei genotipi non permette lo spopolamento totale dei fenotipi compresi tra gli eventuali picchi, dato che la ricombinazione tra due individui con fenotipi distanti dà con buona probabilità un figlio con fenotipo intermedio. Persino nel caso limite di una fitness statica piatta con competizione intensa si continua ad ottenere una distribuzione a campana (anche se molto larga), per effetto della sola ricombinazione in uno spazio degenerare. Questi risultati sono in accordo anche con quanto trovato da Bagnoli e Guardiani [5].

Nel caso in cui si aumenti la pendenza della fitness statica e si intensifichi la competizione si osserva un fenomeno interessante: la distribuzione asintotica consiste in un'unica specie spostata dal centro dello spazio fenotipico verso un estremo (vedi figura 5). È possibile che questo effetto, assente nel caso delle specie asessuate, sia da ricondurre al fatto (già citato) che la ricombinazione casuale tende a ripopolare i genomi intermedi tra quelli parentali; per cui un eventuale sbilanciamento nella densità di popolazione da un lato della distribuzione, favorito da una competizione moderatamente intensa e da una fitness piccata, tenda ad amplificarsi fino ad eliminare un'estremità della distribuzione, lasciando solo un picco laterale. Questa ipotesi sembra confortata dal fatto che un aumento di J riduce notevolmente questo effetto fino quasi ad eliminarlo; si può pensare infatti che un'eccessiva competizione impedisca (o comunque renda più difficile) la formazione del disequilibrio iniziale tra un'estremità e l'altra della distribuzione, impedendo l'accumularsi di individui in un corrispondenza di pochi fenotipi. Inoltre un incremento di R sposta ulteriormente la distribuzione, come è prevedibile seguendo il ragionamento precedente. In ogni caso è evidente la dipendenza da piccole variazioni iniziali, dato che non esiste una direzione preferenziale verso cui si sposti il picco (che però si sposta ogni volta).

Se ora introduciamo l'assortatività basata sul carattere ecologico diventa possibile osservare fenomeni di speciazione. La speciazione in questo caso è spiegabile come conseguenza del fatto che, se la fitness statica non è troppo ripida, la concentrazione di individui nei pressi della sequenza maestra produce una 'buca' nella fitness totale, che favorisce la selezione dei fenotipi assortativi per la scelta dei partner simili, dato che gli individui con fenotipi

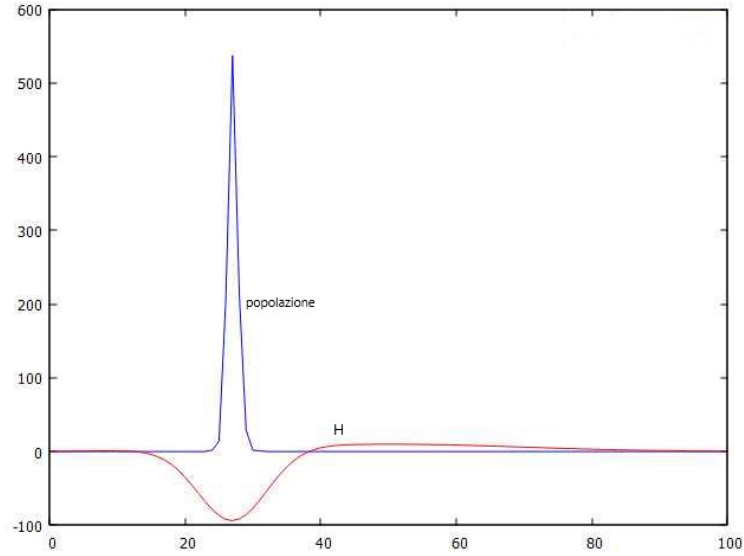


Figura 5: distribuzione della popolazione ($N = 1000$) nello spazio del fenotipo ecologico senza assortatività, con parametri $b = 10$, $r = 20$, $J = 0.1$, $R = 5$, $\mu = 0.01$, dopo $T = 10000$ generazioni.

ecologici intermedi (prodotti principalmente dall'accoppiamento di individui con fenotipi opposti) sono svantaggiati. A sua volta questa selezione favorisce il popolamento delle code della distribuzione, dando luogo al fenomeno.

