UNIVERSIDADE ESTADUAL DE CAMPINAS

Flávia Fuchs de Jesus

Modelos com variação de estrutura populacional no tempo e estudo de suas conseqüências genéticas

Este exemplar corresponde à redação final da tese defendida pelo(a) candidato (a) FLÁVIA FUCHS DE JESUS

e aprovada pela Comissão Julgadora.

Tese apresentada ao Instituto de Biologia para obtenção do título de Doutor em Genética e Biologia Molecular na área de Genética Animal e Evolução

Orientadora: Profª. Drª. Vera Nisaka Solferini

Co-orientador: Prof. Dr. John Wakeley

Campinas, 2006

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

J499m

Jesus, Flávia Fuchs de
Modelos com variação de estrutura populacional no tempo e estudo de suas conseqüências genéticas / Flávia Fuchs de Jesus. -- Campinas, SP: [s.n.], 2006.

Orientadora: Vera Nisaka Solferini.
Co-orientador: John Wakeley.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Genética de populações. 2. Teoria de coalescência. 3. Estruturação genética. 4. Variabilidade genética. 5. Ciclos glaciais. I. Solferini, Vera Nisaka. II. Wakeley, John. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

(scs/ib)

Título em inglês: Models with variation in population structure through time and study of genetic consequences.
Palavras-chave em inglês: Population genetics; Coalescent theory; Genetic structure; Genetic variation; Glacial cycles.
Área de concentração: Genética Animal e Evolução.
Titulação: Doutora em Genética e Biologia Molecular.
Banca examinadora: Vera Nisaka Solferini, Louis Bernard Klaczko, Paulo Alberto Otto, Enrique Pablo Lessa Gallinal, Aluíso de Souza Pinheiro.
Data da defesa: 25/08/2006.
Programa de Pós-Graduação: Genética e Biologia Molecular.

Campinas, 25 de agosto de 2006.

BANCA EXAMINADORA

Profa. Dra. Vera Nisaka Solferini
(orientadora)

Prof. Dr Louis Bernard Klaczko

Prof. Dr. Sérgio Furtado dos Reis

Prof. Dr. Paulo Alberto Otto

Prof. Dr. Enrique Pablo Lessa Gallinal

Prof. Dr. Marcus Aloizio Martinez de Aguiar

Prof. Dr. Aluísio de Souza Pinheiro

Prof. Dr. Fábio de Melo Sene

Dedico esta tese a meu marido, José Andrés, a nossos futuros filhos e a meus pais, Cláudio e Beatriz.

"O maior conforto humano é a luz do sol, mas a maior alegria da mente é a compreensão matemática."

"Aqueles que se enamoram somente da prática, sem cuidar da teoria, ou melhor dizendo, da ciência, são como o piloto que embarca sem timão nem bússola. A prática deve alicerçar-se sobre uma teoria, a qual serve de guia e perspectiva (...)"

Leonardo DaVinci

AGRADECIMENTOS

A Deus, esse Mistério de Amor que se revela, por nos querer como sujeitos da história, nos conceder essa dignidade, e nunca se impor, ao mesmo tempo que nunca nos abandona. Agradeço, especialmente no contexto desta tese e da minha caminhada profissional, porque sua providência permitiu que eu conhecesse pessoas tão maravilhosas, que me ajudaram no caminho, às quais agradeço em seguida. Também, agradeço pela providência que me permitiu, de uma maneira tão inusitada, encontrar um caminho para realizar um desejo antigo, e que já estava quase abandonado: o de trabalhar com modelagem matemática, em genética de populações. E pela paz especial que me permitiu sentir diante de algumas decisões importantes.

A Maria, Nossa Senhora, mãe da humanidade, por estar sempre presente e interceder por alguns pedidos especiais recentemente, com amor de mãe.

Aos meus orientadores: Vera N. Solferini e John Wakeley. À Vera por me orientar desde a graduação, por me dar um apoio constante, por me estimular em muitos sonhos e idéias. Por ter contribuído de inúmeras maneiras em minha formação profissional, e também humana. Ao John, por ter me acolhido em seu laboratório, quando eu não tinha a menor idéia do que era desenvolver modelos, menos ainda em teoria de coalescência. Por me ensinar muito, por compartilhar a alegria das descobertas, por me incentivar, pelo apoio em muitos momentos, pela paciência em outros tantos.

