



MODELS MATEMÀTICS I DE SIMULACIÓ DEL POLIMORFISME CROMOSÒMIC

PRODUÏT PER INVERSIONS

Memòria presentada per
optar al grau de Doctor
en Biologia per la Univer-
sitat de Barcelona per
en

JORDI OCAÑA I REBULL

Vist i Plau
El Director de la Tesi

Dr. ANTONI PREVOSTI I PELEGRIN
Catedràtic de Genètica
Director del Departament de Genètica
Facultat de Biologia
Universitat de Barcelona

Barcelona, Setembre de 1981.



4. MATERIALS I MÈTODES

4. MATERIALS I METODES

4.1. Conceptes generals. Simulació.

En aquest apartat descriurem breument els conceptes generals relacionats amb la simulació de sistemes genètics.

La simulació d'un sistema es podria definir com la representació del seu comportament i propietats mitjançant el funcionament o l'evolució d'un altre. Perquè la simulació tingui sentit cal que es puguin establir certes correspondències entre els elements, relacions entre aquest elements i processos respectius, de manera que el sistema simulador sigui una representació suficientment acurada del sistema simulat.

Hi poden haver moltes classes de simulacions d'un mateix sistema. Una de les més útils i usuals és la construcció de models lògico-simbòlics tals que es puguin representar mitjançant un programa per ordinador, de manera que les variables del programa i relacions lògiques entre elles, representin els elements del sistema real i les seves relacions.

Si seguim les idees de Shah (1973), podríem dir que hi ha tres punts essencials que cal remarcar per classificar un model de simulació a Genètica de poblacions:

a) La unitat d'estudi a Genètica de poblacions és la població o grup. Aquesta població, però, és caracteritzada per l'estat d'unitats més bàsiques que la componen, normalment individus

o gamets (per exemple quan es suposa que hi ha aparellament aleatori). Quan els models caracteritzen les poblacions per les freqüències, proporcions, etc. de pertinença d'individus a determinades classes d'estats i quan el canvi en el temps és expressat directament pels canvis a les proporcions de repartició dels individus en aquestes classes, parlarem de macromodels. Quan el model caracteritzi la població per l'estat d'una sèrie d'individus, amb l'estat de cada un d'ells enregistrat separatament, i quan el canvi al llarg del temps depengui d'esdeveniments que afectin a aquests individus "individualment" (!), com ara aparellament, mort, ..., parlarem de micromodels. Un macromodel típic és el definit per les equacions de canvi de la freqüència gènica p d'un al·lel A d'un locus dial·lèlic, a una població infinita i amb aparellament a l'atzar

$$p' = \frac{wp^2 + p(1-p)}{wp^2 + 2p(1-p) + v(1-p)^2}$$

on w , l , v són les eficàcies dels genotips A/A , A/a i a/a , allò que caracteritza la població és el canvi de la freqüència gènica p , o proporció relativa de gamets que pertanyen a la classe de gamets portadors de l'al·lel A a aquell locus. Uns micromodels típics són els de Fraser i Burnell als quals la població es caracteritza mitjançant un arranjament "d'individus" caracteritzats per dues tires de dígit binaris que codifiquen el seu genotip, amb cada tira corresponent a un dels "gamets" que produïren aquell individu. El canvi de la població es basa en esde-

veniments que afecten separatament a aquests individus, com ara ésser "escollits" com a "progenitors" d'un nou individu.

b) La variable temps pot ser tractada d'una manera discreta o contínua. Considerarem el temps discret quan l'ocurrència dels esdeveniments que canvien l'estat de la població només es pot donar en uns instants determinats i és impossible que passi res entre dos instants d'aquests consecutius. Considerarem el temps continu quan l'ocurrència dels esdeveniments pot ser en qualsevol punt d'un interval de temps. Tractarem bastant adequadament el temps com una variable discreta quan no hi hagi coexistència de generacions de manera que els fets essencials a la substitució d'una generació per l'altra s'esdevinguin quasi simultàniament (l'interval de temps entre ells sigui despreiable comparat amb l'interval de temps sense que no passi res). Altres vegades la consideració discreta del temps serà associada purament a artificis tècnics dels models, com als models als quals tot el canvi produït durant un interval de temps predeterminat t es resumeix en una matriu de transició M_t que permet avaluar les proporcions d'individus a les diferents classes, solament als instants t_0 , $t_0 + t$, $t_0 + 2t$, etc... Malgrat que realment la població canvia contínuament, tot el canvi queda summaritzat als dits instants mitjançant les corresponents operacions algebraïques.

