

**COOPERAÇÃO E CONFLITO EM MODELOS DE VESÍCULAS  
PRÉ-BIÓTICAS**

**Daniel Gomes Marques Silvestre**

Dissertação apresentada ao Instituto de Física de São Carlos, Universidade de São Paulo, para a obtenção do título de Mestre em Ciências, Área de Concentração: Física Aplicada - Biomolecular.

SÃO CARLOS  
São Paulo - Brasil  
Março - 2006

**COOPERAÇÃO E CONFLITO EM MODELOS DE VESÍCULAS  
PRÉ-BIÓTICAS**

**Daniel Gomes Marques Silvestre**

Orientador: Prof. Dr. José Fernando Fontanari

Dissertação apresentada ao Instituto de Física de São Carlos, Universidade de São Paulo, para a obtenção do título de Mestre em Ciências, Área de Concentração: Física Aplicada - Biomolecular.

SÃO CARLOS  
São Paulo - Brasil  
Março - 2006

*Às minhas mulheres,  
Tatiana, Rita e Lúcia.*

## AGRADECIMENTOS

Em primeiro lugar, eu tenho que agradecer à Evolução e a Natureza. Afinal, sem elas eu nem estaria aqui escrevendo! Aliás, essa dissertação nada mais que uma forma de demonstrar meu fascínio por ambas.

Agradeço, em especial, ao prof. Fontanari, por ter visto algum potencial em mim e por ter dado a oportunidade de desenvolvê-lo. Essa é sem dúvida, a melhor forma de consideração que existe. Meu muito obrigado a você !!! Ao Colato, que mesmo sem nada nas mangas salvou o meu dia incontáveis vezes. Giovano, companheiro em todas as discussões e divagações. Sem ele, muito do que será discutido aqui não teria amadurecido o suficiente. E ao Léo, por ter me ensinado os primeiros passos nas formalidades estocásticas. A minha primeira simulação computacional é culpa totalmente sua !!! Ao pessoal do IFSC, Cris, Cristiane, Samira, Wla, companheiras da biblioteca e ao pessoal da padaria Guanabara. Sem esse suporte, tudo seria muito mais difícil.

Agora, o momento família. A minha nova família, Cyra, Rabino, Vivi, Maria Marta, Su, Lu, Helô, Cremilda e Cia.. Nos melhores momentos da minha vida, vocês estavam lá. A minha velha família! Pai, mãe, Corujo, Narizinho, vó Maria, vó Dirceu e o povo da Chácara Flora. Essa dissertação é a minha demonstração de gratidão a todos vocês. Esse é mais um passo naquele sonho de menino, de ser cientista, que todos vocês ajudaram a construir. Finalmente, estou conseguindo. Valeu a pena !!! A minha família. Aquela que vive bem pertinho. Eu sei que já dediquei esta dissertação a vocês. Vocês merecem mais !!! Tati, Ritinha e Lúcia, as mulheres na minha vida. É só alegria agora ! Também, Carol e Fernando que, *pero si, pero no*, agora são da família também.

---

Este trabalho contou com auxílio financeiro do  
CNPq, proc. 131640/2004-6.



# SUMÁRIO

<b>LISTA DE FIGURAS</b>	<b>vii</b>
<b>RESUMO</b>	<b>x</b>
<b>SUMMARY</b>	<b>xi</b>
<b>1 INTRODUÇÃO</b>	<b>1</b>
1.1 Um breve histórico . . . . .	1
1.2 A crise informacional na origem da vida . . . . .	3
1.3 Algumas (possíveis) soluções . . . . .	5
1.3.1 Hiperciclos . . . . .	6
1.3.2 Corretor Estocástico . . . . .	8
1.3.3 GARD . . . . .	9
1.4 Cooperação: um ingrediente essencial . . . . .	10
1.5 O problema da coexistência . . . . .	12
<b>2 DINÂMICA DE VESÍCULAS I</b>	<b>14</b>
2.1 Um modelo determinístico . . . . .	14
2.2 Replicação vesicular . . . . .	15
2.3 Replicação gênica . . . . .	17
2.4 Dinâmica geral . . . . .	18
2.5 Explorando o modelo . . . . .	19
2.5.1 Vesículas independentes . . . . .	20
2.5.2 Vesículas com capacidade infinita . . . . .	21
2.5.3 Coexistência I - a ação da deriva . . . . .	23

---

2.5.4	Coexistência II - deriva e competição . . . . .	24
2.6	Conclusões . . . . .	30
<b>3</b>	<b>DINÂMICA DE VESÍCULAS II</b>	<b>36</b>
3.1	Um modelo computacional . . . . .	36
3.2	Compartimentos: o modelo de Bresch <i>et al.</i> . . . . .	37
3.3	Delineando a simulação . . . . .	39
3.4	Coexistência I - a ação da deriva . . . . .	42
3.5	Coexistência II - o efeito da mutação . . . . .	48
3.6	Conclusões . . . . .	52
<b>4</b>	<b>CONCLUSÕES GERAIS</b>	<b>56</b>
4.1	A emergência da deriva . . . . .	56
4.2	Exclusão sem competição . . . . .	58
4.3	Uma perspectiva . . . . .	59
<b>5</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>60</b>

## LISTA DE FIGURAS

- 1 Freqüências estacionárias das vesículas contendo o tipo menos competitivo em função da intensidade de competição para capacidades  $N = 6, 12$  e  $24$  no caso de vesículas independentes. As linhas sólidas são resultado exatos das equações de recorrência e as linhas pontilhadas são resultado da aproximação difusiva. . . . . 22
- 2 Fração das vesículas com os dois tipos em função da capacidade das vesículas para  $\tau = 0$  (i.e., ambos tipos são competidores equivalentes),  $g = 0.1$  e  $m$  como indicado no gráfico. . . . . 24
- 3 Valor crítico da capacidade das vesículas em função da razão entre os tempos de geração para  $\tau = 0$  e (da esquerda para direita)  $g = 0.1, 0.2, \dots, 0.9$ . Abaixo de  $N_c$ , a fixação de um dos tipos via deriva genética impede a coexistência. . . . . 25
- 4 Paisagem de deriva - a superfície denota o limiar de coexistência - acima desta a coexistência é impossível. A forma da superfície é exclusivamente devida à deriva já que neste caso temos  $g \rightarrow 1$  e  $\tau = 0$ . As curvas de nível projetadas na base do gráfico são para  $\Omega = 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8$ . Os pontos foram obtidos ao término da replicação dos genes, como descrito no modelo. . . . . 26
- 5 Fração de vesículas com os dois tipos de genes em função da capacidade das vesículas para  $\tau = 0.1$ ,  $g = 0.35$  e  $m$  como indicado no gráfico. . . . . 27

---

6	Paisagem de deriva deformada pela competição - a superfície denota o limiar de coexistência - acima desta a coexistência é impossível. Compare a forma da superfície com a da figura 4. Os parâmetros são $g \rightarrow 1$ e $\tau = 0.10$ . As curvas de nível projetadas na base do gráfico são para $\Omega = 0.1, 0.2, 0.4, 0.6$ . Os pontos foram obtidos ao término da replicação dos genes, como descrito no modelo. . . . .	28
7	Valor crítico da capacidade das vesículas em função da razão entre os tempos de geração para $\tau = 0.1$ e $g$ como indicado no gráfico. Para valores fixos de $m$ os dois valores de $N_c$ determinam o intervalo de capacidade que permite a coexistência. . . . .	29
8	Valor crítico da capacidade das vesículas em função da razão entre os tempos de geração para $g = 0.8$ e $\tau$ como indicado no gráfico. . . . .	30
9	Valor crítico da razão entre os tempos de geração em função da pressão de cooperação para $\tau$ como indicado no gráfico. Abaixo de $m_c$ a coexistência é impossível. . . . .	31
10	Valor crítico da intensidade de competição em função da pressão de cooperação para $m = 1$ (principal) e $m = 5$ (inset). A capacidade das vesículas está indicada no gráfico principal. Abaixo de $\tau_c$ a coexistência é impossível. . . . .	32
11	Colina de coexistência - superfície crítica de coexistência - acima desta a coexistência é impossível. As curvas de nível projetadas na base do gráfico são para $\Omega = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5$ (da esquerda para a direita) com $N = 5$ e $m = 1$ . . . . .	33
12	Penhasco de coexistência - superfície crítica de coexistência - acima desta a coexistência é impossível. As curvas de nível projetadas na base do gráfico são para $\Omega = 0.1, 0.2, 0.4, 0.6, 0.8$ (da esquerda para a direita) com $N = 100$ e $m = 1$ . . . . .	34
13	Valor crítico da processividade ( $\Lambda_c$ ) em relação a diversidade de genes ( $d$ ). A reta representa a função $\Lambda_c(d) = Ad^\lambda$ , com $A \sim 0.5$ e $\lambda \sim 2$ . Abaixo da reta, a coexistência é impossível. . . . .	44

---

14	<b>Variação dos expoentes de crescimento populacional como função da processividade <math>\Lambda</math> em populações livres de mutação para <math>d = 4, \dots, 10</math> da esquerda para a direita. A linha zero marca a transição de populações decrescentes para populações crescentes. . . . .</b>	45
15	<b>Densidade populacional esperada em uma linhagem com <math>d = 4, u = 0</math> e <math>\Lambda = 4, 5, \dots, 10</math>, de baixo para a cima. . . . .</b>	46
16	<b>Densidade populacional esperada em uma linhagem com <math>d = 8, u = 0</math> e <math>\Lambda = 8, 12, 16, \dots, 42</math>, de baixo para cima. . . . .</b>	47
17	<b>Probabilidade de sobrevivência para uma linhagem com <math>d = 4, u = 0</math> e <math>\Lambda = 4, 5, \dots, 10</math>, de baixo para cima. . . . .</b>	48
18	<b>Probabilidade de sobrevivência para uma linhagem com <math>d = 8, u = 0</math> e <math>\Lambda = 8, 12, \dots, 40, 42</math>, de baixo para a cima. . . . .</b>	49
19	<b>Densidade populacional esperada de para uma linhagem com <math>d = 2, \Lambda = 2</math> e <math>u = 0.00, 0.05, \dots, 0.50</math>, de cima para baixo. . . . .</b>	50
20	<b>Densidade populacional esperada de para uma linhagem com <math>d = 4, \Lambda = 4, 5, \dots, 14</math> e <math>u = 0.10</math>, de baixo para a cima. . . . .</b>	51
21	<b>Probabilidade de sobrevivência para uma linhagem com <math>d = 2, \Lambda = 2</math> e <math>u = 0.00, 0.05, \dots, 0.50</math>, de cima para baixo. . . . .</b>	52
22	<b>Evolução temporal da fração esperada de genes não-funcionais por vesícula em uma linhagem com <math>d = 2, \Lambda = 2, u = 0.00, 0.05, \dots, 0.50</math>, de baixo para cima. . . . .</b>	53
23	<b>Evolução temporal da fração esperada de um dado gene funcional em uma linhagem com <math>d = 2, \Lambda = 2, u = 0.00, 0.05, \dots, 0.50</math> de cima para baixo. . . . .</b>	54

## RESUMO

A crise primordial da informação genética tal como delineada pelo modelo de quasi-espécies tem permanecido há mais de três décadas como um desafio as teorias de origem da vida e evolução pré-biótica. Apesar das diversas soluções propostas ao longo desse período, não há ainda uma que seja plenamente aceita pela comunidade científica. Dentro dessa gama de soluções, os modelos que usam uma abordagem de seleção em múltiplos níveis têm se destacado como uma alternativa ao paradigma da área, o celebrado modelo de hiperciclos. Denominados modelos de vesículas pré-bióticas, entre os quais destaca-se o corretor estocástico, estes assumem a cooperação, de forma mais ou menos explícita, como ingrediente essencial para vencer a crise da informação genética. A cooperação tem sido vista como capaz de proporcionar um modo de contornar a crise suscitada por Eigen ao permitir a divisão da informação essencial em diversas partes, todas abaixo do limiar da catástrofe de erro. Entretanto, até muito recentemente, havia pouca literatura especializada explorando esta classe de modelos, especialmente no que tange às questões de coexistência. O objetivo desta dissertação é estudar mais profundamente alguns modelos deste tipo dando ênfase a este tipo de problema. Inicialmente, vamos explorar um modelo determinístico de seleção de grupo similar ao corretor estocástico, baseada no famoso *Wright's Island Model* com algumas modificações. Em especial, relaxamos a necessidade de tempos de geração iguais nos dois níveis da dinâmica. Além deste, também vamos estudar uma versão menos restritiva do modelo de compartimentos. Nos dois casos, a ideia geral é buscar as condições que garantam a coexistência dos diversos tipos de genes envolvidos.

## SUMMARY

The primordial genetic information crisis as defined by the Eigen's quasispecies model, which can be used as a paradigm here, has been a challenge to any theory about the origin of life and prebiotic evolution for more than three decades. Despite several tentative solutions proposed along this period, there's no consensual solution to the scientific community. Among that diversity of solutions, models that use a multilevel selection approach have been seen as a viable alternative to the area's paradigm, the hypercycles model. Those called prebiotic vesicle models, whose principal example is the stochastic corrector model, assume cooperation, in a more or less explicit fashion, as an essential ingredient to solve the genetic crisis. Cooperation has been seen as capable of providing means to bypass the genetic crisis by dividing the essential information into several pieces, each one below the error threshold. Until recently, no complete treatment for such a model could be found in the specialized literature, specially in relation to replicators coexistence questions. So, the main purpose of this dissertation is the thorough exploration of some of those models with emphasis in coexistence aspects. Initially, we will explore a deterministic group selection model based in the celebrated Wright's Island Model with some modifications. We relax some of the assumptions of the model to accommodate distinct time scales between the different levels of the model. Besides this model, we will study a less restrictive version of a compartment model. In both cases, we give special emphasis in the search of coexistence conditions that guarantee the survival of the different templates.

# 1 INTRODUÇÃO

*“If the Lord Almighty had consulted me before embarking on creation I should have recommended something simpler.”*

James D. Murray

## 1.1 Um breve histórico

A origem da vida, do ponto de vista científico, tem se demonstrado um dos desafios mais atraentes da atualidade. Passados sessenta anos após a publicação do seminal *O que é Vida?* em 1944 por Erwin Schrödinger [1] ainda buscamos uma definição desta questão. Àquele tempo, Schrödinger buscava um critério que separasse o que definimos como vivo daquilo do que, convencionalmente, chamamos de não-vivo e perguntava se haveriam outras leis/princípios básicas para a matéria viva. Hoje, de posse de parte dessas leis (ao menos em tese) tenta-se responder àquela pergunta que nosso ilustre precursor não fez explicitamente mas deixou como legado científico. E não há nada melhor que uma boa pergunta, mesmo que nas entrelinhas, quando se trata de ciência. Schrödinger não se atreveu a dar uma definição de vida (apesar do nome de seu afamado livro). Não definiu os predicados exclusivos dos seres vivos. Permanece a pergunta. O que significa vida ou estar vivo? De uma maneira simples, a definição atualmente mais aceita afirma que o predicado que melhor define a vida é capacidade de replicação com fidelidade finita de uma dado aglomerado supramolecular [2]. Pragmaticamente: é a capacidade de fazer cópias imperfeitas de si mesmo. Na verdade, essa ainda é uma definição provisória e os últimos sessenta anos de pesquisa tem sido dedicados a avaliar a extensão dessa afirmativa em todas

as suas nuances. E, especialmente, delinear como foi e como é possível a emergência de entidades com essa incrível capacidade. A replicação é tão fantástica que, dadas as condições adequadas (não sabemos exatamente quais são), é capaz de gerar coisas tão díspares e complexas quanto bactérias bioluminescentes simbiote de peixes e lulas, baleias azuis e/ou viróides. Claro, o processo evolutivo não se restringe apenas a fazer réplicas imperfeitas. Esse é apenas o resumo do essencial da ópera da vida. Afinal, o que buscamos aqui nada mais é que acrescentar mais um pouco de luz ao drama vital, o atemporal "De onde viemos?", sob a tutela do método científico. E, como em toda questão desafiadora, não há uma resposta única até o presente momento. Felizmente, apesar da magnitude deste empreendimento, há boas tentativas. Atualmente, as teorias científicas sobre a origem da vida se dividem em duas vertentes básicas. Os adeptos da teoria **Metabolismo Primeiro** (MP) afirmam que a vida se originou a partir de conjuntos de moléculas mutuamente catalíticas organizadas em redes fechadas de reação, à moda de um ciclo metabólico tal como ciclo replicação atualmente conhecido em bactérias, por exemplo. Essas redes catalíticas primordiais, sob certas condições, seriam capazes de sintetizar todos os seus componentes *de novo*. Logo, essa forma primordial de metabolismo seria também a forma primordial de replicação. A diferença entre isto e o que vemos hoje seria apenas quantitativa. Existem várias versões dessa abordagem, cada uma com suas peculiaridades, porém, com a mesma essência (c.f. [3, 4, 5]). E, deve-se ressaltar que esta é a vertente pioneira, datando dos idos de 1930 com os trabalhos de Aleksandr Ivanovich Oparin [6]. A outra vertente, denominada **Genomas Primeiro** (GP), assume que a vida se originou com o aparecimento de uma molécula capaz de dirigir sua própria síntese, i. e., um polímero informacional auto-replicativo. Esta molécula teria cooptado outras moléculas presentes no meio ambiente primordial e servido como guia e centro organizador de informação. Durante os anos de 1970, os adeptos desta abordagem obtiveram resultados muito importantes partindo de primeiros princípios, atraindo para si o foco da comunidade científica e reduzindo drasticamente o brilho dos adeptos da MP. Além disso, certos modelos baseados nas

premissas desta vertente sofreram escrutínio experimental detalhado, notadamente o modelo de quasi-espécies, com resultados surpreendentes e promissores [7, 8]. Por fim, como tudo na vida (e mesmo na evolução), essas duas vertentes não definem as cores do cenário das teorias em evolução pré-biótica e origem da vida. Sem dúvida, existem teorias mistas que se aproveitam das idéias de ambas abordagens [9, 10]. No mais, estas abordagens não são mutuamente excludentes. Não há nada impedindo ambas de estarem simultaneamente corretas e seus resultados sejam compatíveis. O que fica como lição da literatura sobre o assunto é a possibilidade de uma origem múltipla da vida [2]. Por questão de objetividade, usaremos a abordagem GP como paradigma aqui. Em especial, será usada uma formulação com diversas características do modelo do corretor estocástico, mas compatível com os resultados do modelo de quasi-espécies [11, 12]. Antes, cabe aqui uma breve recuperação dos resultados fundamentais do modelo de quasi-espécies, considerado como um paradigma na discussões que seguem, e dos principais problemas que este suscitou na literatura científica.

## 1.2 A crise informacional na origem da vida

A limitação do comprimento de um polímero informacional devido (in)fidelidade de replicação tem se demonstrado um desafio a vertente GP sobre evolução pré-biótica e origem da vida. De acordo com o modelo de quasi-espécies proposto por Eigen na década de 1970 [7], em um cenário puramente competitivo, um dado polímero replicativo (i. e. um replicador e/ou gene) precisará de uma alta fidelidade de replicação por monômero afim de garantir sua presença na próxima geração. Caso contrário, este será completa e inexoravelmente substituído por um mutante. Assim, sob certas condições, existe uma fidelidade de replicação crítica  $q_c$  abaixo da qual parte da informação contida na população de replicadores é irreversivelmente perdida. Esse limiar, denominado limiar da catástrofe de erros, implica em severas limitações na quantidade de informação que um dado sistema genético deste tipo pode conter. Além disso, cria-se um paradoxo bem ao estilo

”Quem veio primeiro: o ovo ou a galinha?”. Apesar deste último ter uma boa resposta do ponto de vista biológico,<sup>1</sup>a versão pré-biótica molecular deste ainda não tem solução em vista. Em termos práticos: para termos replicadores de tamanho arbitrário (e.g. genomas) precisamos de catalisadores especializados (e.g. enzimas tipo polimerases/replicases) para garantir a fidelidade de replicação necessária, mas para termos catalisadores especializados precisamos de replicadores de tamanho arbitrário para codificar os catalisadores[13]. E, como se já não houvesse complicações suficientes, a ação o princípio da exclusão competitiva (c.f.[14]) no cenário do modelo de quasi-espécies impede a coexistência de dois ou mais tipos de replicadores que difiram significativamente em seu conteúdo informacional, a não ser que estes possuam taxas de replicação suficientemente similares[15]. Duas quasi-espécies não podem coexistir. Esses dois resultados consistem na famigerada crise informacional em evolução pré-biótica e são a derrocada definitiva da visão simplista de que um sistema genético complexo pode ser construído a partir de uma coleção de replicadores/genes independentes em uma dinâmica competitiva. Ou seja, apenas seleção natural não permite evolução de longo termo e não garante a estabilidade de informação genética, apesar de ser claramente um elemento primordial de sistemas competitivos. Faltam ingredientes essenciais à estes modelos capazes de lidar com aspectos como origem e estabilidade da informação genética que tornem os cenários propostos mais plausíveis. Também é notável que este limiar não esteja confinado a modelos de polímeros informacionais. Certos modelos de seleção de grupo e de genética de populações apresentam uma formulação distinta, mas um comportamento qualitativamente equivalente (e.g. [12, 16]). Logo, o limiar de erros é um fenômeno associado a estabilidade da informação genética no tempo. Por outro lado, já está claro que este limiar não é um fenômeno generalizado nestas classes de modelos, mas sim, particular de uma classe com severas restrições nas taxas de

---

<sup>1</sup>O ovo como conhecemos, i.e. o ovo amniótico, surgiu no Pennsylvaniano (323 a 290 milhões de anos atrás) enquanto que as aves datam do final do Jurássico (206 a 144 m.a.a.). Ou seja, O ovo está em larga vantagem!

replicação associadas a diferentes tipos replicadores. O limiar depende da paisagem de *fitness* usada. Em especial, o limiar da catástrofe de erros como observado no modelo de quasi-espécies só existe devido à forma degenerada da paisagem de *fitness* usada[17]. Isso deixa claro que a crise informacional no início da vida na Terra pode simplesmente nunca ter existido, ao menos não da forma proposta por Eigen [18]. De qualquer maneira, mesmo que a crise informacional delineada no modelo das quasi-espécies seja apenas mais uma tentativa sem sucesso, o recado foi captado pela comunidade científica: precisamos de modelos mais convincentes sobre a origem da vida!!! E como precisamos! Esse dilema informacional tem persistido por mais de três décadas e, até o presente, nenhuma solução satisfatória, plenamente aceita pela comunidade científica, foi proposta. A maioria das propostas para solucionar esta crise informacional assume como inevitável o limiar de erros e se ocupa de estudar as condições necessárias e suficientes para garantir a coexistência estável de uma dada diversidade de genes com esparsas exceções [19, 20]. Mesmo assim, a problemática acima alimentou diversos programas de pesquisa e gerou apreciável literatura (para uma boa revisão c.f. [2]). Vale destacar as principais soluções e traçar uma breve perspectiva.