Ao Jon F. Wilkins, que também teve um papel de orientador durante alguns meses, especialmente no início. Pelo estímulo inicial, quando era monitor da primeria disciplina que cursei em Harvard, e me encorajou a entrar no doutorado, e iniciar o trabalho com teoria. Pela energia, pelo entusiasmo, pela amizade.

Ao Prof. Aluísio Pinheiro, por me acolher em seu grupo, pelos erros que apontou, pelas importantes sugestões que deu. Em especial, quando eu estava

encontrando erros em uma rotina computacional, foram as suas sugestões que me permitiram resolvê-los.

Ao Prof. Louis Bernard, pela grande admiração que me inspira (e não só a mim!), por ter sido um dos melhores professores que já tive, por ter me ensinado de modo tão elegante os princípios da genética de populações, por ter me estimulado a continuar e não desistir, em mais de um momento – às vezes talvez sem saber. Por ter me dado alguns bons conselhos, e também é claro, pela sua participação na Pré-Banca.

Ao Diogo Meyer, pelo grande apoio e estímulo – crucial em alguns momentos, em especial na volta ao Brasil – pelas trocas de idéias, pelo companheirismo nos projetos de ensino, e mais recentemente pela participação na Pré-Banca.

Aos professores membros da Banca, pela disponibilidade para contribuir.

A todos os professores da Unicamp que contribuíram para a minha formação.

A alguns professores de Harvard que me acolheram sem a menor obrigação, quando eu não era aluna oficialmente. Em especial aos de probabilidade e estatística.

Aos colegas do Wakeley Lab: Alex, Tee, Jon, Guillaume, Richard, Danielle, Bjarki, Hashi, Josh, Jeff e em especial a Tee Muirhead, por todas as discussões tão enriquecedoras, por partilhar seu conhecimento comigo inúmeras vezes, pela amizade.

Aos colegas do Hartl Lab, e aos demais colegas do OEB, pelos grupos de estudo e discussão e pelos momentos de descontração.

Ao Prof. Marcus Aguiar e seus alunos Alexandre, Sabrina, Beth, Fernando, David, e Paulo "Miúdo", por me acolherem em seu grupo de seminários, e pelas trocas de idéias.

Ao Prof. Sergio Furtado dos Reis, pelas palavras de estímulo em diversos momentos.

Aos muitos colegas (e agregados) do Laboratório de Diversidade Genética, de hoje e de antes, por serem tão companheiros, cooperativos, humanos; e também pelas muitas trocas de idéias e boas conversas que tivemos nestes anos: Veridiana, Fabiana, Ju Farinaci, Sónia, Horácio, Ju Félix, Juzinha, Teresa, Terèse, Karla, Tibúrcio, Aluana, Flávio, Claudemir, Kaiser, Camila, Renato, Ricardo, Bruno, Joana, Karina, Kívia, Evandro, Maína, Suzana, Joice,...

Em especial, agradeço à Sónia Andrade, colega desde a iniciação, por ser sempre amiga, disposta a ajudar, "ponta firme". E também à Aluana, por todo o apoio, em especial nestes últimos tempos.

Ao Herberth, seu Pedro, Ricardo, Patrícia e todos os funcionários do Departamento de Genética e Evolução que de alguma forma apoiaram este trabalho. Também, às secretárias da Pós-Graduação, em especial à Lourdes, por toda a ajuda, presteza e paciência.

Às minhas comunidades CVX/CLC ("Comunidade de Vida Cristã"), inclusive aos amigos dos regionais S.Paulo e New England, pela amizade especial que criamos. Em especial à Cambridge CLC ("dear friends": Silvia, Eva, Steve, Martha, Uwe, Marcos, Ramón, Thad, Clarivel, Dario, Jessica, Adrine). Também aos jesuítas da região de Boston, pelo acolhimento tão caloroso. Em especial ao Pe. Carlos Viana sj e ao Marcos Mendes, pela grande amizade. Aos amigos da comunidade atual, em especial os mais presentes hoje: Pat, Sakai, Anali, Marlene, Alessandro. Ao Fábio Papes, pela amizade mesmo na distância.