c) Finalment, un model pot ser estocàstic o determinista. Al darrer cas suposem que la població evoluciona exactament segons

els valors esperats. Aquesta suposició és bàsicament correcta, com a conseqüència de les lleis dels grans nombres, quan la gran dària efectiva de la població és molt gran. Quan l'evolució de la població depèn de l'ocurrència d'esdeveniments que tenen determinada probabilitat però que poden donar-se o no a cada cas individual, parlarem de models estocàstics. Els models de Lotka i Volterra són un exemple clàssic de model determinista, mentre que els models de simulació de Rico (1975) són un cas de model estocàstic. Els primers permeten descriure l'evolució de la població mitjançant equacions diferencials amb les quals, donat l'estat a un instant, es pot avaluar determinísticament l'estat a qualsevol altre instant. Els segons són estocàstics car a cada instant l'evolució del sistema depèn de l'estat actual però també de tries aleatòries (segons les probabilitats definides per l'esmentat estat del sistema), per exemple dels esdeveniments concrets de recombinació que hi hagi a la formació d'un gamet.

4.2. Classificació dels models d'aquesta memòria. Màquines utilitzades.

El primer dels models proposat i estudiat a aquesta memòria (secció 5) es pot qualificar de macromodel amb temps discret i determinista, macromodel ja que allò que caracteritza a la població i es transforma algebràicament d'una generació a l'altra són les freqüències de les diferents classes de gamets (classes de gamets segons els al·lels que presenten L loci, amb $L=5$ a les simulacions realitzades), amb temps discret ja que

solament es considera l'estat del sistema en certs instants (generacions) i el pas d'una generació a la següent es realitza instantàniament des del punt de vista lògic del model (al temps entre generacions "no passa res") malgrat que els càlculs reals puguin ser molt llargs i, finalment, determinista ja que es considera la població molt gran i per tant despreciable la variació aleatòria causada per la deriva, de manera que l'evolució es descriu mitjançant el canvi esperat de les freqüències gamètiques.

A la mateixa secció 5 indiquem el mètode seguit per transformar el model anterior en un macromodel -temps discret- estocàstic, útil per estudiar l'efecte d'un nombre finit d'individus adults.

Els aspectes lògics del model anterior estan exposats detalladament a la secció 5. A l'apèndix 4 descrivim el programa utilitzat per fer les simulacions, les quals han estat realitzades amb dos microordinadors Commodore, model P.E.T. 2001 de 8 i 24 kbytes de memòria i amb un microordinador C.B.M. sèrie 3000, de la mateixa marca, de 32 kbytes.

Totes les simulacions realitzades han representat unes 8000 hores de treball d'aquestes màquines.

El model de simulació de la secció 6 es pot qualificar de micromodel -temps discret- estocàstic. És un micromodel ja que, tal com s'indica a l'apèndix 5, l'evolució del sistema és deter-

minada pels esdeveniments que afecten individualment a unes entitats (individus) representades cada una d'elles per una parella de tires de dígit binaris que codifiquen el seu genotip respecte dels L loci estudiats (16 o 8 segons el cas). El temps es considera discret ja que la variable interna dels programes que mesura el temps (temps lògic, del model, no el temps real que triguen les simulacions) només pot prendre valors enters positius, corresponents al número de la generació que s'està produint, en un moment donat, a partir de l'anterior generació. La producció d'una generació implica tot una sèrie d'esdeveniments (aparellaments, gametogènesis, dispersió de zigots, competicions, etc...) però dins la lògica dels programes aquests esdeveniments ocorren intemporalment en el sentit que no tenen un temps d'ocurrència representat internament als programes, són purament una sèrie de manipulacions que el programa fa seqüencialment per tal de substituir una generació (més tires de dígit) per la següent. A Ocaña, Alonso i Ruíz de Villa (1978) es discuteixen les característiques que hauria de tenir un sistema de simulació genètica amb temps continu. Finalment, el model es pot considerar estocàstic ja que l'evolució del sistema és determinada per l'ocurrència d'esdeveniments que tenen certa probabilitat, segons l'estat de la població, però que poden donar-se o no, aleatòriament, de manera que la trajectòria del sistema és una de les moltes branques possibles d'un procés de ramificació, i canvia d'una rèplica a l'altra.

Una descripció més detallada de model és a l'apèndix 5, ho hem fet així a causa del caràcter sumament tècnic i de l'extensió d'aquesta descripció.