### 1.3 Algumas (possíveis) soluções

Nos mais de trinta anos que sucederam os trabalhos seminais de Eigen e colaboradores [7, 21] muitas soluções foram propostas para a crise genética primordial (cf. [22]), a maioria se baseando fortemente no modelo de quasi-espécies ou numa formulação similar de redes cinéticas autocatalíticas. Formulações distintas usando genética de populações[16] e processo estocásticos [16] também podem ser encontradas. De qualquer forma, o cenário é dominado pelo modelo de *hiperciclos* [23, 24, 25]. Mesmo tendo sido duramente criticado [26, 27, 28] e tendo limitações evidentes e comprometedoras, os hiperciclos tem sido assumidos (e usados) como a melhor solução para a crise da informação genética primordial. Em 1986, surgiu a primeira formulação alternativa e nova de uma solução para o famoso problema: o

modelo do *corretor estocástico* [11, 29]. Na realidade, este modelo não era propriamente uma novidade já que trabalhos similares haviam sido feitos no início dos anos 1980 [27, 30]. À época, apesar do modelos de Bresch e colaboradores confrontarem diretamente o modelo de hiperciclos, parecia claro que cada modelo tratava de aspectos distintos da evolução pré-biótica. Voltaremos a esta história em mais detalhes no capítulo 3 desta dissertação. De qualquer forma, o corretor estocástico se estabeleceu como uma alternativa viável ao modelo de hiperciclos e o cenário científico se viu muito enriquecido nos anos seguintes. É claro que este não passou incólume às críticas da comunidade científica. Na verdade, durante a sua formulação, herdou as críticas feitas por Eigen e colaboradores às primeiras tentativas de modelos de compartimentos [31]. Entretanto, os modelos de compartimentos, em especial o corretor estocástico, vieram para ficar. Mais recentemente, uma nova classe de modelos, de caráter sinérgico, apareceu na literatura e clama por ser capaz de unir a abordagem cinética dos hiperciclos à abordagem genética e de seleção de grupo do corretor estocástico. Ou seja, o melhor dos dois mundos. Este modelos, dos quais se destaca o denominado *Graded Autocatalysis Replication Domain*, ou simplesmente GARD, se denominam capazes de contemplar ambas vertentes em evolução pré-biótica [9]. De fato, uma aplicação recente do mesmo demonstra que é possível o aparecimento e a estabilidade de genomas composicionais (*sensu* Dyson, [2]) a partir de redes metabólicas mutuamente catalíticas. Em adição, o mesmo modelo usa uma abordagem estocástica ao distribuir as diversas populações moleculares do modelo entre as vesículas da população [10]. Por outro lado, este modelo herda os problemas e críticas dos modelos que o inspiraram, mais aquelas da vertente MP. Para fins "didáticos", vamos elaborar um breve resumo das premissas, resultados e críticas destes três modelos exemplares.

### 1.3.1 Hiperciclos

O modelo de hiperciclos [23, 24, 25] propõe que podemos contornar o problema da catástrofe de erros se pudermos garantir que cada replicador aja como sub-

trato/molde para um outro replicador e como replicase tanto para si mesmo quanto para outro replicador subsequente. Logo, poderíamos imaginar a formação de um ciclo de replicação regulado por retroalimentação entre seus componentes. A inserção de novos componentes no ciclo garantiria o aumento de informação, algo necessário para atingir o tamanho suficiente para a codificação de catalisadores especializados [21]. Em sua versão realista, a função de replicase no modelo seria desempenhada por uma "enzima" codificada pelo replicador. Mas, com a ascensão do cenário "RNA world" [32], a função de replicase foi admitida em moléculas de RNA. O modelo viveu momentos de triunfo quando foi estabelecida atividade catalítica específica em alguns tipos de RNA [33]. Porém, este modelo apresenta vulnerabilidades intrínsecas. Dada a ligação que existe entre os componentes do ciclo, fica claro que a falha de um único componente resulta na falha de todos. Em ciclos com mais de quatro tipos de replicadores, essas falhas e o fato desse sistema ser instável, geram bruscas oscilações nas concentrações dos replicadores. Isso implica tornar todo o ciclo muito suscetível à extinção por fatores puramente estocásticos [34]. Ainda mais, a presença de mutantes que não agem como replicases mas servem como molde compromete definitivamente o esquema, gerando um estado absorvente capaz levar à extinção toda a população [35]. Há ainda o risco de mutantes tipo curto-circuito, ou seja, ao invés de cooperar com o membro seguinte este coopera com algum outro, reduzindo o tamanho efetivo do ciclo, e novamente ficamos com problemas de estabilidade. Essas críticas ao modelo foram colocadas logo após a sua publicação (cf. [26, 27]). Trocando em miúdos, por ser sensível apenas a variações na taxas de replicação, a seleção natural é incapaz de fortalecer o vínculo entre os replicadores num cenário como este. Ocorre que o processo de seleção tende a aumentar a afinidade de um dado replicador por sua replicase. O mesmo não acontece com sua atividade de replicase. No fim, acabamos ou como uma única quasi-espécie (i. e., um único tipo de replicador) ou a população se extingue totalmente. Em suma, os hiperciclos não demonstram o potencial evolucionário desejado de uma solução definitiva do problema [28].

### 1.3.2 Corretor Estocástico

Afim de resolver o problema da coexistência de replicadores distintos e visando ser uma alternativa aos hiperciclos, uma série de modelos inspirados nos trabalhos seminais sobre a origem da vida (e.g, [6, 36]) começaram a surgir a partir de 1980. Essa classe de modelos, aqui denominados modelos de vesículas devido a sua inspiração nos coacervados de Oparin (cf. [37]), assume que os replicadores formam grupos isolados (as vesículas e/ou coacervados). Logo, esta é uma versão explicitamente finita, discreta e estocástica dos famigerados "bags of genes" da genética populacional tradicional. Mas, não se trata simplesmente uma versão, como veremos a seguir. O grande problema dos hiperciclos reside no fato de que, *a priori*, estes estão desprovidos de mecanismos intrínsecos para reprimir ou eliminar mutantes não-funcionais, i.e., parasitas e hiperparasitas. Nesse ínterim, a compartimentalização dos replicadores é vista como uma solução ao problema da invasão de mutantes não-funcionais. Em tese, como cada vesícula é uma população isolada de replicadores e imaginando uma distribuição aleatória dos replicadores entre as vesículas descendentes é fácil visualizar que, dentro de certos limites, iremos produzir vesículas que preservam a composição inicial ou que mantiveram uma carga de mutantes baixa o suficiente para garantir a sua viabilidade. Ou seja, há um mecanismo inerente para eliminar parasitas/mutantes da população. Nesta vertente, o modelo de maior destaque é, sem dúvida, o corretor estocástico [11]. Inspirado largamente nos trabalhos de Bresch e colaboradores [27, 30, 38], este assume que a informação que codifica uma replicase não-específica está dividida entre diversos tipos de genes confinados em uma vesícula (uma micela/protocélula). A competição entre os genes de uma vesícula pela replicase garante a ação da seleção natural. Por outro lado, vesículas que não contém todos os componentes da replicase estão condenadas, pois perdem a capacidade de replicação. Dessa forma, cria-se uma maneira de expurgar genes mutantes não-funcionais da população e garante-se a operação da seleção natural [11, 39]. Apesar de, à primeira vista, parecer uma boa solução para o problema, o corretor estocástico também possui problemas intrínsecos. O mecanismo de expurgo

de genes indesejáveis é, certamente, limitado. Vesículas com poucos genes são muito suscetíveis à deriva genética, ou seja, a perda estocástica de informação. É muito claro que a deriva tem um papel central neste cenário. Por outro lado, vesículas capazes de comportar grandes populações e/ou com muitos genes distintos privilegiam enormemente aqueles com taxas de replicação mais elevadas. Logo, a seleção opera em favor de alguns genes e a exclusão competitiva acaba por inviabilizar estas vesículas. Também, não há nenhum mecanismo conhecido pelo qual a compartimentalização ocorreria espontaneamente e a permeabilidade destas supostas protocélulas seria, indubitavelmente, muito baixa [40, 41]. Para completar, um dos cenários mais respeitados de como deveria ser a dinâmica genética entre as protocélulas assume que a transferência lateral de genes desempenhou um papel fundamental na consolidação dos primeiros genes na história da vida na Terra [42]. Sabemos que, na realidade, células não são compartimentos isolados. Ou seja, a dinâmica da informação genética em populações de "vesículas" reais deve ser muito similar a uma dinâmica de metapopulações. Para completar, estes modelos não substituem a proposta dos hiperciclos. De fato, esta parte de primeiros princípios e requer apenas a presença de replicadores [31]. Mas, sem demérito algum, modelos de vesículas podem ser interpretados como um passo subsequente e crucial, pois a organização celular nada mais seria que uma versão mais complexa de uma vesícula pré-biótica [43]. Logo, apesar das críticas, a relevância destes modelos não pode ser diminuída de maneira alguma.

### 1.3.3 GARD

Como dito anteriormente, o ingrediente essencial do processo evolutivo é a replicação com fidelidade finita de espécies moleculares, em geral biopolímeros. Nos dois modelos apresentados anteriormente, os replicadores são, efetivamente, biopolímeros e, no caso do corretor estocástico, facilmente interpretados como genes independentes. Nestes dois casos, conta-se com a atividade de replicação *a priori* ao menos em algum dos componentes do modelo. Em franco contraste com estes, o modelo proposto por Segré e colaboradores [9] parece ser a alternativa que parte

da junção dos modelos de vesículas aos modelos tipo hiperciclo num cenário típico da vertente MP das teorias sobre evolução pré-biótica. A idéia principal do modelo reside em criar conjuntos de espécies químicas mutuamente catalíticas confinadas em um compartimento (uma vesícula) no qual, sob certas condições, estas seriam capazes de reproduzir-se uma a uma *de novo*, i. e. apresentariam confinamento catalítico (*catalytic closure*, *sensu* Kauffman [3]). Entenda confinamento aqui como equivalente é propriedade de conjuntos parcialmente ordenados onde o resultado de uma operação qualquer entre dois elementos do conjunto também é elemento deste conjunto, como numa  $\sigma$ -álgebra. As paredes da vesículas funcionam como barreiras semi-permeáveis, não permitindo que as espécies químicas catalíticas escapem, mas permitindo que espécies que funcionem como substrato (e.g. monômeros) entrem. Em suma, é algo como uma protocélula. Mas, diferentemente da maioria dos modelos de vesículas, este é um modelo cinético similar aos hiperciclos. Neste cenário, a replicação pode ser definida como o processo de produzir cópias adicionais de cada componente catalítico a partir de componentes mais simples. Logo, temos todos componentes dos hiperciclos bem como os do corretor estocástico. O confinamento catalítico equivale ao particionamento da informação essencial. Resultados mais recentes da aplicação deste modelo introduziram a noção de **genomas composicionais**, i.e., "genomas" definidos pelo conjunto de espécies catalíticas contidas numa dada vesícula [10]. Apesar de ser uma alternativa aos outros modelos sobre origem e estabilidade da informação genética, este ainda não foi submetido a uma análise mais apurada da comunidade científica e pouco se sabe acerca de seu comportamento em relação à fidelidade de replicação, à presença de espécies sem atividade catalítica ou mesmo seu comportamento na presença de algo equivalente à taxa de mutação presente nos outros modelos.

## 1.4 **Cooperação: um ingrediente essencial**

Destacados os pontos acima, fica claro o que há de comum entre estes modelos. Todos tem como ingrediente essencial a cooperação, mais ou menos explícita,

entre os replicadores/genes. Há uma vasta literatura acerca da evolução e manutenção da cooperação em sistemas biológicos. A cooperação desafia a interpretação ortodoxa da teoria evolutiva e características ou comportamentos cooperativos são, em primeira instância, contrários à seleção natural. O próprio Charles R. Darwin (1809-1882) já reconhecia este problema em sua famosa afirmação: "*Natural selection cannot possible produce any modification in any one species exclusively for the good of another species*" (p. 228, [44]). Mas, é fato que características desse tipo abundam na Natureza e em todos os níveis (para uma boa revisão cf. [45]). Recentemente, a idéia de cooperação tem sido empregada em contextos moleculares. Por exemplo, proteínas como o complexo GroEL/GroES, cuja única função é auxiliar no enovelamento de outras proteínas, têm sido interpretadas como possuindo um comportamento "altruísta" [46]. De fato, o modelo de hiperciclos, apesar de não o fazer explicitamente, assume a cooperação como uma característica fundamental do processo vivo mesmo num nível tão basal tal qual o molecular. A repressão da competição tem sido vista como uma das formas pela qual a cooperação age promovendo a manutenção da diversidade e atuando na organização dos componentes de um dado sistema biológico [47]. O modelos de vesículas, de um modo geral, fazem uso de um nível a mais de seleção, a seleção de grupo, para contrabalançar os efeitos da seleção natural. Dessa forma, implementa-se a repressão da competição via seleção de grupo. Colocados estes ingredientes, espera-se que haja um balanço entre entre seleção individual e seleção de grupo numa dinâmica de competição-cooperação. Ou melhor dizendo, numa dinâmica de cooperação e conflito entre os agentes do modelo. A seleção de grupo sempre foi muito criticada pela comunidade científica, especialmente entre adeptos do programa adaptacionista em evolução. Passou por um longo período de ostracismo, devido primariamente ao efeito da publicação do controverso "*The Selfish Gene*", no final dos anos 1970 [48]. Mas, retornou com vigor renovado nos últimos anos, agora fortemente embasada em teorias e evidências experimentais [49]. Apesar dos calorosos debates acerca de qual o nível de organização onde a seleção natural efetivamente opera ainda estarem em aberto, características

importantes foram observadas em modelos com diversos níveis de seleção atuando concomitantemente e tem especial valor em evolução pré-biótica[50]. Tudo indica que a seleção de grupo voltou para ficar.

## 1.5 O problema da coexistência

A proposta desta dissertação é, sucintamente, acrescentar um pouco mais de detalhe ao cenário exposto acima. Os modelos apresentados anteriormente têm como intuito principal demonstrar que é possível a coexistência estável de diferentes replicadores/genes ao longo do tempo. Entretanto, essa diferença é apenas, por assim dizer, superficial. Os replicadores não são qualitativamente diferentes entre si. Apesar de existirem variantes que exploram a presença de replicadores/genes intrinsecamente distintos (e.g. [35, 12, 43], estes também buscam responder as mesmas questões. A inspiração aqui difere um pouco. Iremos usar um versão adaptada do corretor estocástico [12] com ingredientes do famigerado *Wright's Island Model* [51] para estudar o problema da coexistência de replicadores/genes com características distintas. Devido ao número de parâmetros envolvidos na especificação de modelos dessa natureza, até muito recentemente faltava na literatura especializada um tratamento mais completo de algum destes. Por exemplo, mesmo os casos mais simples do corretor estocástico não aceitam um tratamento analítico geral e na maioria das situações interessantes sequer é possível uma solução aproximada para o problema. Logo, resta apenas a resolução numérica ou a simulação do modelo. E esse é um dos intuítos desta dissertação: explorar amplamente um modelo dessa estirpe, tanto numericamente quando via simulação. Para completar, nenhum dos modelos citados acima se preocupou em relaxar a premissa pouco razoável de tempos de geração iguais entre os diversos níveis do modelo. Logo, vamos aproveitar a oportunidade para incorporar mais esse ingrediente ao modelo. De qualquer forma, a intenção aqui será explorar amplamente o espaço de parâmetros do modelo visando estabelecer os limites de coexistência entre os diferentes genes envolvidos. Procedendo desta forma, esperamos encontrar uma região no espaço de parâmetros onde o equilíbrio

entre os processos de seleção nos diferentes níveis do modelo garanta a maior diversidade possível, ou seja, o maior conteúdo informacional do sistema. Iremos usar duas formulações similares, mas com características particulares. Primeiramente, vamos explorar numericamente um modelo determinístico entre genes distintos competindo dentro de vesículas. E em segundo, vamos usar uma simulação computacional nos moldes dos modelos de compartimentos da década de 1980 em busca de um tanto a mais de realismo.

## 2 DINÂMICA DE VESÍCULAS I

### 2.1 Um modelo determinístico

*“There are two kinds of generalizations. One is the cheap the other is valuable. It is easy to generalize by diluting a little idea with a big terminology. It is much more difficult to prepare a redefined and condensed extract from several good ingredients”*

George Pólya

Baseados no *Wright’s Island Model* [51], consideramos uma população global composta de um número infinito de populações espacialmente localizadas - as vesículas - cada uma capaz de comportar exatamente  $N$  genes. Essas premissas são típicas de modelos de seleção de grupo e tem sido usadas com sucesso para o estudo da evolução e manutenção de características altruístas na natureza (e.g. [52, 53]). Como dito anteriormente, a intenção aqui é usar o mesmo arcabouço para estudar o delicado problema da coexistência de diferentes tipos de genes numa dinâmica de vesículas. Em particular, focaremos no caso de uma população de vesículas com dois tipos de genes. Sem perda de generalidade, assumimos que um dos tipos possui uma taxa de replicação  $1 - \tau$  onde  $\tau \in [0, 1]$  e o outro tipo possui taxa igual a 1. Logo, o parâmetro  $\tau$  será a medida intrínseca da intensidade de competição individual. Cada vesícula pode ser identificada por sua composição genética, isto é, pelo número  $i = 0, \dots, N$  de genes de um dado tipo que esta contém. Logo, temos  $N + 1$  tipos de vesículas e o estado global da população no instante  $t$  é completamente especificado pelas frequências das vesículas de tipo  $i$  (i.e., vesículas que possuem  $i$  genes de um tipo e  $N - i$  do outro), denotadas aqui por  $Y_i^t$  tal que  $\sum_i Y_i^t = 1$  para todo  $t$ . Dado um mecanismo para a dinâmica interna às vesículas e um para a dinâmica entre as

vesículas o intuito aqui é obter uma equação recursiva para a frequência dos tipos que compõem a população. O ciclo de vida (i.e., uma geração) das vesículas consiste de três eventos distintos - extinção, recolonização e replicação - que ocorrem nesta mesma ordem e são descritos a seguir.