Às comunidades do C&L e dos Focolares na região de Boston, pelo acolhimento, amizade e testemunho de fé. Também às comunidades dos Focolares de Campinas e da Polônia, pelo acolhimento amoroso.

Aos amigos da PU e do grupo "Onde dois ou mais", por todo o carinho: Lívio, Lívia, Gui, Thaís, Priscila, Lucas, Helder, Paula, Karina, André, Felipe, Jose, Raquel...

Aos amigos da Pastoral da Criança, pelo companheirismo, por partilhar um sonho, pela compreensão da minha pouca participação nos últimos tempos: Sakai, Nílson, D. Judite, Anali, Pat, Romaric.

Aos amigos da Casa Sto. Inácio, em especial ao Edison de Lima, sj., "caro amigo", pelo carinho e por seu sorriso constante, e à Marly, pelas orações.

A algumas pessoas que me acompanharam espiritualmente em alguns momentos: Pe. Toninho sj, Maria Elena, MariCristi, Pe. Walter Conlan sj, e em especial ao Pe. Antonio López e ao Pe. Ranieri sj, por conselhos e ponderações importantíssimas, e pelo exemplo de santidade.

Ao senhor Rosendo Augusto Yunes, pelo estímulo e provocações constantes, e por ser um modelo de professor e pesquisador. A toda a família Yunes, por todo o carinho e apoio neste últimos anos. Em especial à Simone, por me ensinar um "segredo" especial para quem está escrevendo tese.

Ao José Andrés Yunes, meu marido, por um estímulo e encorajamento quase irritantes. Por estar ao meu lado em todos os momentos, como um grande e fiel amigo, e me fazer acreditar em possibilidades que já me pareceram impossíveis, e acabaram se concretizando. Por ser um homem de coragem, e me incentivar a ser sempre mais corajosa.

A minha família "de origem", em especial a meus pais, Cláudio e Beatriz, pelo apoio constante e incondicional, pelo estímulo, pela doação, pelas orações. Também a meus queridos irmãos Guto, Gil e Gério, minhas queridas cunhadas Cynthia, Cris e Débora e às minhas tias queridas, pelo apoio e torcida.

Aos muitos amigos, presentes nestes anos todos, pelo carinho, pelas alegrias, pela torcida, pelas partilhas: Renata, Pedrinho, Bel, Carlos, Angela, Marisa, Ju, Alysson, Fernandinha, Nilda, Germano, Hugo, Cíntia, Waltinho, Cleide, Gud, Lívia, Dani, Ná, Alejandra, Sandra, Mônica... Em especial, ao "cumpadres" Élcio e Lucy, com quem compartilho muitos momentos da vida, e ao pequenino João Ignácio que trouxe alegria a todos.

A todos os outros amigos que não citei, que de alguma forma contribuíram para a realização deste trabalho.

À CAPES (PDEE) e à FAPESP, pelas bolsas de doutorado, respectivamente durante o "doutorado sanduíche" e na Unicamp. À FAPESP também pelo financiamento do projeto de auxílio à pesquisa (04/09742-0).

A todos, os meus sinceros agradecimentos.

RESUMO

A estrutura populacional é um dos principais fatores moldando os padrões de variabilidade genética no tempo e no espaço. Devido às flutuações climáticas que ocorreram durante o período Quaternário, muitas espécies podem ter sofrido redução e fragmentação populacional, ficando restritas a "refúgios" durante períodos glaciais e se expandindo novamente durante os interglaciais. Isto tem sido utilizado para explicar alguns padrões encontrados nas espécies atualmente. O presente trabalho consistiu no desenvolvimento e estudo de modelos para auxiliar na compreensão das conseqüências genéticas de mudanças cíclicas na estruturação e tamanho populacionais, como as que teriam ocorrido ao longo das flutuações climáticas do Quaternário. A redução populacional é capaz de causar redução do tamanho efetivo populacional, do tempo médio de coalescência e da variabilidade genética, ao passo que um aumento na subdivisão populacional pode ter o efeito oposto. Para investigar estes efeitos opostos, foram estudados dois modelos, ambos com alternância de duas fases correspondendo aos períodos glaciais e interglaciais. Em ambos os modelos permitiram-se mudanças na estrutura populacional, além de mudanças no tamanho populacional, de uma maneira cíclica. No primeiro modelo, fases totalmente panmíticas alternaram-se com fases totalmente estruturadas. A partir deste modelo, obteve-se uma expressão para a esperança do tempo de coalescência de duas seqüências e, a partir desta, uma expressão para a esperança do número de sítios polimórficos. Tanto o aumento do número de demes quanto da duração das fases estruturadas causaram um aumento do tempo de coalescência e dos níveis de variabilidade genética. Os resultados obtidos foram comparados com os que seriam esperados para uma população panmítica de tamanho constante. Verificou-se que a estruturação pode superar o efeito da redução populacional durante os períodos glaciais. Especificamente, o número médio de sítios polimórficos pode ser maior no modelo proposto, mesmo quando o tamanho populacional é muito reduzido durante as fases estruturadas. No segundo modelo, permitiu-se subdivisão