Les simulacions amb aquest model s'han realitzat a l'ordinador IBM/360-30 del Laboratori de Càlcul de la Universitat de Barcelona. L'antiguetat d'aquest sistema i al mateix temps l'abundant documentació existent fan innecessari descriure'l amb més detall.

Per cloure aquesta secció farem uns breus comentaris sobre els avantatges i inconvenients associats als dos tipus de models estudiats aquesta memòria: els micromodels i els macromodels els quals, deixant de banda el temps i el caràcter aleatori o determinista són les dues classes de models més freqüents a la literatura.

En el primer tipus de simulació no cal posseir un coneixement gaire complet de la dinàmica matemàtica de les poblacions, sinó dels mecanismes genètics i biològics subjacents (recombinació, aparellament, ...) per tal de poder representar-los informàticament. El segon tipus requereix un coneixement bastant concret de les equacions que regeixen la dinàmica del sistema; de fet, les dites simulacions en general es redueixen a problemes de càlcul numèric, per estudiar el comportament d'aquestes equacions quan no és possible fer-ho per mètodes analítics.

Amb un micromodel per temps i espai d'ordinador, la principal restricció resideix en el nombre d'individus de la població, rarament en el nombre de caràcters. Amb un macromodel, donada la representació de la població, el nombre d'individus és irrelevant, en canvi el nombre de caràcters pot arribar a ésser prohibitiu. Això es pot copsar fàcilment amb un exemple: si hom vol estudiar una població amb 50 individus i 100 caràcters (loci, per concretar) per individu, amb dues alternatives o al·lels per caràcter, en el primer cas caldrà usar una variable de dimensió $100 \times 50 = 5000$ (o 10.000 per diploïds), mentre que en el segon caldrà definir una variable de dimensió 2^{100} si es volen considerar tots els genotips possibles en el cas haploïde, o $2^{100} \times (2^{100} + 1)/2$ en el cas diploïde, dimensions fora de la capacitat dels ordinadors més grans.

Des d'un punt de vista purament tècnic-informàtic, serà recomanable usar el primer tipus de simulació amb poblacions petites i bastants caràcters, i el segon amb poblacions grans (infinites, de manera que sigui més fàcil definir les equacions de la dinàmica del sistema) definides per pocs caràcters. Naturalment hi haurà altres limitacions o interessos d'ordre genètic que puguin influir en l'adopció del mètode, o en l'establiment d'una estratègia "mixta", més o menys decantada cap a un dels extrems.

Els nombrosos micromodels de simulació proposats a la literatura genètica tenen característiques de vegades contraposades.

Es convenient repassarles breument per comparar-los amb els d'aquesta memòria.

En tots els treballs de simulació que hem citat es representa a cada individu mitjançant una paraula de díigits binaris, simple o doble segons sigui haploïde o diploïde. Això vol dir que tots els loci es consideren dial·lèlics.

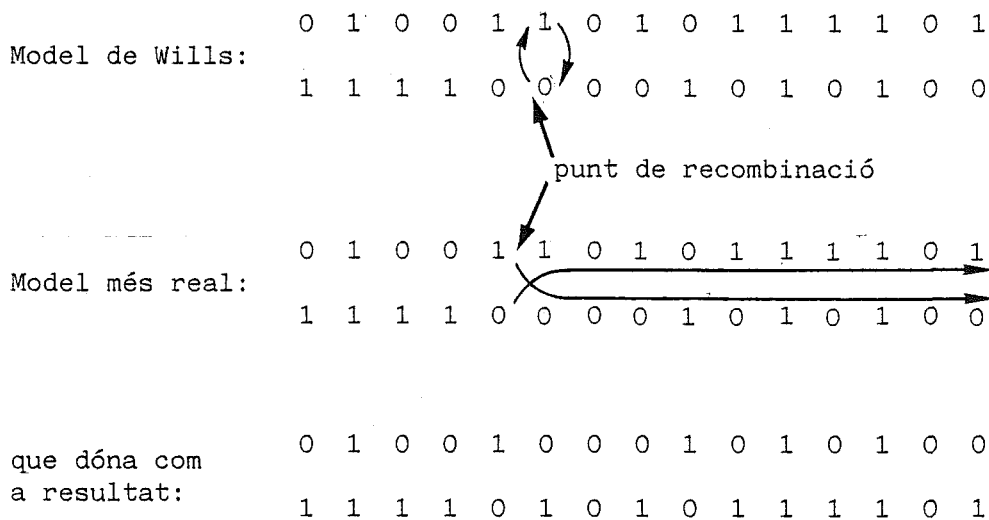
Tots els models consten d'una etapa de selecció, que difereix de l'un a l'altre. Sved (1968) i Fraser, Burnell i Miller (1966) empen un algorisme iteratiu que sembla molt costós quant a temps: es selecciona un pare a l'atzar, que produeix un gamet seleccionant a l'atzar un dels cromosomes (equivalents a cada una de les paraules binàries, perquè només es considera un cromosoma) amb possibilitat de recombinació, es selecciona un segon pare, que produeix un altre gamet, i es forma un individu, l'eficàcia del qual es calcula mitjançant una funció multiplicativa del grau d'heterosi. Aquest individu s'accepta o es rebutja com a membre de la següent generació si la seva eficàcia supera o no cert mínim. Es reparteix el procés fins a completar el nombre original d'individus. Wills i col·laboradors eliminen un determinat percentatge d'individus de la població, corresponent als menys eficaços segons el seu criteri (els més homozigòtics), substituïnt-los per rèpliques de la resta (agafats a l'atzar o agafant els més eficaços per replicar-los). Aleshores es produeixen gamets i es forma una nova població amb aparellament a l'atzar.