## 2.2 Replicação vesicular

Como apontado anteriormente, a coexistência de diversos tipos de genes numa dada população resolve o problema da crise de informação genética se pudermos imaginar que a informação essencial está particionada entre os diversos tipos de genes cada qual abaixo do limiar de erro. Isso implica que cada tipo deve estar presente em qualquer tempo afim de garantir a viabilidade da vesícula. Além do mais, seria interessante se que cada tipo estivesse presente de maneira equânime garantindo assim o balanço informacional. Essa restrição em relação às concentrações dos tipos nos dá uma idéia da medida de probabilidade de sobrevivência,  $\alpha[i]$ , mais adequada ao processo de extinção. Vamos utilizar uma média geométrica adaptada

$$\alpha_i = 1 - g + \frac{2g}{N} [i(N - i)]^{\frac{1}{2}} \quad (1)$$

onde o parâmetro  $g$  nos dá a intensidade da seleção de grupo, como dito anteriormente [43]. De outro modo,  $g$  é a medida da repressão da competição e representa qualitativamente a cooperação entre os genes em cada vesícula. No caso, a despeito do valor de  $g > 0$ , a sobrevivência é garantida (i.e.,  $\alpha_i = 1$ ) para vesículas que possuem uma composição genética balanceada ( $i = N/2$ ). Por outro lado, vesículas que não possuem um dos dois tipos de genes ( $i = 0, i = N$ ) terão probabilidade de sobrevivência  $\alpha_0 = \alpha_N = 1 - g$ . Logo, a intensidade de seleção de grupo age diretamente reprimindo a expansão destes tipos de vesículas. Essa é uma maneira simples de modelar a repressão da competição nos moldes propostos na literatura recente (cf. [47]). Os outros tipos de vesículas que não caem nestes extremos possuem probabilidades de sobrevivência intermediárias de tal modo que quanto mais próximas da composição balanceada maior a probabilidade de sobrevivência. De modo geral,

podemos interpretar a intensidade da seleção de grupo como um vínculo metabólico de balanço gênico, ou seja, uma tendência de coordenar a replicação gênica numa forma rudimentar de desequilíbrio de ligação ou de um *operon* primitivo entre genes essenciais. Um desequilíbrio de ligação reduzido (ou a ausência de parte de um *operon*) num cenário como este, diminui drasticamente a capacidade metabólica de uma dada vesícula e, por conseguinte, a sua probabilidade de sobrevivência. O processo de extinção descrito anteriormente tem como consequência direta a perda de uma fração igual à  $1 - \sum_i \alpha_i Y_i$  das vesículas da população. O espaço proporcionado é, então, imediatamente recolonizado por novas vesículas de acordo com suas respectivas proporções após o processo de extinção. As novas frequências são

$$\frac{\alpha_i Y_i^t}{\sum_j \alpha_j Y_j^t} \quad (2)$$

para  $i = 0, \dots, N$  e fica claro que podemos interpretar a probabilidade de sobrevivência  $\alpha_i$  como a medida da taxa de replicação de uma dada vesícula. Logo, a extinção e recolonização podem ser vistas conjuntamente como o processo de replicação das vesículas. É importante notar que o processo de recolonização descrito acima, apesar de ser amplamente utilizado na literatura, implica na reposição dos níveis populacionais de modo instantâneo não importando a fração de vesículas que sobreviveram (excetuando-se a extinção completa). Como consequência, por menor que seja a fração de vesículas sobreviventes (mesmo uma única vesícula) na geração seguintes teremos uma nova população com o mesmo tamanho que a população inicial [54, 55, 12, 43]. Apesar de raramente explicitada, essa premissa aparece em quase todos os modelos populacionais determinísticos sem sobreposição de gerações. No caso acima é simples evitar situações bizarras e indesejáveis como aquelas apenas controlando o valor de  $g$  de tal modo que a fração  $1 - \sum_i \alpha_i Y_i$  não seja próxima de 1. De qualquer modo, não deixa de ser importante a exploração de casos extremos, especialmente se desejamos entender as relações entre os parâmetros do modelo [56].

## 2.3 Replicação gênica

Agora que já descrevemos a replicação das vesículas, precisamos descrever o que acontece no interior de cada uma destas. Como a capacidade de cada vesícula é finita e fixada em  $N$  genes precisamos de uma abordagem estocástica para representar os eventos no interior destas. Como de praxe, assumimos que o número de cópias que um dado gene fornece à próxima geração é proporcional a sua taxa relativa de replicação (*fitness*) e não há sobreposição de gerações na dinâmica interna. Então, dado que existem  $j$  genes do tipo  $A$  em uma vesícula e  $N - j$  do tipo  $B$ , ou seja, dada uma vesícula de tipo  $j$  a probabilidade que um dado gene seja escolhido para contribuir com uma cópia para a próxima geração é dada por:

$$w_j = \frac{j(1 - \tau)}{j(1 - \tau) + N - j} \quad (3)$$

Esta equação é prontamente identificada como sendo a taxa relativa de replicação (*fitness*) do gene  $A$  numa vesícula de tipo  $J$ . De modo similar, a probabilidade de se escolher o gene tipo  $B$  é o complementar desta, i.e.,  $1 - w_j$ . Logo, a probabilidade  $T_{kj}$  que uma vesícula de tipo  $j$ , em uma única geração, torne-se uma vesícula de tipo  $k$  é dada pela probabilidade de que exatamente  $k$  genes tipo  $A$  e  $N - k$  genes tipo  $B$  sejam escolhidos para a replicação. Isso não é novidade, já que este é o aclamado sorteio do modelo clássico de Wright-Fisher em genética populacional. Assim, a distribuição binomial

$$T_{kj} = \binom{N}{k} (w_j)^k (1 - w_j)^{N-k} \quad (4)$$

nos dá cada um dos elementos da matriz de transição  $T_{kj}$ . Entretanto, existe mais um fator a se considerar na dinâmica interna. Se temos como cenário para o modelo o famigerado *RNA world* então é necessário acrescentar mais um ingrediente ao modelo. Tradicionalmente, modelos como este assumem que os tempos de geração entre vesículas e dentro das vesículas são equivalentes. Cada geração da dinâmica entre vesículas corresponde a uma geração na dinâmica interna das vesículas. Essa premissa, apesar de razoável para a maioria dos genomas atuais, não é apropriada para um cenário pré-biótico. Os genes primitivos na ausência de mecanismos de

regulação devem se comportar como certos vírus de RNA e viróides atuais. Isso implica em diversas gerações gênicas para cada geração entre vesículas. Então, nada mais acertado que inserir esse elemento no modelo. Dessa forma, a probabilidade de que uma vesícula de tipo  $j$  torne-se uma vesícula de tipo  $k$  em  $m$  gerações da dinâmica interna pode ser escrita da seguinte forma:

$$R_{ij} = \sum_{k_{m-1}}^N \cdots \sum_{k_1}^N T_{i,k_{m-1}} T_{k_{m-1},k_{m-1}} \cdots T_{k_2,k_1} T_{k_1,j} = (\mathbf{T}^m)_{ij} \quad (5)$$

Fica claro que  $\sum_i R_{ij} = 1$  para todo  $j \in [0, 1, \dots, N]$ . O parâmetro  $m$  pode ser entendido como a razão entre dos tempos de geração das vesículas e dos genes, ou seja, a relação temporal entre as dinâmicas interna e externa. Se tomarmos o tempo de geração das vesículas como sendo a unidade, o tempo de geração na dinâmica dos genes nada mais será que a fração  $1/m$  deste tempo. Há uma limitação clara no modelo assim definido, já que não poderemos explorar situações para  $m < 1$ . Porém, esta limitação é facilmente contornável se imaginarmos que o significado de  $m < 1$  nada mais é que a aplicação da dinâmica interna a cada  $t \bmod (1/m)$ . Entretanto, estes casos não são relevantes no cenário pré-biótico aqui proposto. Mais detalhes sobre este aspecto podem ser conferidos em [56].

## 2.4 Dinâmica geral

Agora que já descrevemos como se processarão as dinâmicas entre vesículas e entre genes podemos reunir todos os processos numa única equação de recorrência:

$$Y_i^{t+1} = \frac{\sum_{j=0}^N R_{ij} \alpha_j Y_j^t}{\sum_{j=0}^N \alpha_j Y_j^t} \quad (6)$$

É notável que esta formulação, baseada no arcabouço do modelo de Wright para populações espacialmente estruturadas, contrasta com a de outros modelos de seleção de grupo que fazem uso do reservatório de propágulos (*propagule pool*, *sensu* Slatkin [57]), em especial com a formulação de Wilson [58] onde os grupos são desfeitos a cada geração para gerar o reservatório e em seguida reconstituídos a partir do mesmo.

Esse tipo de formulação, com grupos transientes e reservatório de propágulos favorece sensivelmente a coexistência. Imaginado que a fase de reconstituição se dá por amostragem do reservatório é fácil perceber que se esta for aleatória iremos privilegiar uma distribuição equânime dos elementos da população em questão. Por outro lado, ficamos com uma verdadeira dinâmica de metapopulações e o reservatório propicia um meio para a migração de elementos entre os grupos com todas as implicações deste fato. Porém, vale adiantar aqui que modelos como este não toleram altas taxas de migração, já que estas levariam a população se comportar como se fosse uma única vesícula. A formulação aqui apresentada não leva em conta nem migração ou mutação. Deixaremos a mutação para o próximo capítulo.

## 2.5 Explorando o modelo

Apesar de bastante simples, a equação 6 não possui uma solução fechada possível. Mesmo no regime estacionário  $t \rightarrow \infty$ , a situação não é diferente. Logo, a abordagem aqui é definir um condição inicial adequada e iterá-la diretamente até as frequências  $Y_i$  atingirem o regime estacionário. Mas, como dito anteriormente, o problema estudado aqui é encontrar as condições para a coexistência dos dois tipos de genes. Temos quatro parâmetros envolvidos: a capacidade das vesículas ( $N$ ), a intensidade de competição ( $\tau$ ), a intensidade da seleção de grupo ( $g$ ) e a razão entre o tempo de geração das duas dinâmicas ( $m$ ). Logo, o que estamos buscando é traçar um limiar que separe coexistência de não-coexistência. Fortuitamente, devido as características da formulação do modelo, podemos atacar a questão da coexistência de modo bastante direto. Por exemplo, dados  $N, m$  e  $g$  fixos, perguntamos qual o valor de  $\tau = \tau_c$  que garante a coexistência de ambos tipos de genes? Em termos mais práticos, dados  $N, m$  e  $g$ , existe  $\tau = \tau_c$  tal que  $\Omega = 1 - (Y_0^\infty + Y_N^\infty) > 0$ ? É fácil imaginar um algoritmo capaz de lidar com essa situação. Resolver este problema é equivalente a encontrar numericamente a raiz de uma dada equação. Mais simples que isso. Para  $\tau > \tau_c$  implica em  $\Omega = 0$  (a competição coíbe a coexistência) e para  $\tau < \tau_c$ ,  $\Omega > 0$ . Logo, o ponto  $(\tau_c, \Omega(\tau_c))$  é o equivalente à raiz de uma equação

simples. Então, basta usarmos uma adaptação de um método simples como o da bissecção para encontrarmos  $\tau_c$  com a precisão desejada. Vamos assumir que o regime estacionário é atingido após dez mil gerações, o que para modelos desse tipo é mais que suficiente (cf. [59]). Uma análise prévia do modelo, ao menos uma formulação similar na literatura, corroboram essa premissa [12]. Por fim, é importante ressaltar que, apesar de possuir elementos de um modelo estocástico, a dinâmica aqui proposta é um caso limite de um processo estocástico. No limite de de vesículas infinitas, os elementos  $Y_i^t$  do vetor  $Y$ , são efetivamente o valor esperado daquele tipo de vesícula em uma população assim descrita. Portanto, não há nenhuma variável aleatória envolvida. Ainda assim, os efeitos oriundos da população finita de genes em cada vesícula não desaparecem, como demonstram os resultados a seguir.

### 2.5.1 Vesículas independentes

Afim de apreciar melhor os obstáculos envolvidos da delicada questão da coexistência, vamos nos ater um momento no problema mais simples onde a probabilidade de coexistência das vesículas independe da sua composição genética. Isso é fácil de garantir, basta definir  $g = 0$  na equação 1 e teremos o efeito desejado. Além disso, fazendo  $m \rightarrow \infty$  permitimos que a deriva genética atue fixando um dos tipos em uma dada vesícula ( $i = 0$  ou  $i = N$ ) e como  $\alpha_0 = \alpha_1 = 1$  com  $g = 0$  temos que a competição é efetivamente eliminada do modelo. É claro que isso só é possível porque estamos lidando com uma população infinita. Caso fosse uma população finita, a deriva, agora atuando no nível das vesículas, acabaria por fixar apenas um dos tipos de vesículas na população eliminando um dos genes da população definitivamente. Isso é evidente, já que ao fixarmos um dos genes na vesícula esta passa a se comportar como se fosse um gene e perdemos o significado do parâmetro  $N$ . Usando a equação proposta em [56],

$$Y_N^\infty = \frac{1 - \exp(-N\tau p)}{1 - \exp(-N\tau)} \quad (7)$$

válida no limite de  $N \rightarrow \infty$  e  $\tau \rightarrow 0$  tal que o produto  $\tau N$  seja finito, de acordo com os modelos clássicos em genética de populações (cf. [60]). O parâmetro  $p$  é a frequência

inicial do gene A em cada vesícula. Como  $g = 0$ , sem dúvida,  $Y_0^\infty = 1 - Y_N^\infty$ . Usando a idéia proposta anteriormente, vamos comparar essa aproximação analítica com o resultado numérico da iteração direta da equação 6. Usamos como condição inicial  $Y_{N/2}^0 = 1$  para  $N = 2, 4, \dots$ . Portanto,  $Y_i^0 = 0$  para todo  $i \neq N/2$  e  $p = 1/2$ . A figura 1 demonstra os resultados. Fica claro que para valores pequenos da intensidade de competição  $\tau$  a aproximação analítica é razoável. Entretanto, nota-se que  $Y_N^\infty$  decresce consideravelmente mais rápido que  $\exp(-\tau N/2)$  ao aumentarmos  $\tau$ . O ponto importante aqui é enfatizar que a deriva genética é uma faca de dois gumes nas teorias em evolução pré-biótica. Ao mesmo tempo que esta permite que genes menos competitivos sejam fixados, ao menos em parte da população (i.e.,  $Y_N^\infty > 0$ , dado  $\tau$  suficientemente pequeno), também acaba por dificultar amplamente a coexistência de ambos tipos numa mesma vesícula. Se considerarmos um população finita o cenário fica ainda mais complicado já que teremos a deriva agindo nos dois níveis, entre genes e entre vesículas. Mesmo na ausência de competição a coexistência é ameaçada. Parece não haver saída para a crise informacional primordial. Fica claro, então, que questões de coexistência são muito mais desafiadoras (e muito menos estudadas) do que se pensa à primeira vista.

### 2.5.2 Vesículas com capacidade infinita

É importante observarmos mais atentamente os efeitos determinísticos implicados pelo parâmetro de intensidade de competição  $\tau$ . No limite de vesículas com capacidade infinita,  $N \rightarrow \infty$ , para todo  $\tau > 0$ , o tipo mais competitivo é fixado na vesícula independentemente do valor da intensidade de seleção de grupo  $g < 1$ . De fato, a razão pela qual uma população uniforme de vesículas balanceadas, como definido na condição inicial a seção anterior, é nada mais que a amplificação das flutuações da dinâmica interna às vesículas e acaba por resultar em uma população com os dois tipos opostos de vesículas. Essas flutuações são inerentes à natureza estocástica da dinâmica de populações finitas, assim como a deriva genética. Mas, para o limite  $N \rightarrow \infty$  perdemos estas flutuações e, dessa forma, acabamos com toda a

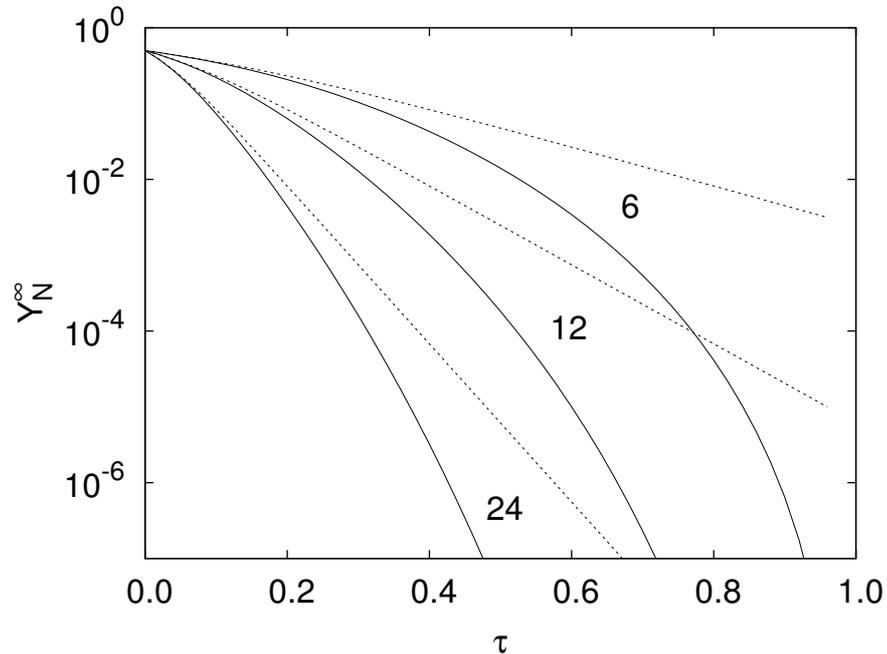


Figura 1: Frequências estacionárias das vesículas contendo o tipo menos competitivo em função da intensidade de competição para capacidades  $N = 6, 12$  e  $24$  no caso de vesículas independentes. As linhas sólidas são resultado exatos das equações de recorrência e as linhas pontilhadas são resultado da aproximação difusiva.

população de vesículas com o gene mais competitivo fixado. A exceção se faz no caso (degenerado) de  $\tau = 0$  onde, efetivamente, desligamos a competição e voltamos a ter a população dividida igualmente entre os dois tipos de genes em cada vesícula. Vale ressaltar aqui que, como o modelo é determinístico, as flutuações estão incorporadas nos elementos da equação 5, no sentido que a medida que  $N$  cresce, a dimensão da matriz  $\mathbf{T}$  também cresce. Como a soma de cada linha deve ser igual à unidade, isso implica em dividir a expectativa total por um número cada vez maior de transições possíveis. Como algumas transições tem expectativa maior que outras, no limite de população gênica infinita, de acordo com a formulação do modelo, a transição para uma vesícula com um só tipo gênico acaba por absorver toda a expectativa. As implicações destes limites são claras. A coexistência dos dois tipos de genes está restrita a uma faixa limitada do espaço de parâmetros, em particular da capacidade

$N$  de cada vesícula. Logo, nas próximas seções vamos buscar determinar o mais explicitamente possível estes limites de coexistência. Por fim, uma última observação. Apesar deste modelo não apresentar migração entre as vesículas, podemos imaginar facilmente o que aconteceria. Para taxas de migração pequenas, o isolamento das vesículas é eficiente e todos os resultados observados aqui se mantêm. Porém, para taxas de migração suficientemente grandes a população de vesículas se comportaria como uma única vesícula. Logo, o gene mais competitivo eliminaria por completo as outras variantes e terminariamos com um só tipo de vesícula [61].

### 2.5.3 Coexistência I - a ação da deriva

A partir de agora o nosso parâmetro de interesse será a fração  $\Omega$  de vesículas no regime estacionário que apresentam os dois tipos de genes independentemente de suas respectivas redundâncias, i.e., a fração  $\Omega = 1 - (Y_0^\infty + Y_N^\infty)$ . Como dito anteriormente, no caso de intensidade nula de competição ( $\tau = 0$ ) o único contratempo à coexistência é a deriva genética. Sem dúvida, para populações de vesículas com pouca capacidade, digamos  $N < N_c^-$ , a deriva não permitirá a coexistência dos dois tipos de genes e teremos  $\Omega = 0$ . Para complementar a situação temos ainda que avaliar o efeito das diferentes temporalidades das dinâmicas do modelo, expressas pela razão  $m$ . É fácil imaginar o efeito de  $m$  na ausência de competição. Como temos  $m$  gerações na dinâmica interna para cada geração de vesículas estamos efetivamente ampliando o efeito da deriva na população, ou seja, estamos acelerando a ação da deriva. E é exatamente isso que vemos na figura 2. Ainda podemos ressaltar outros achados importantes. Nessa situação, é possível notar que para  $N \rightarrow \infty$  temos que  $\Omega \rightarrow 1$ . Além disso, ao fixar um valor para a intensidade de seleção de grupo  $g$  acabamos por determinar um valor mínimo para a capacidade de cada vesícula,  $N = N_c^-$ , tal que abaixo deste a coexistência é impossível. Esse limiar é uma clara demonstração da ação da deriva genética na dinâmica interna das vesículas. Para completar, o limiar é sensível ao aumento de  $m$ , ou seja, aumentar  $m$  acaba por amplificar o efeito da deriva pelo simples fato de acelerar a dinâmica, como foi predito.