populacional de acordo com o modelo de finitas ilhas em ambas as fases, com migração. O tamanho populacional, a taxa de migração e o número de demes variaram entre as fases. Para este modelo, além de uma expressão para a o tempo médio de coalescência, obteve-se também uma expressão para a distribuição dos tempos de coalescência de duas seqüências. As distribuições observadas foram muito diferentes do que seria esperado para uma população panmítica de tamanho constante. Um tamanho populacional reduzido durante os períodos glaciais causou descontinuidades e picos múltiplos na distribuição dos tempos de coalescência, bem como uma redução dos tempo médios. O aumento da estrutura populacional, através da redução da taxa de migração, aumentou os tempos médios e atenuou os picos da distribuição. O tempo médio de coalescência, em geral, também aumentou em decorrência de um maior número de demes durante os períodos glaciais. Os resultados encontrados ajudam na compreensão das conseqüências genéticas de ciclos glaciais e, em especial, da importância da estrutura populacional na manutenção da variabilidade genética. Além disso, oferecem uma possível explicação para padrões genéticos observados em muitas espécies em que genealogias gênicas muito longas são econtradas, com o ancestral comum mais recente antecedendo em muito ao último período glacial.

ABSTRACT

Population structure is one of the major factors shaping the patterns of genetic variation across time and space. Due to the climatic fluctuations of the Quaternary, several species may have suffered population reduction and fragmentation, becoming restricted to refugia during glacial periods and expanding again during interglacials. This has been used to explain some patterns currently observed in several species. The present work consisted in the development and study of models to help understand the genetic consequences of cyclic changes in population structure and size, such as the ones that may have occurred throughout the climatic fluctuations of the Quaternary. Population reduction may cause reduction in population effective size, mean coalescence time and genetic variation, whereas an increase in population subdivision may have the opposite effect. In order to investigate these two opposite effects, two models were studied, both with two alternating phases, corresponding to the glacial and interglacial periods. Both models included changes in population structure, besides those in population size, in a cyclic manner. In the first model, completely panmictic phases were alternated with completely structured ones. Based on this model, an expression was derived for the expectation of coalescence times of two sequences and, from this, an expression for the expectation of the number of segregating sites. Both an increase in the number of demes and in the duration of the structured phases caused an increase in coalescence times and levels of genetic variation. The results obtained were compared to what would be expected for a panmictic population of constant size. It was verified that population structure may outhweigh the effect of population reduction during glacial periods. Specifically, the mean number of segregating sites can be greater in the proposed model, even when population size is quite reduced during the structured phases. In the second model, population subdivision was allowed in both phases - according the finite island model – with migration. Population size, migration rate and number of demes varied

between phases. For this model, besides an expression for the mean coalescence time, an expression for the distribution of coalescence times was also obtained. The distributions observed were quite different from what would be expected for a panmictic population of constant size. Population reduction during glacial periods caused discontinuities and multiple peaks in the distribution of coalescence times, as well as a reduction in the expected times. An increase in population structure, through reducing migration rates, increased the mean times and attenuated the peaks of the distribution. Mean coalescence times, in general, also increased with a greater number of demes during glacial periods. The results obtained help understand the genetic consequences of glacial cycles, and, especially, point to the importance of population structure for the maintenance of genetic variation. Besides, they offer a potential explanation for the genetic patterns observed in several species, for which long gene genealogies are observed, with the most recent ancestor predating by far the last glacial period.