Con assortatività basata sul carattere non ecologico si può osservare speciazione per parametri simili al caso ecologico (vedi figura 6). In questo caso il fenomeno può essere spiegato pensando che, nel momento in cui la distribuzione inizia a spopolarsi nella zona centrale per effetto del minimo della fitness, un (anche piccolo) sbilanciamento nella presenza di fenotipi marker di un certo tipo nella distribuzione nello spazio del fenotipo ecologico provochi il meccanismo di selezione del caso precedente; ma, essendo il fenotipo ecologico e quello marker correlati (anche se del tutto casualmente), la selezione sul secondo finirà per agire anche sul primo. Quello che succede, in pratica, è che l'azione combinata della selezione naturale e di quella sessuale basata su un carattere 'inutile' di per sé ma inizialmente correlato ad uno utile (per dettagli sulla selezione dei caratteri sessuali si veda [7], [8]) mette in moto un processo di divergenza che porta alla speciazione.

Come mostrato già da Higgs e Derrida [9] per il caso assortativo, la distribuzione finale è, in generale, influenzata dalla scelta dei genotipi iniziali (sistema *non ergodico*), anche con differenze nel numero di picchi che si osservano (ove presenti).

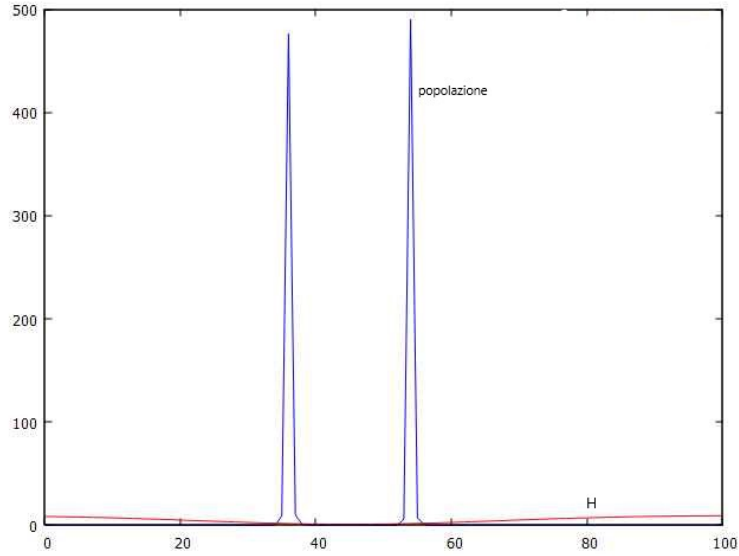


Figura 6: distribuzione della popolazione ($N = 1000$) nello spazio del fenotipo ecologico con assortatività basata su carattere non ecologico, con parametri $b = 10$, $r = 400$, $J = 0.01$, $R = 20$, $\mu = 0.01$, dopo $T = 10000$ generazioni.

4 Considerazioni finali

Abbiamo studiato il fenomeno della speciazione simpatrica per specie aploidi che si riproducono asessualmente, sia da un punto di vista deterministico prendendo come unità fondamentale la popolazione del fenotipo, che da un punto di vista probabilistico schematizzando il comportamento dei singoli membri della specie. I risultati ottenuti mostrano come, in un contesto di coesistenza, la competizione intraspecifica sia un elemento necessario per la formazione di nuove specie, e come la distribuzione delle risorse rispetto al fenotipo condizioni fortemente questo processo. L'ipotesi di degenerazione dello spazio fenotipico non sembra modificare sostanzialmente il risultato, rappresentando solo una leggera pressione selettiva verso la sequenza con degenerazione massima.