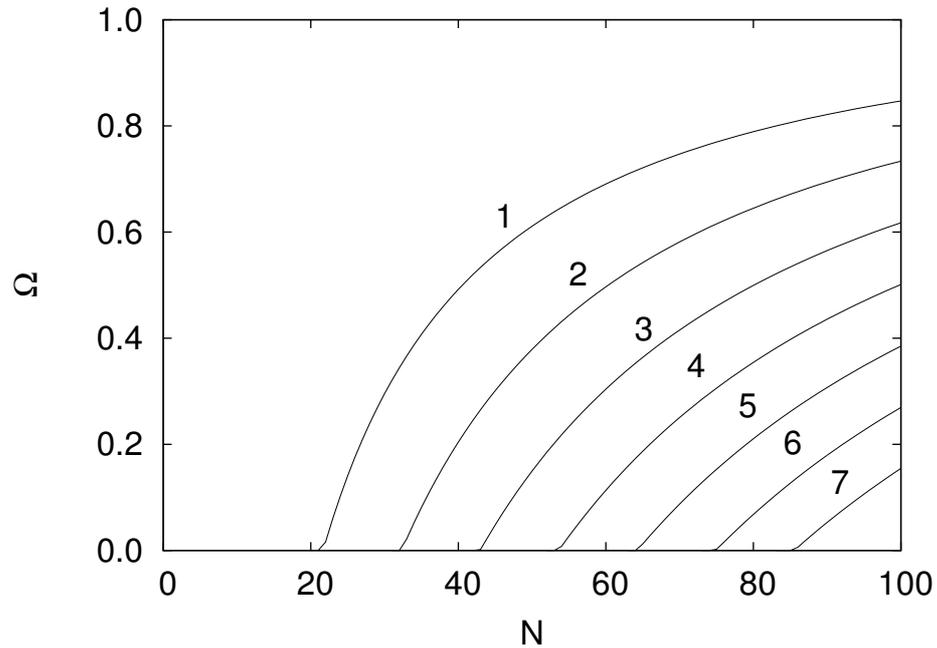


Figura 2: Fração das vesículas com os dois tipos em função da capacidade das vesículas para  $\tau = 0$  (i.e., ambos tipos são competidores equivalentes),  $g = 0.1$  e  $m$  como indicado no gráfico.

Fica claro, então, que acelerar a dinâmica interna inibe a coexistência e desloca o limiar  $N_c^-$  para valores mais elevados. De fato, a figura 3 reitera essa observação e mostra que  $N_c^-$  aumenta linearmente com  $m$  para  $g$  fixo. Como é de se esperar, o incremento em  $g$  reduz a inclinação, de um modo não-trivial, da relação entre  $N$  e  $m$ . Nessa situação também é notável que a relação entre  $N$  e  $m$  seja linear para  $\Omega$  fixo. A figura 4 denota a paisagem gerada pela deriva genética na ausência de competição. A superfície delimita o limiar de coexistência. Guarde bem este resultado.

#### 2.5.4 Coexistência II - deriva e competição

Agora, vamos complicar as coisas definitivamente. O foco agora é nos efeitos da intensidade de competição aliado aos efeitos da deriva. Os resultados anteriores demonstram claramente que a deriva impõe um limite mínimo na capacidade da vesículas, fixados os outros parâmetros do modelo. Mas, na seção 2.5.2 afirmamos

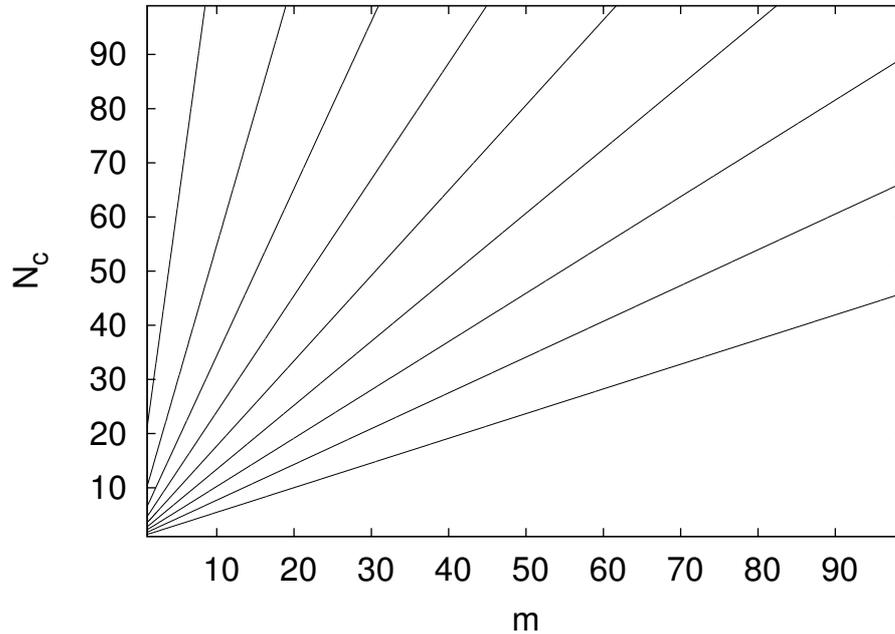


Figura 3: Valor crítico da capacidade das vesículas em função da razão entre os tempos de geração para  $\tau = 0$  e (da esquerda para direita)  $g = 0.1, 0.2, \dots, 0.9$ . Abaixo de  $N_c$ , a fixação de um dos tipos via deriva genética impede a coexistência.

que no limite de  $N$  grande favorecemos o gene com maior vantagem seletiva (o gene B no caso). Ou seja, dada uma intensidade de competição  $\tau > 0$  espera-se que o princípio da exclusão competitiva mostre seus efeitos fixando o gene B em parte da população de vesículas. Isso nos leva a imaginar que agora teremos um máximo de capacidade de cada vesícula acima do qual o efeito determinístico da competição coíbe a coexistência dos dois tipos de genes. Não podemos esquecer da razão entre os tempos de geração neste caso. Como  $m$  efetivamente representa o número de geração internas para cada geração de vesículas podemos esperar que o aumento de  $m$  intensifica os efeitos da competição. Isso é evidente, já que a seleção de grupo atua apenas no nível das vesículas e seu efeito mitigador sobre a competição só pode manifestar após a ação da competição entre os genes. A figura 5 representa exatamente essas previsões. Fica claro que para  $N > N_c^+$  a coexistência é impossível. E um aumento em  $m$  altera drasticamente esse limiar superior, como esperávamos. Por outro lado,

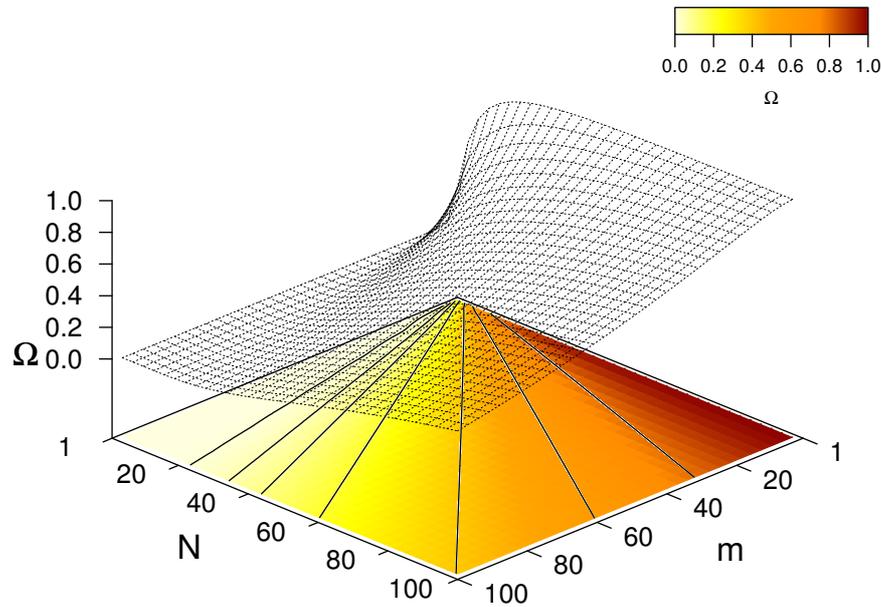


Figura 4: Paisagem de deriva - a superfície denota o limiar de coexistência - acima desta a coexistência é impossível. A forma da superfície é exclusivamente devida à deriva já que neste caso temos  $g \rightarrow 1$  e  $\tau = 0$ . As curvas de nível projetadas na base do gráfico são para  $\Omega = 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8$ . Os pontos foram obtidos ao término da replicação dos genes, como descrito no modelo.

o limiar inferior parece ser pouco afetado pela presença de competição. Assim, de forma sucinta caracterizamos a presença de dois limiares de coexistência  $N_c^-$  e  $N_c^+$  em relação a capacidade de cada vesícula. Ou seja, mostramos que o conteúdo informacional de um sistema como este, ao menos em relação à redundância, possui claras limitações. Mas, ainda há mais para se explorar. A figura 6 mostra que a presença de competição muda drasticamente aquela paisagem de deriva apresentada na seção anterior. Compara os gráficos. Mesmo a relação entre  $N$  e  $m$  já não é mais linear. E pela mudança observada podemos dizer que a limitação imposta pela competição é muito severa. Agora, para um dado nível de coexistência  $\Omega$  fixado, um aumento arbitrário simplesmente leva à extinção de uma parte da população. Também podemos perceber um máximo em  $m$  que nos mostra que a amplificação dos efeitos determinísticos devido à uma dinâmica interna acelerada tendem a se tornarem des-

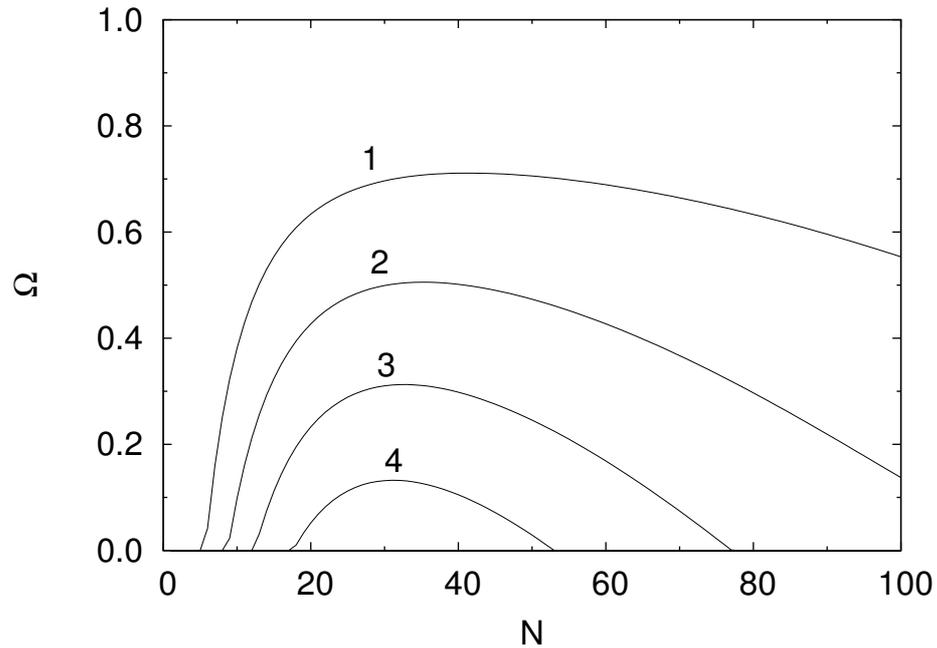


Figura 5: Fração de vesículas com os dois tipos de genes em função da capacidade das vesículas para  $\tau = 0.1$ ,  $g = 0.35$  e  $m$  como indicado no gráfico.

proporcionalmente severos com o aumento de  $N$ . Vamos observar agora o quanto a seleção de grupo pode compensar os efeitos da deriva e da competição com especial atenção na ação da razão entre os tempos de geração das dinâmicas. Como colocado acima, a seleção de grupo só pode atuar no nível das vesículas. Logo, em relação a competição (que atua incessantemente) e seleção de grupo fica em desvantagem nesta formulação. Isso significa que devemos esperar que um aumento na intensidade de competição exija um aumento ainda maior da intensidade de seleção de grupo para garantir o mesmo limiar de coexistência. A figura 7 expõe uma parte deste fato. Vemos que um aumento da intensidade de seleção de grupo compensa muito pouco o efeito de  $m$  na dinâmica. Um aumento considerável em  $g$  desloca comparativamente pouco o valor limite de  $m$  para além do qual a coexistência é impossível independentemente da capacidade das vesículas. E, um aumento de  $g$  também tem pouco efeito em relação a valores elevados de  $N_c$ . Contrastando com estes resultados, a figura 8 mostra que um aumento sutil na intensidade de competição desloca drasti-

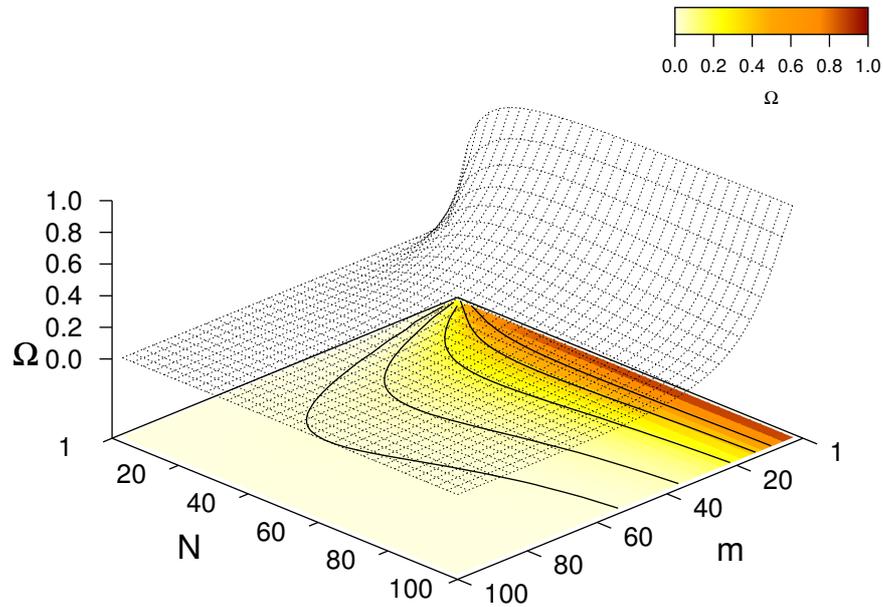


Figura 6: Paisagem de deriva deformada pela competição - a superfície denota o limiar de coexistência - acima desta a coexistência é impossível. Compare a forma da superfície com a da figura 4. Os parâmetros são  $g \rightarrow 1$  e  $\tau = 0.10$ . As curvas de nível projetadas na base do gráfico são para  $\Omega = 0.1, 0.2, 0.4, 0.6$ . Os pontos foram obtidos ao término da replicação dos genes, como descrito no modelo.

amente o valor limite de  $m$  como também muda a posição deste em relação à  $N_c$ , ou seja, desloca o valor ótimo de capacidade. Estes resultados estão resumidos na figura 9 onde é possível ver como estes valores críticos  $m = m_c$  se comportam em relação aos parâmetros  $g$  e  $\tau$ . A presença de competição, mesmo que suavemente, exige muito da intensidade de seleção de grupo para um dado  $m_c$  afim de garantir a coexistência. Para finalizar, analisemos mais intimamente a relação entre a intensidade de seleção de grupo  $g$  e a intensidade de competição  $\tau$ , deixando a razão entre os tempos de geração  $m$  para segundo plano. Isso também vai deixar mais claro o quanto a seleção de grupo pode fazer em relação à deriva genética. A figura 10 deixa muito claro que para vesículas com capacidade pequena precisamos de um valor elevado de seleção de grupo para garantirmos a coexistência dos dois tipos de genes. Mesmo na ausência total de competição. Portanto, para vesículas pequena

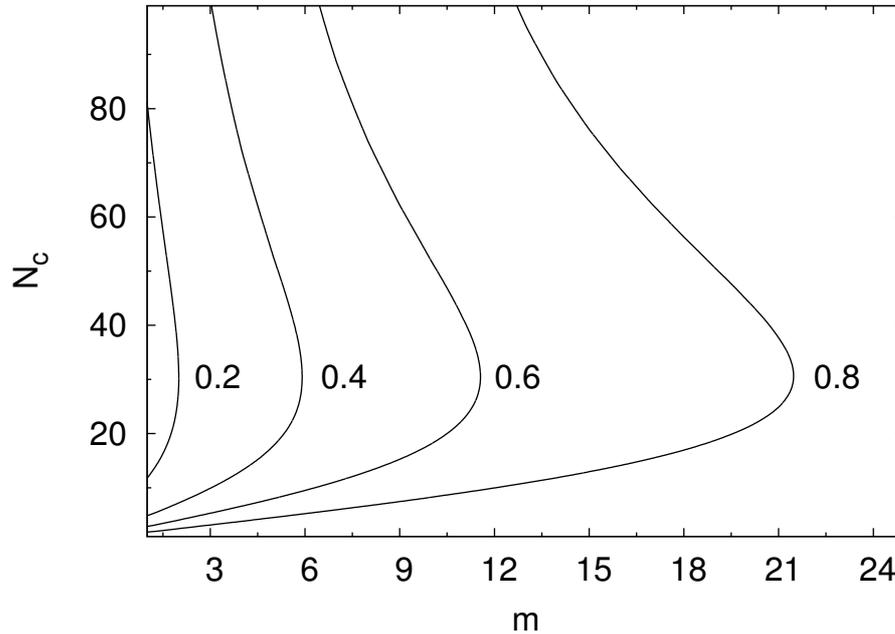


Figura 7: Valor crítico da capacidade das vesículas em função da razão entre os tempos de geração para  $\tau = 0.1$  e  $g$  como indicado no gráfico. Para valores fixos de  $m$  os dois valores de  $N_c$  determinam o intervalo de capacidade que permite a coexistência.

a grande dificuldade é realmente a deriva genética e não a competição. O início de cada uma das curvas até o primeiro ponto de inflexão deixam esse ponto ainda mais claro. Para uma vesícula de capacidade  $N$  fixada existe um  $g_c$  mínimo para garantir a coexistência com  $\tau_c = 0$ . Após esse mínimo, um aumento na intensidade de competição corresponde a um aumento pequeno na intensidade de seleção de grupo. Após a inflexão, temos uma região "tampão" entre os parâmetros, em especial para vesículas com capacidades maiores onde este efeito é mais notável. Porém, logo chegamos a segunda inflexão da curva e adentramos uma região onde a seleção de grupo já se mostra ineficaz para compensar a competição. Completando a análise, vemos que um aumento em  $m$  realmente amplifica a ação da competição. Em especial, para vesículas pequenas, uma aumento em  $m$  reduz ou elimina a região de tamponamento ou de linearidade da relação entre  $g$  e  $\tau$ . As figuras 11 e 12 apenas ilustram melhor esses pontos. Podemos ver claramente que na figura 11 a relação entre  $\tau$  e  $g$  desenha

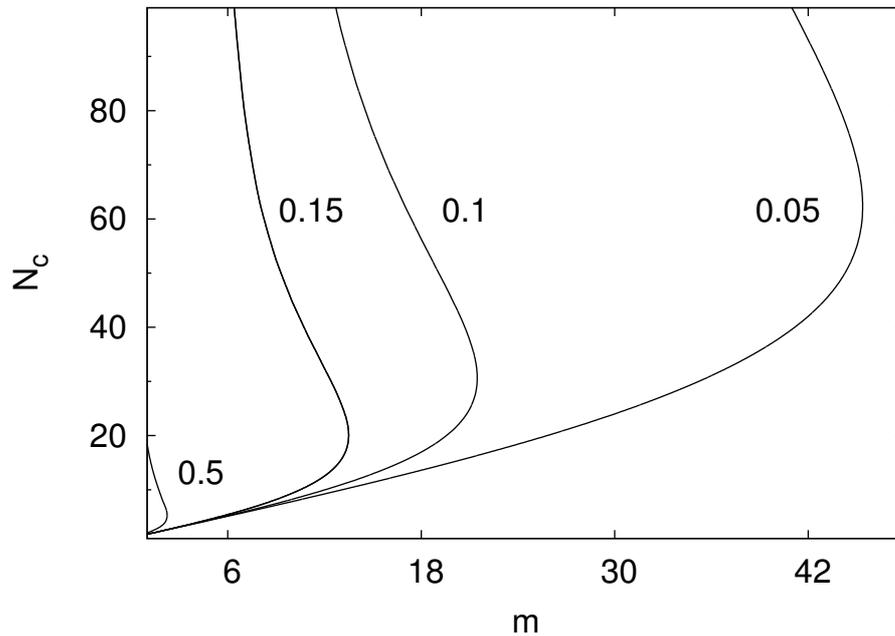


Figura 8: Valor crítico da capacidade das vesículas em função da razão entre os tempos de geração para  $g = 0.8$  e  $\tau$  como indicado no gráfico.

uma paisagem bastante suave, digamos uma colina de coexistência. Já na figura 12 onde a intensidade de competição está fortemente amplificada temos uma paisagem muito mais íngreme, um penhasco de coexistência.