ÍNDICE

INTRODUÇÃO GERAL

O período Quaternário teve início há aproximadamente 2 milhões de anos, e vem sendo marcado por grandes variações climáticas, com alternância de períodos glaciais e interglaciais. Vários estudos paleoclimáticos indicam que essas flutuações ocorrem ciclicamente, relacionadas a mudanças periódicas na órbita da Terra (Tzedakis et al. 1997, Petit et al. 1999, Jouzel 2003). Cada ciclo completo dura aproximadamente 100.000 anos, com períodos glaciais longos interrrompidos por períodos interglaciais mais curtos (de menos de 10.000 anos geralmente), além de algumas flutuações menores aninhadas em cada período. Atualmente vivemos um período interglacial, que teve início há aproximadamente 11.000 anos (Petit et al. 1999).

Essas flutuações climáticas parecem influenciar o padrão de distribuição de diversos organismos (Webb & Bartlein 1992, Burnham & Graham 1999). Muitas espécies podem sofrer redução e fragmentação populacional, ficando restritas a "refúgios" durante os períodos glaciais. Durante os interglaciais podem, então, se expandir novamente. Essas mudanças de tamanho e distribuição populacional têm sido bem investigadas, especialmente no Hemisfério Norte, em regiões de alta latitude em que as alterações climáticas foram muito pronunciadas. Mas há também estudos e discussões acerca das conseqüências das flutuações climáticas do Quaternário para espécies tropicais (Hewitt 2000, 2004; Lessa et al. 2003). Nos trópicos, o clima mais frio e seco durante os períodos glaciais permitiu a expansão de savanas e campos, com uma redução e possível fragmentação das florestas.

Mudanças demográficas podem, a princípio, deixar sinais detectáveis a partir de seqüências de DNA. Já existe uma grande quantidade de estudos, a partir de dados genéticos, de espécies que teriam passado por essas mudanças cíclicas (Hugall et al. 2002, Lessa et al. 2003, Petit et al. 2003, Galbreath & Cook 2004, McLachlan et al. 2005, entre outros). Na maioria desses trabalhos busca-se responder questões acerca do passado populacional das espécies, tentando-se

esclarecer a existência de refúgios, bem como seu número e sua importância para a variabilidade genética de populações atuais. Em geral, procura-se enfocar as conseqüências genéticas do último ciclo apenas; entretanto, para muitas dessas espécies, o ancestral comum mais recente dos genes estudados parece ocorrer anteriormente ao último período glacial (Hewitt 2004). Por exemplo, em duas espécies de besouros do gênero *Moneleima*, que ocorrem em regiões desérticas da América do Norte, estimou-se que o tempo até o ancestral comum mais recente de alguns genes mitocondriais teria ocorrido há mais do que um ou até dois milhões de anos atrás (Smith & Farrell 2005). Outro exemplo é o caso do lagarto Ibérico, *Lacerta schreiberi*, em que linhagens de DNA mitocondrial parecem anteceder o período Quaternário (Paulo et al. 2001). Além disso, estudos de filogenia indicam que vários *taxa* teriam se originado anteriormente ao Pleistoceno (por ex. Klicka & Zink 1997, para aves norte-amercianas; Moritz et al. 2000, para faunas de florestas tropicais). Esses estudos indicam que grande parte das espécies existentes hoje podem ter sofrido múltiplos ciclos glaciais e, possivelmente, carregariam algum sinal genético desta história.

É interessante, portanto, um modelo teórico que descreva as conseqüências genéticas de tais mudanças cíclicas. Apesar da ampla quantidade de estudos empíricos, ainda há necessidade de estudos teóricos, de modelos que sejam adequados para descrever as conseqüências genéticas dos ciclos glaciais-interglaciais. Dois processos muito importantes, que têm que ser levados em conta, são: redução de habitat e fragmentação de habitat. Espera-se que esses processos afetem o número de indivíduos que o habitat é capaz de suportar (tamanho populacional) bem como a distribuição dos indivíduos no ambiente (estrutura populacional).