Abbiamo esaminato lo stesso fenomeno anche per specie sessuate, mettendo in evidenza come in questo caso l'assortatività sia una premessa indispensabile alla speciazione simpatrica, e mostrando come questa caratteristica non debba necessariamente essere legata a un fenotipo ecologico.

A Dettagli sul calcolo della degenerazione

A.1 Calcolo dei coefficienti di degenerazione

Data la distribuzione di degenerazione possiamo scegliere di calcolare due coefficienti per ogni fenotipo. Il primo coefficiente è relativo alle mutazioni verso il fenotipo più vicino alla coda della distribuzione di degenerazione, l'altro a quelle verso il fenotipo più vicino al picco. Ricordando l'effetto della degenerazione sull'evoluzione del sistema, definiamo (in modo arbitrario: avremmo potuto considerare regole più complicate e accurate) il primo coefficiente:

$$c_x^1 = \frac{d(x)}{D}, \quad \text{(manca un fattore 2 al denominatore)}$$

dove $d(x)$ è la degenerazione del fenotipo x e D la degenerazione massima. Dovendo essere la probabilità totale di mutazione uguale a μ , l'altro coefficiente è dato da:

$$c_x^2 = 1 - c_x^1.$$

Fanno naturalmente eccezione a questa regola il primo e l'ultimo fenotipo, per cui:

$$\begin{aligned} c_x^1 &= 0, \\ c_x^2 &= 1. \end{aligned}$$

A.2 Metodo approssimato per calcolo della degenerazione

Data la schematizzazione del genoma, il suo rapporto con il fenotipo e la degenerazione massima, possiamo utilizzare un metodo statistico per stimare la distribuzione di degenerazione. Lo abbiamo usato solo per il caso già analizzato di stringa di valori booleani, ma l'implementazione in casi più complicati non dovrebbe presentare grosse difficoltà.

Il procedimento è semplice. Per prima cosa si sceglie un campione di N individui ognuno caratterizzato dal proprio genoma; naturalmente più il campione è vasto e più il metodo è preciso. In seguito si assegna ad ogni bit di ogni genoma un valore booleano casuale con probabilità uguale. Infine si calcola il fenotipo di ogni individuo sommando tutti i bit del relativo genoma, e si costruisce la distribuzione della popolazione nello spazio fenotipico. A questo punto possiamo stimare la degenerazione di ogni fenotipo come l'arrotondamento all'intero più vicino (ma in ogni caso maggiore di zero) della quantità:

$$d(x) = \frac{n(x)}{n_{max}} D,$$

dove $n(x)$ è il numero di individui con fenotipo x , n_{max} il numero di individui con il fenotipo più popolato e D la degenerazione massima.

Almeno nel caso di stringhe binarie il metodo sembra funzionare piuttosto bene.

Può risultare utile per genomi lunghi, schematizzazioni complicate, e in genere ogni volta che il calcolo esatto della degenerazione risulti troppo laborioso (a patto che si riesca a calcolare o stimare con ottima approssimazione la degenerazione massima).

Riferimenti bibliografici

- [1] Richard Dawkins. *Il fenotipo esteso. Il gene come unità di selezione*. Zanichelli, 1986.
- [2] Richard Dawkins. *Il gene egoista*. Zanichelli, 1979.
- [3] M. Higashi, G. Takimoto, N. Yamamura. *Letters to Nature*, **523**. 1999.
- [4] Franco Bagnoli, Michele Bezzi. *Physical Review Letters*, **3302**. 1997.
- [5] Franco Bagnoli, Carlo Guardiani. *Lecture Notes in Computer Science*. 2004.
- [6] Ulf Dieckmann, Michael Doebeli. *Letters to Nature*, **354**. 1999.
- [7] Ronald A. Fisher. *The evolution of sexual preference*. 1915.
- [8] Andrew Pomiankowski, Yoh Iwasa. *Proceedings of the National Academy of Sciences*, **5106**. 1998.
- [9] Paul G. Higgs, Bernard Derrida. *Journal of Physics*, **L985**. 1991.