## 2.6 Conclusões

Após essas análises mais pormenorizadas e uma boa exploração do espaço de parâmetros do modelo, temos condições de expor algumas conclusões importantes. Primeiramente, a compartimentalização de replicadores ou genes apesar de ser reconhecidamente um passo crucial na origem da vida celular [27, 31, 29] levanta questões dilemas ainda sem muita solução à vista. Ficou muito claro aqui que a compartimentalização gera um grande obstáculo a diversidade e estabilidade da informação genética - a deriva genética. Como demonstrado, a deriva inibe a coexistência de diversos tipos de genes numa dada população de protocélulas ao promover a fixação

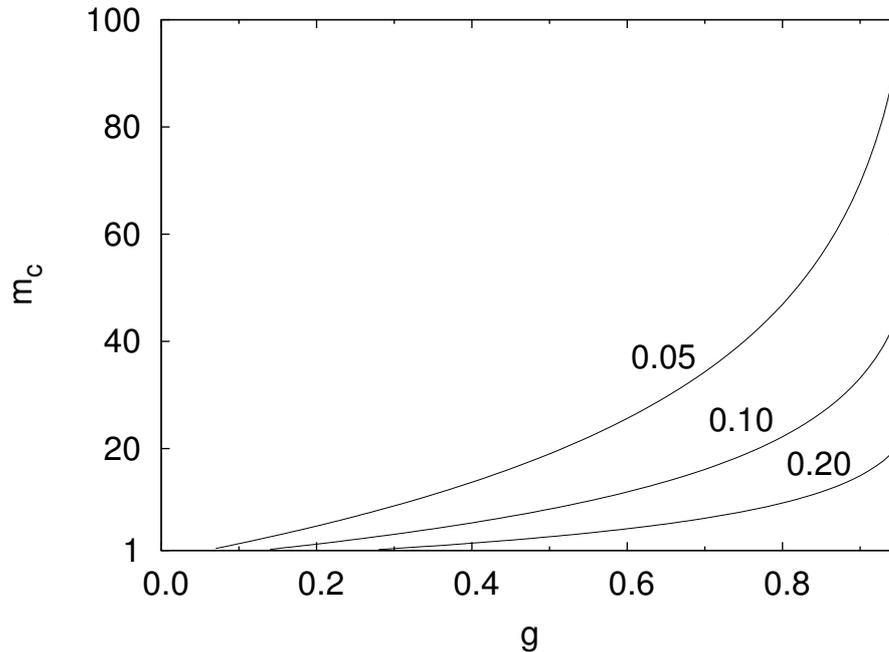


Figura 9: Valor crítico da razão entre os tempos de geração em função da pressão de cooperação para  $\tau$  como indicado no gráfico. Abaixo de  $m_c$  a coexistência é impossível.

de um dos tipos. Mais do que isso. Em protocélulas com capacidade reduzida os efeitos da deriva são ainda mais severos, o que nos deixa com um sério problema na transição de genes livres para genes confinados. Isso porque se logo de início não contarmos com a presença de compartimentos com a capacidade adequada, os efeitos positivos da compartimentalização sequer serão experimentados. Trocando em miúdos, é improvável a ocorrência de uma transição gradual de genes livres para genes em compartimentos sucessivamente maiores. No outro extremo da questão, como é inevitável a competição entre entidades tão similares como os genes primordiais. Ora, então também ficamos impedidos de transitarmos para protocélulas com grande capacidade, pois os efeitos determinísticos da competição coibiriam qualquer possibilidade de coexistência de genes distintos. Estamos restritos a uma faixa bem delimitada de capacidade das protocélulas que permita a coexistência. Uma faixa que depende sensivelmente de diversos parâmetros com relevância biológica indiscutível como a intensidade de competição, intensidade de seleção de grupo, ve-

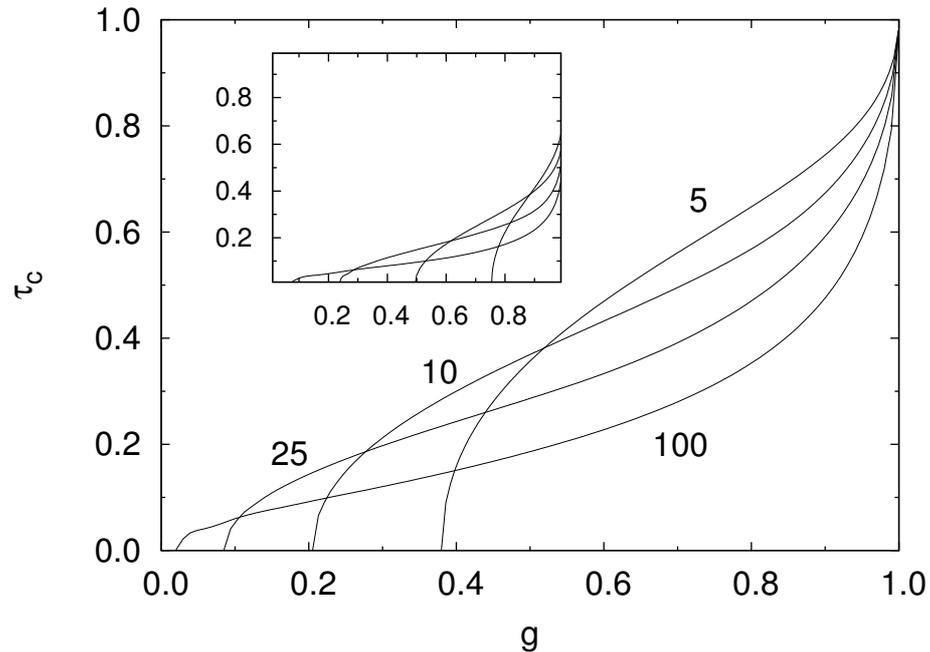


Figura 10: Valor crítico da intensidade de competição em função da pressão de cooperação para  $m = 1$  (principal) e  $m = 5$  (inset). A capacidade das vesículas está indicada no gráfico principal. Abaixo de  $\tau_c$  a coexistência é impossível.

localidade de replicação só para citar os aqui estudados. Biologicamente, a intensidade de competição nada mais é que a diferença nas taxas de replicação oriunda de diferenças nas afinidades dos substratos de replicação, atividade da replicase, entre outros fatores. A velocidade de replicação é consequência direta da atividade e da processividade (i.e., o número de cópias produzidas por unidade de tempo) da replicase. Assim, esses parâmetros tem uma interpretação muito direta. Já a seleção de grupo é muito mais facetada e pode aparecer sob vários aspectos. Aqui imaginamos que a seleção de grupo é a forma pela qual a cooperação se impõe na dinâmica. E, no caso, atua como repressão da competição promovendo a eliminação das vesículas não-balanceadas. Não é fácil imaginar um paralelo em evolução pré-biótica para tal mecanismo. Estudos recentes tem demonstrado que certos tipos vesículas lipídicas de diversas formulações são plenamente capazes de estabelecerem uma dinâmica própria de incorporação de recursos (e.g., fosfolipídeos livres), replicação [40, 41] e mesmo

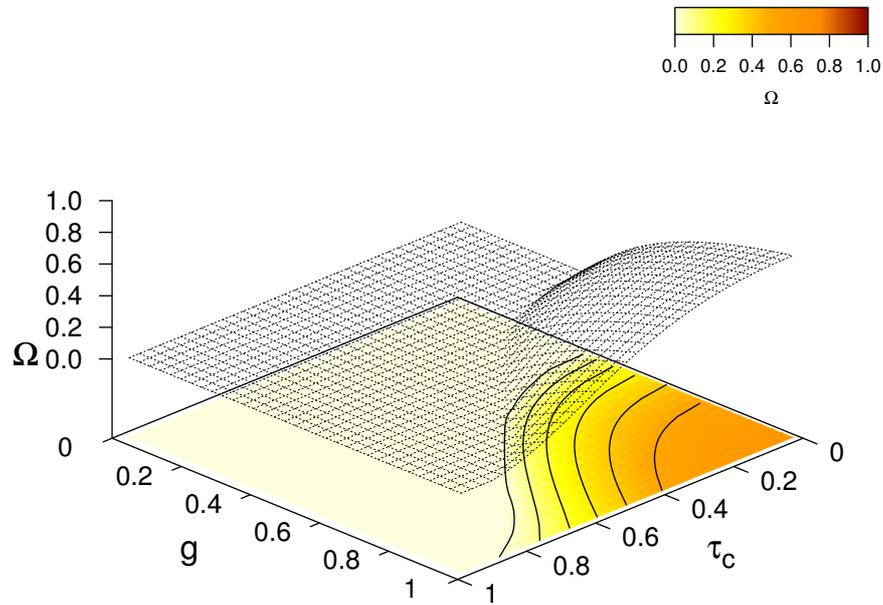


Figura 11: Colina de coexistência - superfície crítica de coexistência - acima desta a coexistência é impossível. As curvas de nível projetadas na base do gráfico são para  $\Omega = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5$  (da esquerda para a direita) com  $N = 5$  e  $m = 1$ .

de competição competição [62]. Logo, as vesículas conseguem “sobreviver” bem na ausência de qualquer aparato genético. Isso é um problema para o cenário aqui proposto pois, idealmente, a seleção de grupo só opera devido a composição genética de cada vesícula na população. Para tanto, é preciso que a replicação gênica, de alguma forma, promova a replicação das vesículas com maior diversidade gênica. Até o término desta dissertação não pudemos encontrar um exemplo onde a replicação de vesículas lipídicas esteve vinculada aos seu conteúdo molecular. Entretanto, tentativas de se construir sistemas experimentais com essas características estão já bem encaminhadas e muito em breve teremos resultados mais concretos disponíveis [63]. De qualquer maneira, até o momento não há como propor alternativas efetivas de modelos em evolução pré-biótica que tenham resultados promissores sem a presença (implícita ou explícita) de cooperação. É curioso imaginar, tomando como base essas vesículas independentes de genética, um sistema tipo corretor estocástico construído

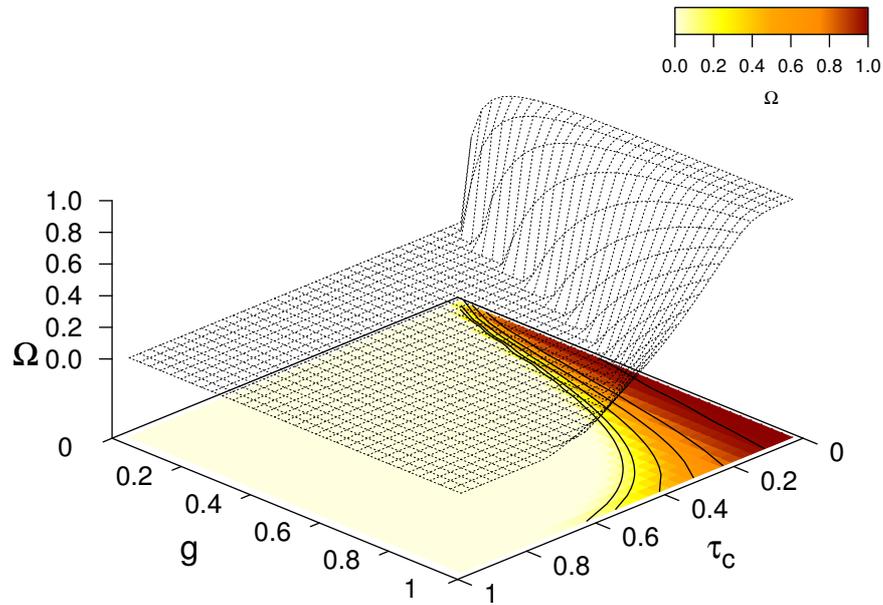


Figura 12: Penhasco de coexistência - superfície crítica de coexistência - acima desta a coexistência é impossível. As curvas de nível projetadas na base do gráfico são para  $\Omega = 0.1, 0.2, 0.4, 0.6, 0.8$  (da esquerda para a direita) com  $N = 100$  e  $m = 1$ .

com essas exigiria uma intensidade de seleção de grupo mínima, apenas o suficiente apenas para compensar os efeitos iniciais da deriva e da competição. Por outro lado, para termos vesículas com grandes populações de genes estes devem ser todos competidores ineficazes ( $\tau$  pequeno). Mas, é claro que nessa situação a seleção natural operaria dentro de cada vesícula, tornando algum tipos de genes competidores mais eficazes e, conseqüentemente, eliminando os outros por exclusão competitiva. E é exatamente neste ponto que reside a vantagem deste tipo de formulação. Como o competidor está fixado e isolado em uma vesícula e como, por algum mecanismo, a seleção de grupo é capaz de perceber vesículas assim configuradas, estabelecemos uma forma de reprimir a propagação deste mutante mais eficaz na população de vesículas. Para dar um ponto final, ressaltamos que o objetivo de toda a teoria em evolução pré-biótica é nada mais que estabelecer um cenário razoável com mecanismos plausíveis pelos quais fiquemos convencidos de que é possível sairmos de uma

realidade abiótica para o domínio da biologia plena. Porém, cada vez que nos atemos à questão mais complexa esta se apresenta e novas sutilezas do processo fazem parecer ainda mais improvável o fato de existir a biosfera tal esta é hoje. A maior contribuição exposta aqui foi levantar e dar um pouco mais de atenção ao complexo problema da coexistência. Em especial, reforçamos a transição para a vida celular tem que enfrentar um desafio ainda maior que aquele exposto na crise primordial da informação genética, segundo Eigen [7]. O aparecimento de genes compartimentalizados dá início a era da deriva genética e com ela uma nova crise informacional tem início. Uma nova crise, ainda sujeita a problemas como o limiar de fidelidade de replicação e a exclusão competitiva mas ainda mais pernicioso. Essa nova crise é capaz de atuar na ausência de competição e apresenta aos menos dois limiares. É a genética se equilibrando entre Cila e Caríbdis, entre a deriva e a exclusão competitiva.

## 3 DINÂMICA DE VESÍCULAS II

*“Computers are to biology what mathematics is to physics”*

Harold J. Morowitz

### 3.1 Um modelo computacional

No capítulo anterior, analisamos os principais desafios que devem presentes nos primeiros passos da evolução para organismos celulares. Vimos também que a deriva genética desponta como o maior problema a ser enfrentado pelas protocélulas ou qualquer outro sistema genético confinado em compartimentos. Porém, o modelo anteriormente explorado possui algumas limitações importantes que devem ser ressaltadas. Primeiramente, como a população de vesículas é infinita, perdemos a chance de observar possíveis efeitos devidos ao tamanho da população. Além disso, ficamos praticamente impossibilitados de ver extinções de populações inteiras (linhagens), a não ser em casos mais patológicos como, por exemplo, na presença de competição e seleção de grupo tendendo à 1. Em segundo lugar, deixamos de lado a possibilidade de mutação. Terceiro, não permitimos variações na capacidade das vesículas, nem mesmo flutuações desta. E, por fim, usamos apenas dois genes distintos o que é muito pouco para conclusões mais amplas sobre questões de coexistência. É possível encontrar na literatura modelos mais gerais com as mesmas características. Por exemplo, no modelo proposto por Fontanari e colaboradores [43], explora-se uma situação muito similar à anteriormente apresentada, mas, com uma diversidade maior de genes e permite-se mutações. Mesmo assim, ainda é usada a premissa de uma população infinita de vesículas com capacidade fixa. Logo, nada mais acertado que

tentar agregar um pouco mais de realismo as conclusões anteriores usando um modelo com menos exigências. E nada melhor que uma simulação de uma dinâmica como a anterior, ao menos em seus princípios. Na verdade, vamos buscar inspiração um pouco mais atrás na história dos modelos de evolução pré-biótica.

### 3.2 Compartimentos: o modelo de Bresch *et al.*

No início dos anos 1980, um modelo se destacou como sendo a primeira alternativa real ao modelo de hiperciclos. Denominado simplesmente modelo de compartimento (*package model*), proposto e sucintamente explorado por Bresch e colaboradores [27, 30], este tem sido uma importante referência nas discussões sobre estabilidade da informação genética. Por ser um modelo estocástico de populações com tamanho finito, estão presentes todos os efeitos relacionados, i.e., a deriva genética, extinções de linhagens, flutuações, etc. Entretanto, o modelo apresenta alguns problemas que reduzem consideravelmente o impacto dos seus resultados. A concepção do modelo é bastante simples. Temos uma população inicial de vesículas com um certo número de genes de  $d$  tipos distintos. À época, estes genes eram considerados moléculas de RNA à maneira imaginada por Eigen e colaboradores em seus trabalhos sobre hiperciclos [23, 24, 25]. Entretanto, com uma diferença sutil mas crucial: a afinidade dos RNAs é não-específica. Ou seja, diferentemente da formulação dos hiperciclos aqui a função de replicase pode ser desempenhada por apenas um único tipo de molécula ou por todas. Os compartimentos são pensados como sendo feitos de algum material (não especificado no modelo) que os torna capazes de serem fissionados por forças não-específicas. Dessa forma, cria-se uma maneira de reproduzir os compartimentos. Durante a fissão, idealmente, o conteúdo do compartimento é distribuído aleatoriamente entre os compartimentos resultantes. Porém, é estritamente necessário para a replicação no interior do compartimentos a presença de ao menos uma cópia de cada um dos  $d$  tipos de genes. Caso contrário, aquele compartimento estará condenado, pois não irá possuir a capacidade de replicação. Essa é exatamente a mesma receita usada pelo modelo do corretor estocástico, anos mais tarde. Olhando do

ponto de vista aqui apresentado, essa é uma forma de seleção de grupo idêntica àquela imposta pelo parâmetro de seleção de grupo do modelo do capítulo anterior. Curiosamente, na descrição de 1981 do modelo de compartimentos, [30] os autores consideraram que a seleção natural como fator evolutivo prevalente no modelo. Em nenhum momento os autores citam explicitamente idéias que remetam à seleção de grupo ou cooperação. Mas, é claro que formulado desta maneira o modelo implica em cooperação através de uma forma severa de seleção de grupo por repressão de competição. Em outras palavras, o fato do modelo exigir a presença de ao menos um gene de cada tipo para garantir a capacidade de replicação, efetivamente, impede a fixação de qualquer gene, seja por competição seja por deriva. Além de tudo, são permitidas mutações e acidentes. As mutações permitidas no modelo levam genes funcionais a se tornarem genes parasitas ou genes letais. Os parasitas não contribuem para a replicação e os letais, como o próprio nome indica, destroem completamente o compartimento. Não são permitidas mutações reversas. Por fim, os acidentes nada mais são que a eliminação não-específica de compartimentos, imitando catástrofes naturais. Mas, apesar dessa formulação simples e criativa o modelo peca na implementação da simulação da dinâmica populacional. A versão apresentada em Niesert e colaboradores [30] é bastante simples. Temos três fases na dinâmica: processamento, avaliação e fissão. Durante o processamento, um gene é escolhido aleatoriamente dentro de uma dada compartimento. Este passa pelos testes de mutação de verificar se irá ou não se tornar um parasita ou letal. Logo após, o gene original e sua cópia (mutante ou não) são devolvidos à população de genes daquele compartimento (i. e. sorteio com reposição). Esse processo é repetido um número de vezes equivalente à processividade da replicase. Na formulação, a processividade é representada pelo parâmetro *NORM*, i.e., *Number Of Replicated Molecules*. É importante ressaltar que a processividade é o parâmetro crucial do modelo, já que esta determina o tamanho esperado das populações de genes em cada compartimento. Após o esgotamento da processividade da replicase naquela vesícula, repete-se o processo nas outras vesículas da população. Na próxima fase, cada vesícula é avaliada em relação ao seu conteúdo

gênico. Isso também lembra a forma pela qual implementamos a seleção de grupo no modelo do capítulo anterior. Mas, a avaliação aqui se processa de uma forma diferente. Avalia-se a presença de genes letais no compartimento e a ausência de algum gene essencial. Caso algum destes eventos aconteça o compartimento é desmantelado e passamos para o próximo e repete-se o processo. Isso é feito com toda a população. Os compartimentos que sobreviveram à avaliação passam agora para a fase de fissão. E é aqui que reside a falha do modelo. Uma falha que compromete seriamente as conclusões da simulação. Usando a limitação computacional como motivo, os autores introduzem aqui um parâmetro denominado valor prospectivo, com efeito muito similar à um parâmetro de seletividade (leia-se *fitness*), afim de evitar uma explosão populacional. Abaixo de um dado valor crítico (não especificado na simulação) o compartimento é classificado como inviável e eliminado da população. Entretanto, a maneira como esse parâmetro é determinado força desnecessariamente o resultado da simulação. Para simplificar, o valor prospectivo nada mais é que o valor de uma função cujos parâmetros são as frequências dos genes de cada compartimento (cf. [30]). Particularmente, a maneira como este é implementado impõe que compartimentos com um certo número de parasitas sejam equivalentes à um contendo gene letal. Logo, fica claro que para um dado nível de seletividade e uma certa probabilidade de mutação, iremos impor um limite máximo na carga parasitária dos compartimentos. E, com isso, perdemos parte da originalidade dos resultados, já que selecionamos *a priori* o tipo de vesículas que desejamos.. Seria muito mais interessante que estes resultados emergissem apenas dos processos microscópicos implementados na simulação. E este é exatamente o ponto aqui. Vamos reconstruir o modelo de Bresch sem apelar para nenhum artifício para controlar a dinâmica populacional.