Cap d'aquests mecanismes de selecció sembla prou flexible per simular els processos reals de selecció: la manera d'escollir els individus, eliminant els que tinguin una eficàcia per sota de cert valor, converteix l'improbable en l'impossible, tal com fa notar Crosby (1973). En aquest aspecte és superior el model de Rico (1975), el qual, a més de donar compte de la majoria de tipus de selecció artificial, es pot adaptar fàcilment a l'anàlisi de la selecció natural, associant els individus de la població unes probabilitats, dependents del genotipus o del fenotipus, de participar en la formació de la generació següent. Un cop fet això, es repliquen els individus segons aquestes probabilitats (com si es prengué una mostra aleatòria) i amb la nova població obtinguda es produeixen gamets. El model de Rico (1975) també supera els altres en el fet que la selecció s'exerceix sobre els fenotips dels individus, considerats equivalents als valors genotípics més una perturbació aleatòria ambiental, mentre que en els altres models esmentats es selecciona directament el valor del genotip, avaluat segons una funció determinista (com ara grau d'heterozigosi, nombre d'al·lels dominants, etc...).

Fraser i Burnell (1970) i Crosby (1973) fan sengles discussions generals dels tipus possibles de selecció en simulació genètica finita.

La gametogènesi i sobre tot la recombinació són un altre punt a discutir en els models de simulació. Wills i col·labora-

dors simulen la recombinació "bombardejant" a l'atzar el cromosoma. En els loci "tocats" s'intercanvien els dígit. Això no correspon realment a un sol entrecreuament, sinó a un doble entrecreuament que inclogui aquell locus: si es volgués representar la recombinació amb més realisme caldria que, a partir del punt en què s'ha produït l'entrecuament, s'intercanviessin tots els dígit, fins a trobar, eventualment, un altre intercanvi. Gràficament es podria indicar:



Wills i col.laboradors preveuen la possibilitat de entrecreuaments múltiples (que pel que s'ha comentat serien dobles entrecreuaments múltiples), fent que la variable aleatoria "nombre d'entrecreuaments per cromosoma" segueixi aproximadament una distribució de Poisson. Això equival a no considerar cap mena d'interacció entre quiasmes, suposició simplificadora que no sembla gaire certa, segons mostren nombroses dades experimentals (recopilades fins l'any 1975 a Sybenga (1975)), que fan

pensar en l'existència de fenòmens d'agregació i de rebuig, sembla que aquesta variable segueixi una distribució binomial negativa o alguna de semblant. Cap dels models que hem repassat inclou la interacció entre quiasmes, d'anàlisi teòrica sumament complexa (Lewontin (1964 a), Shikata (1964)).

A Fraser i Burnell (1970), Crosby (1973), i Rico (1975) tampoc no es considera la interacció entre quiasmes, encara que la recombinació es simula en el sentit més realista especificat a l'inici d'aquests paràgrafs.

En tots els models de simulació es pren directament com a vector de probabilitats de recombinació el de les distàncies genètiques entre loci, perquè l'intercanvi es realitza entre els dos cromosomes homòlegs de cada individu, sense duplicació de cromàtides i intercanvi entre dues mentre les altres resten intactes.