Os efeitos genéticos de mudanças de tamanho populacional têm sido o objeto de numerosos estudos teóricos, ao passo que mudanças em estrutura populacional têm sido menos estudadas. A seguir, passamos a uma breve revisão

da literatura pertinente, ao final da qual descreve-se o que foi realizado no presente trabalho.

Mudanças em tamanho populacional

Um tipo de mudança em tamanho populacional que é de grande interesse em genética de populações é o "gargalo": uma redução populacional, geralmente seguida de uma expansão. É interessante notar que um dos primeiros estudos acerca do efeito de gargalos sobre a variabilidade genética foi motivado por flutuações populacionais cíclicas. Um dos resultados amplamente conhecidos, de Sewall Wright (1938), é que quando os ciclos têm curta duração – em gerações – a média harmônica dos tamanhos da população ao longo do tempo é uma boa aproximação para seu tamanho efetivo. Posteriormente, foram descritos os efeitos de um único gargalo sobre a heterozigosidade e o número de alelos (Nei et al. 1975, Maruyama & Fuerst 1984, 1985a). O efeito de gargalos periódicos sobre a heterozigosidade foi estudado por Maruyama & Fuerst (1985b), a partir de um modelo com duração fixa dos ciclos.

No contexto da teoria de coalescência (Kingman 1982a,b, Hudson 1983, Tajima 1983), e pressupondo um modelo mutacional de infinitos sítios (Kimura 1969), Tajima (1989b) estudou a esperança do número de sítios polimórficos e o número médio de diferenças par-a-par em um modelo em que uma população sofre uma mudança repentina em seu tamanho populacional, ou sofre um gargalo. Slatkin & Hudson (1991) estudaram a distribuição do número de diferenças par-a-par em um modelo de crescimento exponencial e Rogers & Harpending (1992) investigaram esta distribuição para um modelo de expansão repentina. Em ambos, verificou-se que a distribuição após um evento de expansão é unimodal e bem distinta da que seria esperada para uma população em equilíbrio. Além disso, foram desenvolvidos métodos computacionais para a estimativa de parâmetros de tamanho e crescimento populacional a partir de

informações das genealogias gênicas (Griffiths & Tavaré 1994, Kuhner et al. 1998).

Estrutura populacional

Apesar de os efeitos de mudanças em estrutura populacional terem sido relativamente pouco estudados, as conseqüências genéticas da estrutura populacional em si têm sido objeto de numerosos estudos. Entre os modelos de subdivisão populacional e migração, o modelo de ilhas (Wright 1931) talvez seja o mais estudado, devido em grande parte à sua simplicidade matemática. Neste modelo, a população é subdividida em sub-populações, ou "demes", sendo que todos os demes podem trocar migrantes entre si. Em geral, uma taxa de migração reduzida entre demes resulta em uma redução da diversidade genética dentro de cada deme e, ao mesmo tempo, em um aumento da diferenciação genética entre os demes. A partir do modelo de ilhas, Wright (1951) concluiu que uma população subdividida é capaz de preservar um maior número de alelos quando comparada a uma população panmítica de mesmo tamanho total. Ou seja, uma possível conseqüência da subdivisão é a acumulação de diversidade genética adicional.

O modelo de ilhas formulado por Wright, originalmente, tinha como pressuposto um número infinito de sub-populações. Modificações posteriores passaram a considerar um número finito (Maruyama 1970, Latter 1973). Slatkin (1991) obteve expressões para coeficientes de endogamia a partir de tempos de coalescência par-a-par, para o modelo de finitas ilhas e diversos outros modelos. Nei & Takahata (1993) obtiveram uma fórmula para o tamanho efetivo populacional no modelo de finitas ilhas, a partir de tempos médios de coalescência, e descobriram que este tamanho efetivo é maior em populações com taxas de migração menores.