### 3.3 Delineando a simulação

Diferentemente a implementação da simulação em [30] onde a abordagem era simular a dinâmica populacional usando-se uma matriz constantemente atualizada,

vamos usar um esquema de filas duplamente encadeadas com inserções sempre ao final da fila e deleções sem restrição de posição. Cada elemento da fila é uma estrutura representando uma vesícula e contém um vetor  $L$  que armazena as concentrações  $n_l$ , com  $l = 1, \dots, d$ , de cada um dos  $d$  tipos de genes. A replicação é, explicitamente, um processo cooperativo. Idealmente, a replicase forma-se como resultado da junção de todos os tipos de genes. Ou seja, se temos  $d$  tipos gênicos distintos e funcionais, a replicação só será possível se houver no interior de uma dada vesícula ao menos uma cópia de cada dos  $d$  tipos de genes, i.e., se  $\prod_{l=1}^d n_l \geq 1$ . Caso contrário, esta vesícula está condenada. Isso é factível se imaginarmos que a replicase primordial se comporta-se como o ribossomo moderno. O ribossomo atual é composto largamente de RNA e sua principal função catalítica é desempenhada por RNAs. As proteínas do ribossomo parecem possuir funções estruturais e/ou regulatórias. O ribossomo também é automontado. Logo, o cenário proposto aqui é totalmente condizente com as premissas do *RNA world* [32]. De qualquer forma, essa exigência de ação conjunta dos genes nada mais é que uma forma natural de seleção de grupo via repressão de competição. Vamos permitir mutações durante o processo de replicação. No caso, não há muito interesse em permitir que haja mutações para tipos distintos de genes funcionais, já que isso apenas facilitaria a coexistência entre estes. Por isso, as mutações aqui permitidas serão para um tipo-não funcional de gene que não contribui para a replicase primordial. Dessa forma, qualquer mutação leva ao mesmo tipo de gene não-funcional. As implicações são claras. Como todos os outros sofre mutação para este tipo, então a frequência de replicação deste é maior que a frequência de replicação de qualquer outro tipo de gene num cenário de concentrações iniciais iguais. Este mutante é um competidor mais eficiente. A replicação de uma dada vesícula não está, a princípio, vinculada ao seu conteúdo gênico. Isso ficará mais claro logo adiante. Está estabelecido o conflito. A simulação consiste nos seguintes estágios:

- Replicação - neste estágio cada vesícula da população tem seus genes replicados. Seleccionada uma vesícula, escolhemos aleatoriamente um de seus genes

e o replicamos. Durante a replicação podem ocorrer mutações com probabilidade  $u$ . Todos os mutantes (genes do tipo  $p$ ) são não-funcionais, ou seja, não contribuem para o processo de replicação. Ambos genes, molde e cópia, são devolvidos à população daquela vesícula. Repete-se, então, o processo com esta nova população de genes. Isso é feito até que esgotar-se a processividade  $\Lambda$  da replicase. A processividade é, literalmente, o número de cópias gênicas que a replicase consegue produzir em um intervalo de tempo de simulação. Esse processo é feito em todas as vesículas da população.

- Avaliação - durante este processo cada vesícula da população é avaliada em relação ao seu conteúdo gênico. Aquelas que não possuem aos menos uma cópia de cada gene funcional são eliminadas da população. Essa eliminação é apenas por motivos técnicos. Como durante a fissão o conteúdo das vesículas parentais é dividido entre as vesículas descendentes, aquelas incapazes de replicação irão diluir o seu conteúdo até ficarem totalmente vazias. Estas vesículas condenadas são removidas da população e este é o único meio de eliminá-las.
- Fissão - cada vesícula que sobreviveu à avaliação dá origem a exatamente duas novas vesículas. O conteúdo da vesícula parental é distribuído aleatoriamente entre suas duas descendentes. O processo é repetido em cada vesícula.

Um intervalo de tempo na simulação consiste num ciclo com estes três processos. Fica claro que não há sobreposição de gerações. Também, é importante dizer que, nessas condições, a linhagem corre risco de extinção total. Além disso, devido ao processo de replicação e de fissão as vesículas estão duplamente arriscadas à perda de genes essenciais. Todos os ingredientes que desejamos estão presentes. Mas, há um pequeno detalhe aqui que deve ser melhor explicado. No processo de replicação, foi dito que a replicase se forma da junção dos genes contidos na vesícula, daí a necessidade de uma cópia de cada, ao menos. Só que a replicase, por definição, usa os próprios genes como substrato, ou seja, como moldes. Então, no mínimo precisaríamos de uma cópia de cada gene para compor a replicase mais uma cópia servindo de

molde. Na verdade, a situação é mais complicada do que realmente parece. Em tese, cada conjunto de  $d$  genes distintos poderia formar uma replicase. Se, por exemplo, uma dada vesícula possuísse dez cópias de cada gene esta deveria ser capaz de produzir dez replicases e ter uma processividade acumulada dez vezes maior que uma vesícula com apenas uma replicase. Mas, nessa situação, as dez replicases consumiriam todas as cópias e não restaria nenhuma como molde. Pior ainda, só os genes não-funcionais se beneficiariam. Todos esses detalhes tornariam a simulação excessivamente complicada e certamente obscureceria a análise de resultados, tornando mais difícil relacionar as possíveis causas de extinção ou de sucesso, por exemplo. Por isso, vamos simplificar a situação. Iremos nos contentar com apenas uma cópia de cada gene para garantir a replicação. Isso não é tão restritivo quanto exigir ao menos duas cópias de cada gene mas não deverá mudar qualitativamente os resultados da simulação. Os detalhes em relação à replicase serão ignorados. Para todos os efeitos, apenas uma replicase estará em ação. Por fim, devido à natureza estocástica da simulação e em razão da imaterialidade da escala temporal estamos em total acordo com as premissas de uma simulação estilo Monte Carlo. A descrição da simulação sugere que teremos um crescimento/decrescimento exponencial da linhagem, i. e., teremos um processo de ramificação. Assim sendo, não será possível usar o mesmo número de amostras em todas as rodadas de simulação. Mas, vamos nos assegurar para que isto não comprometa os resultados e a análise.

### 3.4 Coexistência I - a ação da deriva

Como de praxe, começamos nossa análise pela ação da deriva na dinâmica de vesículas. Novamente, para isolarmos os efeitos da deriva, restringimos os parâmetros da simulação. Nessa situação, não há mutação e iremos observar o efeito populacional de acréscimos na processividade da replicase. A processividade é, pragmaticamente, a medida da capacidade média das vesículas. Independentemente do valor inicial da população de genes, a população média de genes em cada vesícula deve convergir rapidamente para algo muito próximo de  $\Lambda$ . Por exemplo, seja  $n_t$  o número esperado

total de genes em uma dada vesícula no tempo  $t$ . Assim sendo, fica claro que a relação entre o número esperado total de genes da vesícula parental no tempo  $t - 1$  e desta é dado pela simples recorrência,

$$n_t = (n_{t-1} + \Lambda)/2 \quad (8)$$

cujas soluções são igualmente simples,

$$n_t = (n_{t-1} - \Lambda)/2^t + \Lambda. \quad (9)$$

Logo, quando  $t \rightarrow \infty$  implica em  $n_\infty = \Lambda$ . Na prática, atingimos algo muito próximo desse valor quase exponencialmente. Como consequência, independentemente da condição inicial, iremos ter rapidamente uma linhagem com vesículas de capacidade homogênea, em média. Os desvios em relação a este valor são devidos puramente ao processo de fissão.

Partindo para a exploração do modelo, o primeiro efeito da deriva que esperamos encontrar é a exigência de um acréscimo na processividade afim de garantir a sobrevivência de linhagens com diversidade sucessivamente maiores. A primeira vista isto é óbvio. Se temos  $d$  genes diferentes na população, precisamos uma processividade no mínimo igual a este valor. Mas, se a deriva for capaz de influenciar na dinâmica, então, a essa exigência deve aumentar de modo não-trivial. Podemos conferir esse efeito na figura 13.

Foram determinados os valores de  $\Lambda_c$  que garantem a sobrevivência de uma linhagem com  $d$  genes distintos. Em outras palavras, o valor de processividade para o qual a probabilidade de sobrevivência  $P_t$  torna-se constante, no longo prazo ( $t \gg t_c$ ). Como previsto, conforme aumentamos a diversidade amplificamos a ação da deriva. Empiricamente, o valor de  $\Lambda_c$  vai aproximadamente com  $d^2/2$ . Essa estimativa foi obtida a partir dos dados ajustados por uma função do tipo lei de potência,

$$\Lambda_c(d) = Ad^\lambda \quad (10)$$

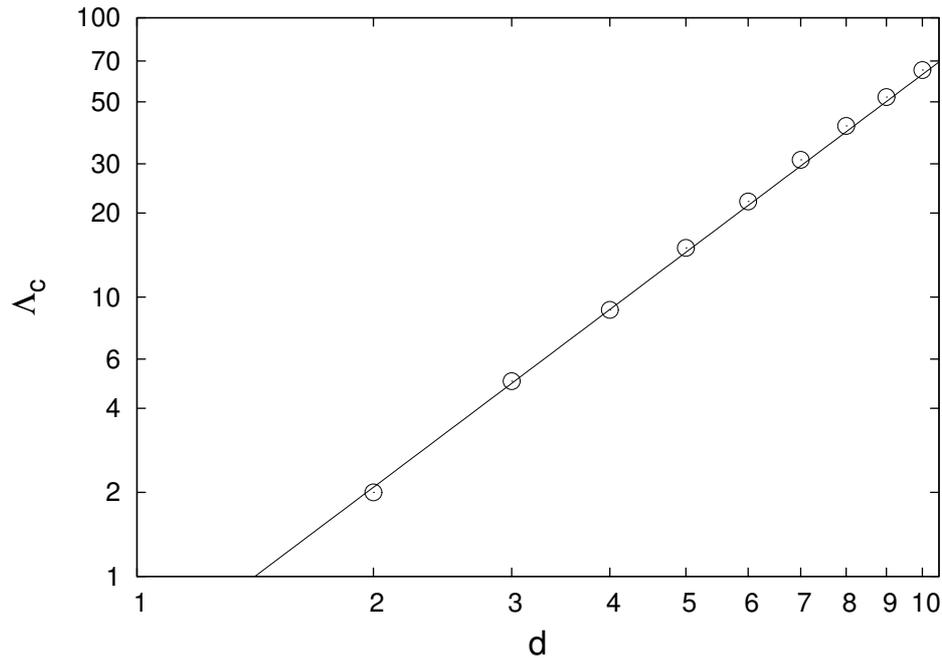


Figura 13: Valor crítico da processividade ( $\Lambda_c$ ) em relação a diversidade de genes ( $d$ ). A reta representa a função  $\Lambda_c(d) = Ad^\lambda$ , com  $A \sim 0.5$  e  $\lambda \sim 2$ . Abaixo da reta, a coexistência é impossível.

onde o ajuste foi feito usando-se mínimos quadrados não-lineares segundo o esquema de minimização de Marquardt-Levenberg. Em termos práticos, evidencia-se que o efeito da deriva aumenta radicalmente com o aumento da diversidade gênica. Mas, não é só isso. A deriva também afeta o vigor (i. e., o potencial biótico) da população. Como dito na definição da simulação, a única fonte de mortalidade é a perda de algum tipo essencial de gene. A deriva nada mais é que a perda estocástica de um dado tipo de gene. Então, a maneira como uma linhagem cresce ou decresce está intrinsecamente relacionada à intensidade da deriva. Um aumento na diversidade gênica não apenas implica em uma exigência maior de processividade como também em um crescimento populacional menos vigoroso. Para assegurar-nos de este é realmente o caso, obtivemos diversas curvas de densidade populacional média  $\langle N \rangle_t$  para valores distintos de processividade (e. g. fig. 17).

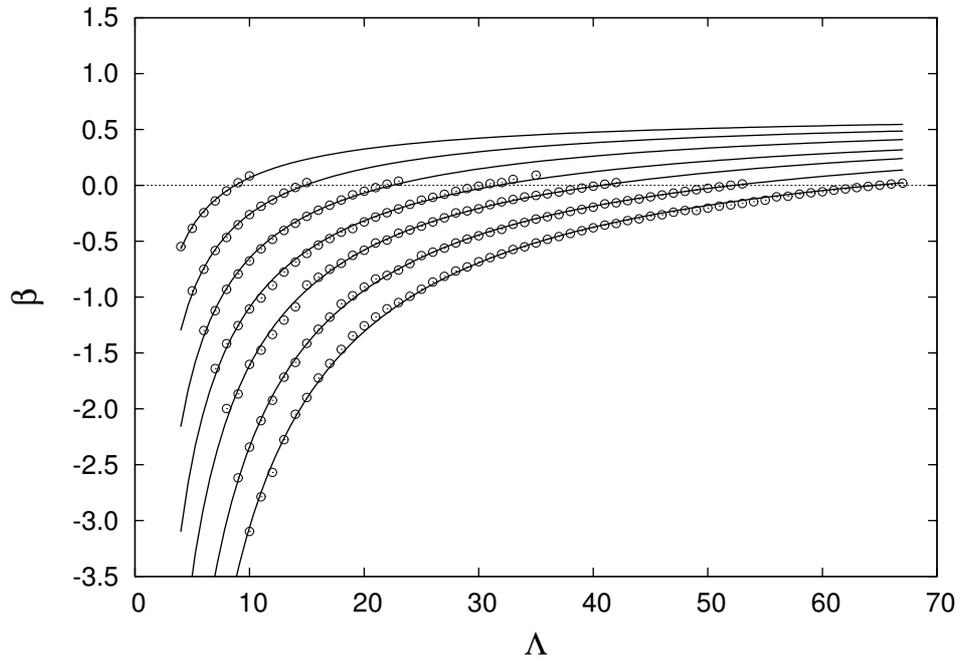


Figura 14: **Varição dos expoentes de crescimento populacional como função da processividade  $\Lambda$  em populações livres de mutação para  $d = 4, \dots, 10$  da esquerda para a direita. A linha zero marca a transição de populações decrescentes para populações crescentes.**

As curvas resultantes mostram um processo exponencial de crescimento/decrescimento populacional, como esperado em qualquer processo de ramificação. Logo, é fácil obter as taxas de crescimento intrínsecas,  $\beta$ , da população usando o mesmo esquema que fizemos anteriormente, só que usando a exponencial

$$\langle N \rangle_t = N_0 e^{\beta t} \quad (11)$$

no lugar da lei de potência. A constante  $N_0$  obtido apenas reflete uma estimativa do tamanho inicial da população caso o processo partisse de uma população representativa. Como em todas as rodadas de simulação linhagens sempre começam com um único indivíduo, vamos ter um crescimento/decrescimento muito mais vigoroso no

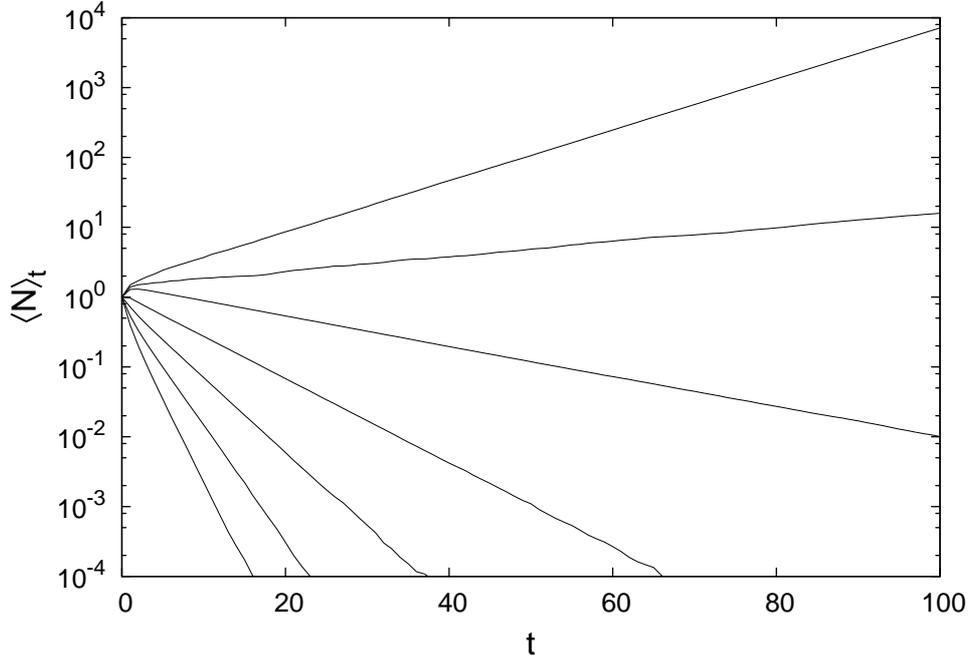


Figura 15: **Densidade populacional esperada em uma linhagem com  $d = 4$ ,  $u = 0$  e  $\Lambda = 4, 5, \dots, 10$ , de baixo para a cima.**

início do processo até atingirmos aquela situação prevista anteriormente para cada vesícula. Este efeito devido à condição inicial desaparece rapidamente e pode ser seguramente descartado durante o ajuste das curvas. Confira nas figuras 15, 16. A figura 14 é uma demonstração clara deste aspecto. Essas curvas empíricas são o resultado do ajuste da função

$$\beta(\Lambda) = \beta_{max} - C_1 \Lambda^\theta \quad (12)$$

aos pontos obtidos. Podemos notar que os expoentes  $\theta$  das curvas apresentam um acréscimo em módulo, ou seja, as curvas vão se tornando mais suaves, já que, no máximo, poderemos ter linhagens que dobram de tamanho em cada intervalo de tempo (i.e.,  $\beta = \beta_{max} = \ln(2)$  para  $\Lambda \rightarrow \infty$  e  $u = 0$ ). Entretanto, todas possuem pontos acima de zero. O zero marca, seguramente, o valor de taxa de crescimento acima do qual a probabilidade de sobrevivência de uma dada linhagem é constante para  $t$  suficientemente grande. De qualquer forma, demonstra-se que o vigor de crescimento populacional também é alvo da deriva e de uma forma não-trivial. Linhagens

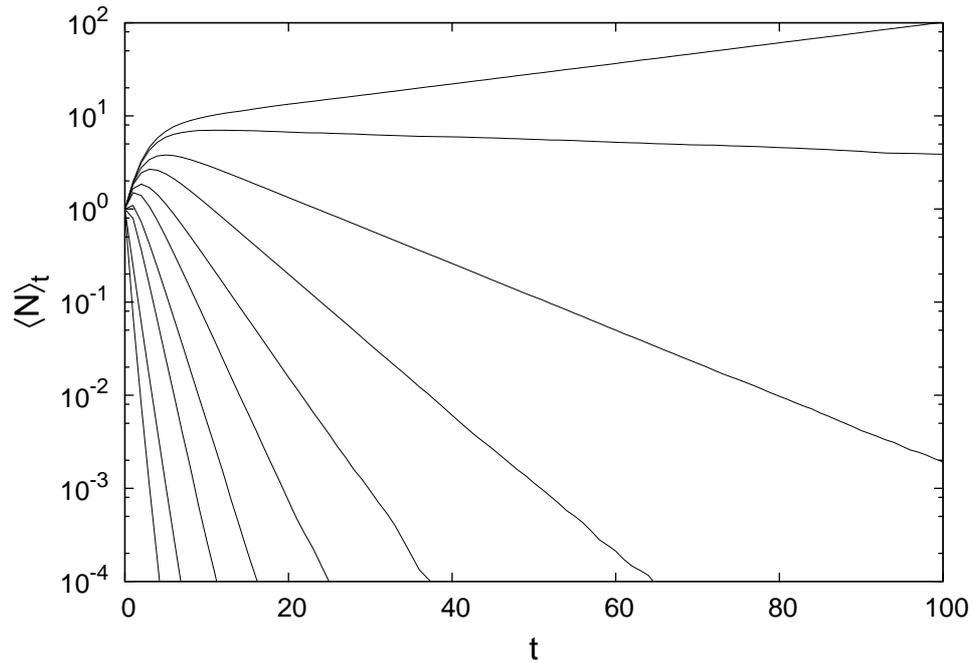


Figura 16: **Densidade populacional esperada em uma linhagem com  $d = 8$ ,  $u = 0$  e  $\Lambda = 8, 12, 16, \dots, 42$ , de baixo para cima.**

com diversidade gênica grande devem possuir elevado grau de redundância para garantir o mesmo vigor de crescimento que uma população com menor diversidade. Por fim, podemos notar que o número esperado de vesículas viáveis produzidas por vesículas por geração,  $\mu$ , deve ser uma função complicada dos parâmetros do modelo. Na ausência de mutação, podemos dizer que  $\mu(d, \Lambda_c) = 1$ , de acordo com o que vimos anteriormente. Logo, para um dado  $d$ , temos duas fases no modelo. Para  $\Lambda > \Lambda_c$ , a linhagem possui probabilidade de sobrevivência  $P_t > 0$  com  $P_t = C_{d,\Lambda}$ . E, se  $\Lambda \leq \Lambda_c$  esta se extingue com probabilidade 1 para  $t$  suficientemente grande. De qualquer forma, fica claro que podemos encontrar condições de superar a ação da deriva se pudermos garantir a processividade adequada. E isso vale para qualquer número de genes distintos.

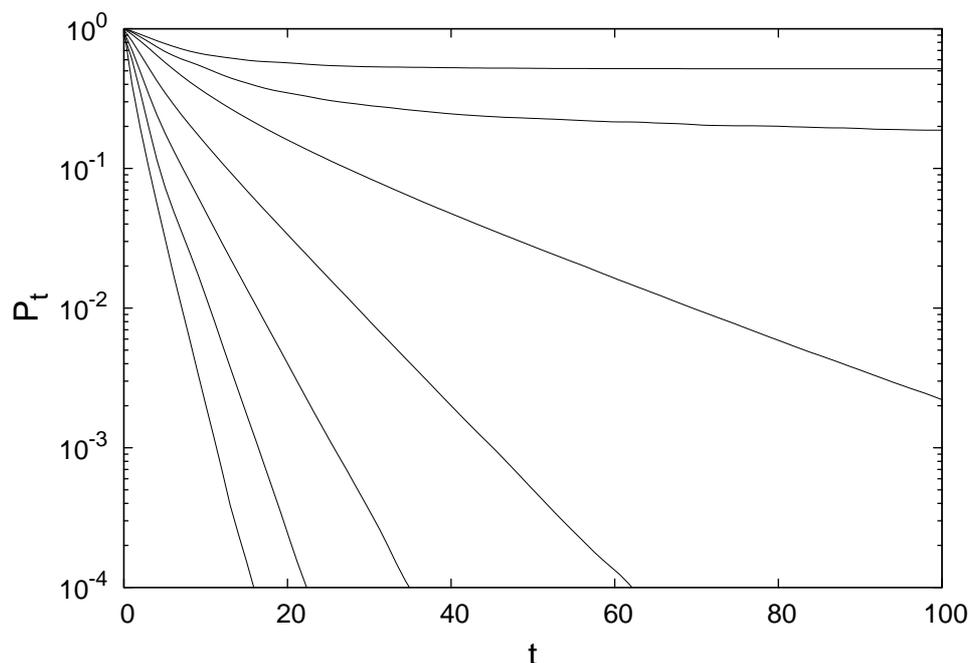


Figura 17: **Probabilidade de sobrevivência para uma linhagem com  $d = 4$ ,  $u = 0$  e  $\Lambda = 4, 5, \dots, 10$ , de baixo para cima.**

### 3.5 Coexistência II - o efeito da mutação

Passamos agora a observar o efeito das mutações na dinâmica. A maneira pela qual implementamos o processo de mutação implica em termos dois tipos qualitativamente diferentes de genes na população. Podemos classificar a variante não-funcional como sendo não-cooperativa. Por outro lado, essa variante não-funcional é, sem dúvida, um competidor mais eficiente. Logo, teremos um conflito interessante. Para garantir a própria permanência na população, os genes não-cooperativos não devem excluir genes essenciais via competição. Mas, como a replicase é não-específica, corre-se o risco desta copiar apenas os mutantes não-funcionais e, por conseguinte, condenando ao menos uma das vesículas descendentes. Além disso, como os mutantes não contribuem para a replicação, a presença destes reduz o tamanho médio da população de genes funcionais numa dada vesícula. Em outras palavras, para todos os efeitos, a presença de mutantes diminui a redundância efetiva de uma dada vesícula. Nesse sentido, uma linhagem com replicase de elevada processividade se

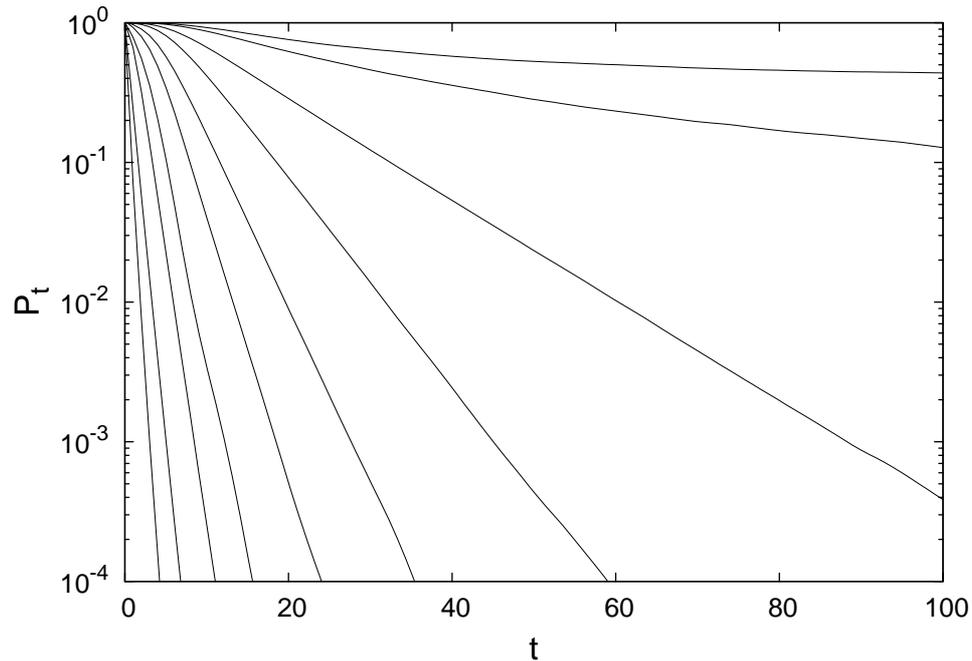


Figura 18: **Probabilidade de sobrevivência para uma linhagem com  $d = 8$ ,  $u = 0$  e  $\Lambda = 8, 12, \dots, 40, 42$ , de baixo para a cima.**

comporta como uma linhagem com replicase de processividade somente devido à probabilidade de mutação. É algo como se houvesse um desperdício catalítico. Podemos ver isso com clareza na figura 19. Se imaginarmos que a probabilidade de mutação  $u$  age como um modificador da processividade  $\Lambda$ . Quanto maior for  $u$  menor será  $\Lambda$  usada, efetivamente, para copiar genes funcionais. Mas, isso irá afetar a linhagem de um outro modo além do citado acima. Como visto anteriormente, linhagens com processividade reduzida são mais suscetíveis a ação da deriva. E populações com probabilidade de mutação elevadas tem a processividade efetiva reduzida. Logo, estas devem ser mais sensíveis a ação da deriva. Isso fica claro se compararmos as figuras 15 e 20.

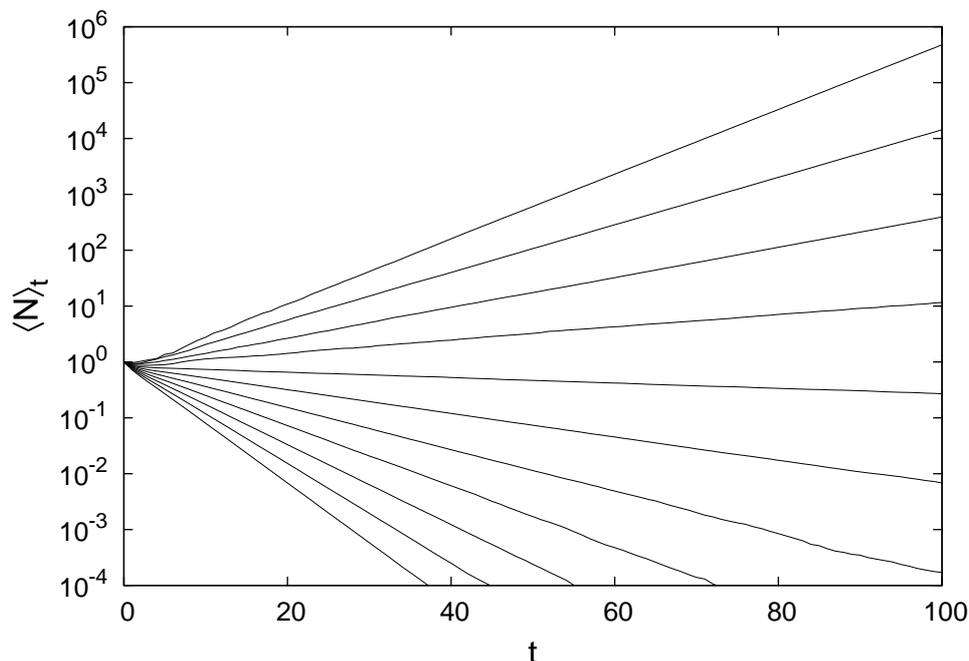


Figura 19: **Densidade populacional esperada de para uma linhagem com  $d = 2$ ,  $\Lambda = 2$  e  $u = 0.00, 0.05, \dots, 0.50$ , de cima para baixo.**

Criamos um círculo vicioso. A probabilidade de mutação aqui, então, é duplamente danosa para a população. Num sentido, cria uma forma severa de competição da qual apenas o mutante não-funcional pode sair beneficiado. E, em outro, possibilita uma ação intensa da deriva genética. Quando observamos temporalmente a fração esperada de genes não-funcionais por vesícula,  $\langle n \rangle_p$ , como ilustrado na figura 22, percebemos que, independentemente da probabilidade de mutação definida, a linhagem torna-se homogênea também na sua composição interna. O mesmo pode ser observado com fração esperada de um dado gene funcional (e.g.  $\langle n \rangle_1$ ), como vemos na figura 23. Isso significa que no começo do processo temos muitas vesículas com poucos genes mutantes. Mas, rapidamente, estes se espalham pela população. E essa é uma situação sem retorno. Diferentemente do modelo do capítulo anterior, neste caso a única forma de seleção de grupo está na exigência da presença de todos os tipos gênicos para a efetivação da replicação. E é exatamente neste ponto que reside a catástrofe do modelo. Naquele modelo, havia uma exigência maior em relação ao

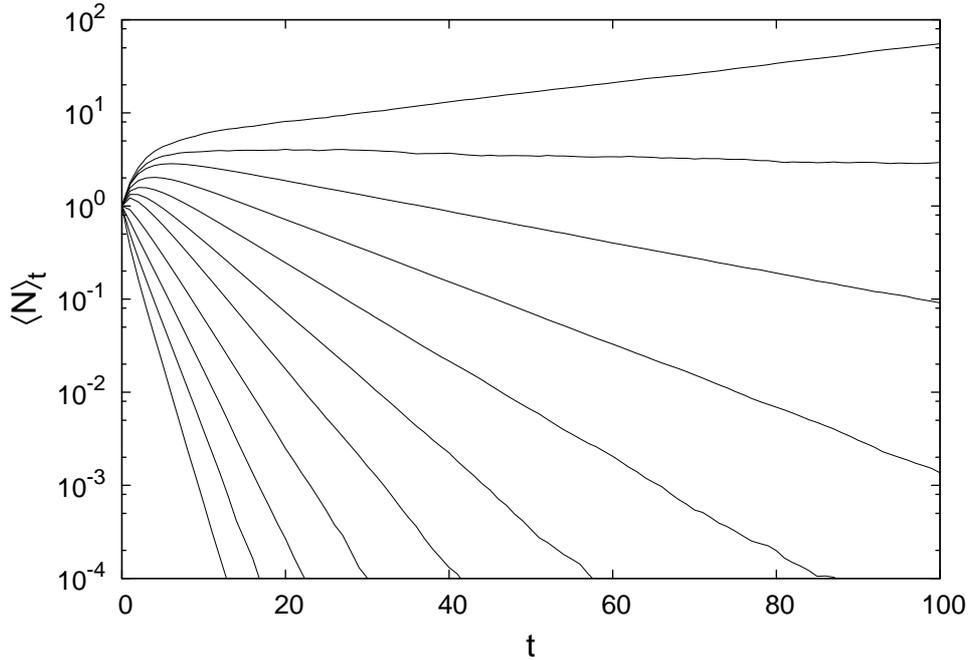


Figura 20: **Densidade populacional esperada de para uma linhagem com  $d = 4$ ,  $\Lambda = 4, 5, \dots, 14$  e  $u = 0.10$ , de baixo para a cima.**

conteúdo das vesículas. Havia um vínculo entre a probabilidade de sobrevivência e o conteúdo gênico. Aqui a situação é diferente. A exigência está relacionada a capacidade de replicação. Ou seja, enquanto que no modelo do capítulo 2 a seleção de grupo primava pelo tipo ideal de vesícula, aqui a dinâmica interna leva a um tipo primariamente determinado pela probabilidade de mutação. No fim, para uma linhagem relativamente grande, teremos a dominância um tipo de vesícula com aquelas frações esperadas ilustradas anteriormente. Se estas vesículas tiverem uma processividade efetiva,  $\bar{\Lambda}$ , menor que  $\Lambda_c$  calculado anteriormente, a linhagem estará fadada à extinção. Na prática,  $\bar{\Lambda}$  nada mais é que  $\Lambda \sum_l^d n_l$ . Por exemplo, para  $u = 0.5$  e  $\Lambda = 2$  temos que  $\Lambda_c = 2$  e  $\bar{\Lambda} \sim 1.4$ . Logo,  $\bar{\Lambda} \leq \Lambda_c$  e a linhagem é inviável. Infelizmente, até o término desta dissertação, não foi possível estabelecer nenhuma forma analítica de determinar  $\Lambda_c$ , nem a distribuição de equilíbrio  $\langle n \rangle_t$ .

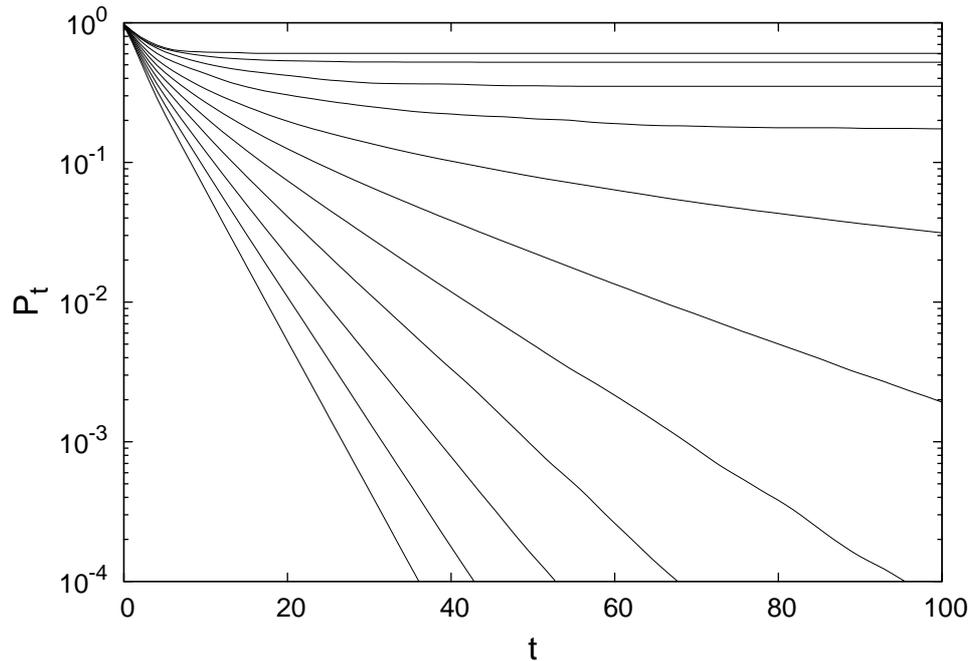


Figura 21: **Probabilidade de sobrevivência para uma linhagem com  $d = 2$ ,  $\Lambda = 2$  e  $u = 0.00, 0.05, \dots, 0.50$ , de cima para baixo.**

### 3.6 Conclusões

Garantir as condições necessárias para o início da vida na Terra é, realmente, uma tarefa espinhosa. Mais uma vez, usando uma abordagem distinta chegamos à conclusões similares às do capítulo anterior. E desta vez, a deriva mostrou ser o efeito preponderante na determinação do destino das populações de vesículas. Isso aponta para algumas conclusões e críticas. Começemos pelas críticas. Reiterando que a abordagem aqui é inspirada em trabalhos considerados clássicos da literatura em evolução. Não esperávamos encontrar novidades em uma formulação já explorada. O intento era afrouxar um pouco as restrições das propostas de Bresch e colaboradores [27, 30, 38] e dar um pouco mais de crédito às suas conclusões. Talvez a mais importante conclusão destes autores tenha sido que o número inicial de genes distintos nas protocélulas deve ter sido algo em torno de três. Porém, é difícil sustentar essa afirmação. Após confrontar as simulações apresentadas por estes com dos achados desta dissertação, nota-se que esta não se sustenta. Está muito claro

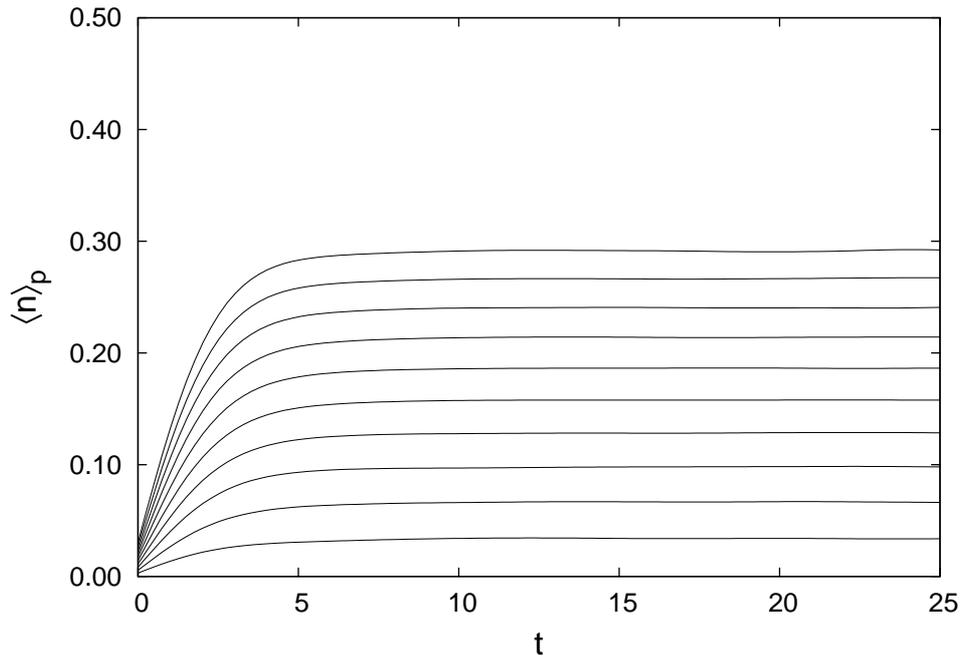


Figura 22: **Evolução temporal da fração esperada de genes não-funcionais por vesícula em uma linhagem com  $d = 2$ ,  $\Lambda = 2$ ,  $u = 0.00, 0.05, \dots, 0.50$ , de baixo para cima.**

que boa parte da razão de ser das conclusões de Bresch e colaboradores reside em critérios demasiado arbitrários. Em especial, o valor prospectivo usado em [30], por ser um critério de parâmetro livre, torna plenamente possível estabelecer condições mais restritivas quanto condições mais amenas. Isso mina definitivamente qualquer afirmativa em relação ao número inicial de genes nas primeiras protocélulas. E, por fim, definir um conjunto de condições como favoráveis com base em algo como a probabilidade de sobrevivência de uma dada amostra não é um critério convincente, pois, depende muito do intervalo de tempo observado e do número de amostras usadas. Aliás, o número usado originalmente é, certamente, insuficiente para conclusões dessa monta. Apesar de tudo, não podemos menosprezar o pioneirismo destes autores. O que encontramos aqui, baseado no que aqueles autores fizeram, sugere algumas coisas realmente importantes. Como a deriva é, sem dúvida, o maior desafio a qualquer sistema genético compartimentalizado, a demonstração (mesmo que de forma

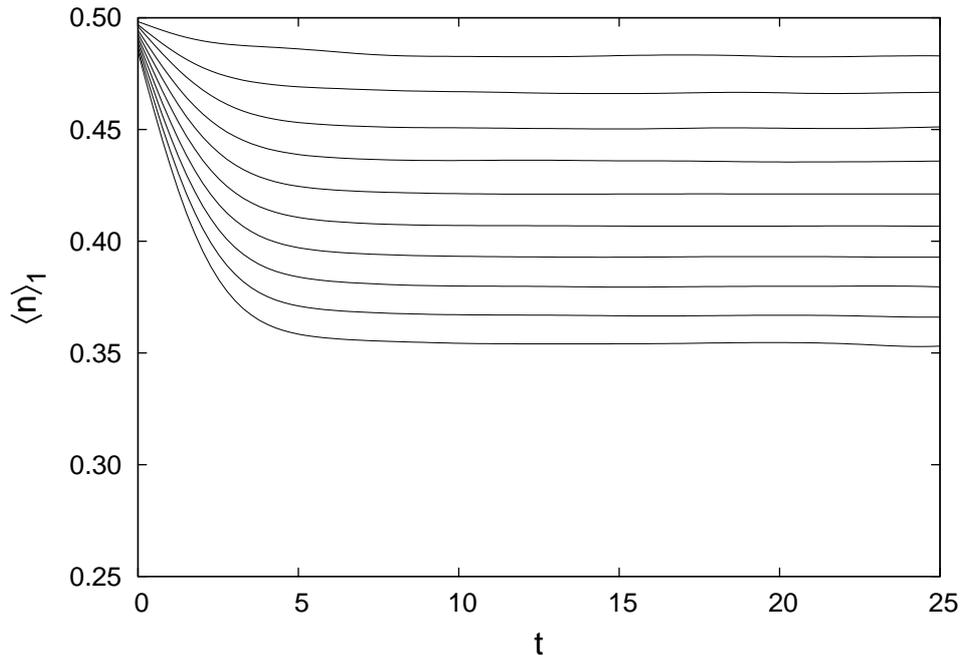


Figura 23: **Evolução temporal da fração esperada de um dado gene funcional em uma linhagem com  $d = 2$ ,  $\Lambda = 2$ ,  $u = 0.00, 0.05, \dots, 0.50$  de cima para baixo.**

muito simplificada) de exigências crescentes para se manter diversidades maiores é preocupante. No caso, demonstramos que esta exigência cresce de forma não-linear. Se imaginarmos que as medidas aqui são realistas (o que, claramente, não é o caso), um organismo protocelular com um conjunto mínimo de genes, algo em torno de 60 tipos distintos [64, 65], tem que ser capaz de comportar uma população de 1600 genes!!! Isso considerando-se apenas a deriva. Quando acrescentamos mutações ao cenário, esse número deve ser muito maior. E, como os organismos reais têm, no mínimo, mais que 480 genes, a situação é inviável. Não há como conceber uma replicase primordial com essas características. <sup>2</sup>A questão que fica é: como manter este nível de diversidade, digamos 300 genes distintos, com um número tão reduzido de cópias? No cenário proposto aqui fica claro que a simples exigência de cooperação não é suficiente para garantir tal diversidade. Como se não bastasse, deixamos claro

<sup>2</sup>A bactéria *Mycoplasma genitalium*, apesar de não tem o menor genoma em tamanho, possui somente 480 *CDSs* em seu genoma, o menor número encontrado em um organismo celular, cf. [66]

que a ação da deriva mais competição reduzem drasticamente o potencial de crescimento da linhagem para um dado conjunto de parâmetros. Então, mesmo que uma dada linhagem consiga se equilibrar entre a deriva e a competição interna de seus genes, esta será incapaz de sobreviver a competição com linhagens com menor diversidade gênica e maior processividade da replicase, já que estas últimas teriam taxas de crescimento intrínsecas maiores. Logo, por princípio, a linhagem com maior diversidade (e menor taxa de crescimento) seria excluída competitivamente no longo prazo.

## 4 CONCLUSÕES GERAIS

*“Dixi, et animan levavi!”*

Nyukhin

O objetivo de todas as teorias sobre a origem da vida e evolução pré-biótica é, simplesmente, elencar os processos primordiais envolvidos na transição do mundo abiótico para a fantástica diversidade vista hoje em dia. Na realidade, a grande questão é tentar responder como é possível manter informação genética numa escala geológica de tempo. Na atualidade, resolver tais questões tem se demonstrado uma tarefa árdua. E esta dissertação, no intuito de trazer mais luz ao cenário, não consegue escapar desta sina. Os modelos de vesículas tem sido vistos como um passo inevitável do processo evolutivo como conhecemos. O que surpreende é concluir que mesmo neste cenário não temos como escapar de uma crise na informação genética. Esta não é mesma crise ensejada por Eigen e colaboradores, o que não diminui a sua importância. Estamos diante de uma crise onde a informação genética é muito mais frágil. As vesículas pré-bióticas estão sujeitas a uma catástrofe de erros, condenadas a ter de lidar com a deriva genética e não escapam da competição. E tudo ainda poder ser mais complicado. Populações de vesículas irão sofrer ao menos com a competição. Parece que estamos em um situação sem saída.

### 4.1 A emergência da deriva

Ao longo do desenvolvimento desta dissertação foi possível mostrar como a deriva genética é importante em cenários sobre a origem da vida celular. Na verdade, a deriva é uma caso particular de uma classe mais geral de processos estocásticos

em genética de populações finitas [67]. Aliás, por definição, a deriva genética nada mais é que o acúmulo sucessivo de erros de amostragem no pool genético de uma população, culminando na perda estocástica de um ou mais genes tipos de genes. O aparecimento da vida celular implica estar em constante exposição ao processo. Uma célula, do ponto de vista genético, nada mais é que um compartimento com uma população finita de replicadores. A vida celular se tornou o que é hoje porque, de alguma forma, desenvolveu mecanismos para tolerar a deriva sem pagar os altos custos de um aumento na redundância de informação genética. O que ficou claro aqui é que estes custos aumentam de forma exorbitante com o aumento da diversidade gênica, em cenários mais realistas. Porém, não podemos ignorar a ecologia da situação. Esses altos custos em redundância tornam-se proibitivos em qualquer comunidade ecológica primordial, já que não deveriam haver produtores como hoje [68]. Logo, o conflito inicial das protocélulas com a deriva genética exigiu uma solução eficiente que consumisse poucos recursos. Ainda não sabemos ao certo qual é essa solução. As explorações aqui apresentadas apontam que qualquer que tenha sido a solução encontrada esta deve ter exigido a vinculação entre o conteúdo genético da protocélula e sua capacidade de replicação. Ora, isso é a própria definição de seleção natural só que num nível de complexidade maior. É o que chamamos de seleção de grupo. Desse ponto de vista parece muito natural a emergência da seleção de grupo em evolução pré-biótica. Podemos reiterar que o fracasso prático do modelo de Bresch e colaboradores, bem como da versão deste aqui explorada, reside na ausência de um vínculo explícito deste tipo. Não basta dividir a informação essencial, é preciso garantir que sua concentração seja mantida. E é exatamente neste ponto que a seleção de grupo, na forma apresentada no capítulo 1.1, tem se mostrado crucial na elaboração das teorias sobre origem da vida e evolução pré-biótica. Fica a questão de encontrar uma forma natural de realizar aquele mesmo mecanismo.

## 4.2 Exclusão sem competição

Um outro aspecto importante visto aqui, raramente citado na literatura especializada, é a possibilidade de perda de diversidade sem competição. E como não há recombinação nem diploidia envolvidos nos cenários aqui explorados, não cabe discutir a denominada *inbreeding depression*[69]. A crise primordial da informação genética, como dito na introdução desta dissertação, se baseia em dois princípios para seu delineamento: fidelidade finita de replicação e exclusão competitiva. Desse dois elementos, o único realmente essencial para a crise é a fidelidade finita de replicação. Em sua versão original, modelo de quasi-espécies é, essencialmente, imune à deriva genética. Por outro lado, qualquer versão de população finita deste irá imediatamente estar sujeita à deriva genética. Ou seja, podemos abdicar da exigência de competição e ainda assim teremos perda de informação genética, sob certas condições. Lembrando do caso de vesículas independentes do capítulo 1.1, temos ali um exemplo de como isso é possível. A realização deste fato fica ainda mais evidente nas simulações do capítulo 2.2. Na ausência de mutações precisamos de uma processividade mínima para preservar todos genes em cada vesícula. É notável o fato destes genes serem indistintos no que tange à replicase. Isso é equivalente a dizer que todos eles são competidores igualmente eficientes. Nessa mesma situação, o modelo de quasi-espécies garante a sobrevivência de todos os tipos de replicadores. O caso não é bem esse aqui. Sem uma processividade mínima aliada à exigência da presença de todos tipos de genes a população se tornaria uma distribuição uniforme de vesículas com apenas um único tipo de gene em cada. Este é o outro lado da deriva amplamente explorado na teoria neutra de evolução molecular. A fixação de um dado tipo gênico numa população finita, na ausência de quaisquer outros processos fora replicação, é uma questão de tempo apenas.

### 4.3 Uma perspectiva

No fim, mesmo com esse cenário desolador apresentado não há motivo para desistirmos de questões tão essenciais quantos as suscitadas aqui. Os modelos sobre como a vida começou na Terra ainda não estão maduros o suficiente para nos darem um cenário convincente. A vida encontrou um forma de superar todos estes obstáculos que parecem intransponíveis aos nossos modelos. O real intuito de qualquer modelo nada mais é que tentar eliminar todas as complicações e complexidades que for possível e captar aquilo que há de essencial em um processo ou fenômeno. O estudo da evolução parece desafiar esse paradigma de modelagem. Quanto mais a estudamos mais ingredientes essenciais aparecem. Qualquer modelo em evolução pré-biótica que busque ser minimamente completo precisa lidar com competição, mutação, diversidade e deriva. Esta dissertação e a literatura recente sugerem que mais um ingrediente deve ser acrescentado à lista: a cooperação. E, quem sabe, este não seja o ingrediente que estava faltando. O que mais nos fascina no mundo natural, a luta pela sobrevivência, nada mais é que uma versão mais sofisticada daquele que consideramos o conflito essencial: a busca por condições de coexistência. No mais, ainda existem muitas questões a serem exploradas em dinâmicas de cooperação e conflito em cenários pré-bióticos. Só para citar algumas, a recombinação e a transferência lateral de genes têm sido consideradas na literatura recente como processos primordiais [65, 70, 42]. Muito pouco se sabe sobre o efeitos destas em dinâmicas como as apresentadas aqui. Enfim, há muito que fazer ainda. Não poderia ser melhor. Para nós, cientistas, a melhor perspectiva sempre é o desconhecido.

## 5 REFERÊNCIAS BIBLIOGRÁFICAS

- [1] E. Schrödinger. *What is Life?: The Physical Aspect of the Living Cell*. Cambridge University Press, New Ed edition, 1967.
- [2] F. J. Dyson. *Origins of Life*. Cambridge University Press, 2 ed., 1999.
- [3] S. A. Kauffman. *The Origins of Order : self-organization and selection in evolution*. Oxford University Press, U.S.A., 1993. A mais completa referência sobre evolução molecular e assuntos correlatos.
- [4] G. Wächtershäuser. Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci. U.S.A.*, 87(1):200–204, Jan 1990.
- [5] A. G. Cairns-Smith. The origin of life and the nature of the primitive gene. *J. Theor. Biol.*, 10(1):53–88, Jan 1966.
- [6] A. I. Oparin. *The chemical origin of life*. Dover Publications, 2003.
- [7] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, Oct 1971. Esta é a mais importante referência em evolução pré-biótica e origem da vida.

- [8] M. Eigen. Error catastrophe and antiviral strategy. *Proc. Natl. Acad. Sci. U.S.A.*, 99(21):13374–13376, Oct 2002.
- [9] O. Kedem & Y. Pilpel D. Segre, D. Lancet. Graded Autocatalysis Replication Domain (GARD): kinetic analysis of self-replication in mutually catalytic sets. *Orig. Life Evol. Biosph.*, 28(4-6):501–514, Oct 1998.
- [10] D. Segré & D. Lancet. Composing life. *EMBO Rep.*, 1(3):217–222, Sep 2000.
- [11] E. Szathmáry & L. Demeter. Group selection of early replicators and the origin fo life. *J. Theor. Biol.*, 128(4):463–486, October 1987.
- [12] A. T. C. Silva & J. F. Fontanari. Deterministic group selection model for the evolution of altruism. *Eur. Phys. J. B*, 7(3):385–392, February 1999.
- [13] J. Maynard-Smith. Models of Evolution. *Proc. R. Soc. Lond. B*, 219:315–325, October 1983.
- [14] E. R. Pianka. *Evolutionary Ecology*. Benjamin Cummings, 6 edition, 2000.
- [15] J. Swetina & P. Schüster. Self-replication with errors. A model for polynucleotide replication. *Biophys. Chem.*, 16(4):329–45, Dec 1982.
- [16] D. Alves. & J. F. Fontanari. Population genetics approach to the quasispecies model. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 54(4):4048–4053, Oct 1996.
- [17] T. Wiehe. Model dependency of error thresholds: the role of fitness functions and contrasts between the finite and infinite sites models. *Genet. Res.*, 69(2):127–136, 1997.
- [18] M.. Eigen & P. Schüster. Stages of emerging life—five principles of early organization. *J. Mol. Evol.*, 19(1):47–61, 1982.
- [19] I. Scheuring. Habitat diversity and stability in a metapopulation. *J. Theor. Biol.*, 206(2):229–233, Sep 2000.

- [20] T. Czárán & E. Szathmáry P. Szabó, I. Scheuring. In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature*, 420(6913):340–343, Nov 2002.
- [21] I. R. Epstein & M. Eigen. Selection and self-organization of self-reproducing macromolecules under the constraint of constant flux. *Biophys. Chem.*, 10(2):153–160, Sep 1979.
- [22] P. F. Stadler & J. C. Nuño. The influence of mutation on autocatalytic reaction networks. *Math. Biosci.*, 122(2):127–160, August 1994.
- [23] M. Eigen & P. Schüster. The hypercycle. a principle of natural self-organization. part a: Emergence of the hypercycle. *Naturwissenschaften*, 64(11):541–65, Nov 1977.
- [24] M. Eigen & P. Schüster. The hypercycle. a principle of natural self-organization. part b: Abstract nypercycle. *Naturwissenschaften*, 65(1):7–41, 1978.
- [25] M. Eigen & P. Schüster. The hypercycle. a principle of natural self-organization. part c: Realistic hypercycle. *Naturwissenschaften*, 65(7):341–369, 1978.
- [26] J. Maynard-Smith. Hypercycles and the origin of life. *Nature*, 280(5722):445–6, Aug 1979.
- [27] U. Niesert & D. Harnasch C. Bresch. Hypercycles, parasites and packages. *J. Theor. Biol.*, 85(3):399–405, Aug 1980.
- [28] E. Szathmáry. A hypercyclic illusion. *J. Theor. Biol.*, 134(4):561–563, October 1988.
- [29] M. Santos & C. Fernando E. Szathmáry. Evolutionary Potential and Requirements for Minimal protocells. *Top. Curr. Chem.*, 259:167–211, August 2005.
- [30] D. Harnasch. & C. Bresch U. Niesert. Origin of life between Scylla and Charybdis. *J. Mol. Evol.*, 17(6):348–53, 1981.

- [31] W. C. Gardiner & P. Schüster M. Eigen. Hypercycles and compartments. Compartments assists—but do not replace—hypercyclic organization of early genetic information. *J. Theor. Biol.*, 85(3):407–11, Aug 1980.
- [32] W. Gilbert. Origin of life: The rna world. *Nature*, 319(6055):618–618, February 1986.
- [33] J. A. Doudna & J. W. Szostak. RNA-catalysed synthesis of complementary-strand RNA. *Nature*, 339(6225):519–22, Jun 1989.
- [34] J. C. Nuño & P. Tarazona. Lifetime of small catalytic networks. *Bull. Math. Biol.*, 56(5):875–898, September 1994.
- [35] J. F. Fontanari & P. F. Stadler P. R. A. Campos. Error propagation in the hypercycle. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 61(3):2996–3002, Mar 2000.
- [36] S. L. Miller & H. C. Urey. Organic compound synthesis on the primitive earth. *Science*, 130(3370):245–251, Jul 1959.
- [37] R. Kafri & D. Lancet D. Segré, E. Shenhav. The molecular roots of compositional inheritance. *J. Theor. Biol.*, 213(3):481–491, Dec 2001.
- [38] U. Niesert. How many genes to start with? A computer simulation about the origin of life. *Orig. Life Evol. Biosph.*, 17(2):155–69, 1987.
- [39] E. Szathmáry & I. Gladkih. Sub-exponential growth and coexistence of non-enzimatically replicating templates. *J. Theor. Biol.*, 138(1):55–58, May 1989.
- [40] S. M. Fujikawa & J. W. Szostak M. M. Hanczyc. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science*, 302(5645):618–622, Oct 2003.
- [41] M. M. Hanczyc & J. W. Szostak. Replicating vesicles as models of primitive cell growth and division. *Curr. Opin. Chem. Biol.*, 8(6):660–664, Dec 2004.

- [42] C. Woese. The universal ancestor. *Proc. Natl. Acad. Sci. U.S.A.*, 95(12):6854–6859, Jun 1998.
- [43] M. Santos & E. Szathmáry J. F. Fontanari. Coexistence and error propagation in pre-biotic vesicle models: A group selection approach. *J. Theor. Biol.*, 239(2):247–256, March 2006.
- [44] C. R. Darwin. *On the Origin of Species By Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. W. Clowes & Sons, 1859. *Fac-Símile* disponível em <http://www.esp.org/books/darwin/origin/facsimile/>.
- [45] T. P. Wilcox & J. J. Bull J. L. Sachs, U. G. Mueller. The evolution of cooperation. *Q. Rev. Biol.*, 79(2):135–60, Jun 2004.
- [46] A. T. C. Silva & J. F. Fontanari D. Alves, P. R. A. Campos. Group selection models in prebiotic evolution. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 63(1):011911–, December 2000.
- [47] S. A. Frank. Perspective: repression of competition and the evolution of cooperation. *Evolution*, 57(4):693–705, Apr 2003.
- [48] R. Dawkins. *The Selfish Gene*. Oxford University Press, New York, 1976.
- [49] M. E. Borrello. The rise, fall and resurrection of group selection. *Endeavour*, 29(1):43–47, March 2005.
- [50] P. Hogeweg & N. Takeuchi. Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. *Orig. Life Evol. Biosph.*, 33(4-5):375–403, Oct 2003.
- [51] S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 1931.
- [52] S. A. Boorman & P. R. Levitt. Group selection on the boundary of a stable population. *Theor. Popul. Biol.*, 4(1):85–128, Mar 1973.

- [53] S. A. Boorman & P. R. Levitt. A frequency-dependent natural selection model for the evolution of social cooperation networks. *Proc. Natl. Acad. Sci. U. S. A.*, 70(1):187–189, Jan 1973.
- [54] I. Eshel. On the neighbor effect and the evolution of altruistic traits. *Theor. Popul. Biol.*, 3(3):258–277, Sep 1972.
- [55] K. Aoki. A condition for group selection to prevail over counteracting individual selection. *Evolution*, 36(4):832–842, 1982.
- [56] D. G. M. Silvestre & J. F. Fontanari. Template coexistence in prebiotic vesicle models. *Eur. Phys. J. B*, 47(3):423–429, 2005.
- [57] M. Slatkin & D. S. Wilson. Coevolution in structured demes. *Proc. Natl. Acad. Sci. U.S.A.*, 76(4):2084–2087, Apr 1979.
- [58] D. S. Wilson. A theory of group selection. *Proc. Natl. Acad. Sci. U.S.A.*, 72(1):143–146, Jan 1975.
- [59] D. Hartl & A. G. Clark. *Principles of Population Genetics*. Sinauer Associates, 1997.
- [60] M. Nei. *Molecular Population Genetics and Evolution*. American Elsevier Pub. Co, 1983, reimpressão.
- [61] A. Fog. Simulation of evolution in structured populations: The open source software package altruist. *Biotech Software & Internet Report*, 1(5):226–229, 2000.
- [62] R. W. Roberts & J. W. Szostak I. A. Chen. The emergence of competition between model protocells. *Science*, 305(5689):1474–1476, Sep 2004.
- [63] D. P. Bartel & P. L. Luisi J. W. Szostak. Synthesizing life. *Nature*, 409(6818):387–390, Jan 2001.

- [64] A. Mushegian. The minimal genome concept. *Curr. Opin. Genet. Dev.*, 9(6):709–714, Dec 1999.
- [65] E. V. Koonin. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.*, 1(2):127–136, Nov 2003.
- [66] E. V. Koonin. How many genes can make a cell: the minimal-gene-set concept. *Annu. Rev. Genomics Hum. Genet.*, 1:99–116, 2000.
- [67] M. Kimura. *The neutral theory of molecular evolution*. Cambridge University Press, reimpressão, 1985. A mais importante e bem-sucedida teoria em evolução molecular.
- [68] S. Lifson. On the crucial stages in the origin of animate matter. *J. Mol. Evol.*, 44(1):1–8, Jan 1997.
- [69] J. A. Vucetich. & T. A. Waite. Erosion of heterozygosity in fluctuating populations. *Conserv. Biol.*, 13(4):860–868, August 1999.
- [70] N. Lehman. A case for the extreme antiquity of recombination. *J. Mol. Evol.*, 56(6):770–777, Jun 2003.