Wakeley (1998) obteve expressões para a esperança e a variância do comprimento total da árvore para a genealogia de uma amostra de seqüências, a

partir do modelo de finitas ilhas, pressupondo um grande número de ilhas. Esta solução utiliza uma abordagem de separação de escalas de tempo, em que a história da amostra é dividida em duas fases. A história recente da população é modelada com consideração explícita da distribuição geográfica das amostras. Para a história mais remota da população – que representa a maior parte da profundidade da genealogia – o processo genealógico apresenta um comportamento muito semelhante ao do processo de coalescência em uma população panmítica de tamanho constante ("standard coalescent", Kingman 1982 a,b), após uma correção da escala do tempo, a partir do tamanho efetivo populacional. Estes resultados foram extendidos para um modelo que inclui uma mudança demográfica no passado (Wakeley 1999) e para um modelo mais geral, que permite migração assimétrica entre demes e tamanhos desiguais para os demes (Wakeley 2001).

Muitas outras contribuições foram feitas para o estudo de populações subdivididas. O trabalho de Wakeley (2004) inclui uma revisão sobre o modelo de ilhas, bem como algumas novas perspectivas. Hey e Machado (2003) apresentam uma revisão mais geral sobre a relação entre estrutura genética e geográfica; e Charlesworth et al. (2003) apresentam uma revisão completa da literatura sobre genética de populações subdivididas.

Expansão com estrutura populacional

Alguns estudos recentes consideraram modelos de expansão ou colonização populacional incluindo também estrutura populacional. Austerlitz et al. (1997) estudaram colonização a partir de um único deme, em um modelo "stepping-stone" linear. Eles obtiveram a distribuição e a esperança dos tempos de coalescência de dois genes, e estudaram também a mudança dos valores esperados para o índice de fixação F_{ST} ao longo do tempo. Eles verificaram que o efeito fundador (redução da diversidade genética devido ao pequeno número de indivíduos colonizadores) tende a aumentar com a distância do primeiro deme, e

diminuir quando a taxa de migração ou a taxa de crescimento populacional são aumentadas. Estes autores também estudaram o efeito de um gargalo populacional em um único deme, e observaram um pico na distribuição dos tempos de coalescência de dois genes, correspondendo ao momento de ocorrência do gargalo. Austerlitz et al. (2000) simularam o processo de colonização em um modelo "stepping-stone" de duas dimensões. Eles estudaram também o impacto de gerações sobrepostas e de uma fase juvenil longa sobre a distribuição da variabilidade genética, visando explicar os padrões genéticos observados em plantas anuais e árvores.

Ray et al. (2003) fizeram um estudo de simulação de expansão no espaço – iniciada a partir de um único deme – em um modelo "stepping-stones" bi-dimensional, registrando vários aspectos da diversidade molecular intra-dêmica. Eles observaram que para valores altos de *Nm* (número de migrantes por geração) – acima de 20 – as genealogias gênicas têm forma de estrela e as distribuições "mismatch" são unimodais, como é o padrão esperado em expansões sem estrutura populacional (Slatkin & Hudson 1991, Rogers & Harpending 1992). Mas quando o número de migrantes é menor, as distribuições "mismatch" se tornam multimodais, e testes estatísticos comumente utilizados para detectar desvios de equilíbrio demográfico, como o teste *D* de Tajima, passam a ter pouco poder para detectar as expansões. Excoffier (2004) deu continuidade a este estudo, obtendo, analiticamente, expressões para a distribuição do número de diferenças entre pares de seqüências amostradas após uma expansão com subdivisão, sendo esta modelada a partir do modelo de infinitas ilhas. Os resultados encontrados foram concordantes com os que haviam sido obtidos anteriormente por Ray et al. (2003) a partir de simulações baseadas no modelo do tipo "stepping-stones". Muito recentemente, Ewing & Rodrigo (2006) apresentaram um método para estimativa de parâmetros, permitindo mudanças de diversos parâmetros populacionais ao longo do tempo, inclusive do número de demes. O método foi desenvolvido para amostras seriais,

ou seja, amostras obtidas em diferentes momentos no tempo, e pode ser usado para estudar colonização – e outros processos – quando este tipo de amostra estiver disponível.

O presente trabalho

O presente trabalho consistiu no desenvolvimento e estudo de modelos teóricos com alternância de duas fases, correspondendo a períodos glaciais e interglaciais. Teve como objetivo auxiliar na compreensão das conseqüências genéticas de mudanças populacionais cíclicas, como as causadas pelas flutuações climáticas do período Quaternário.

Foram estudados dois modelos, ambos com mudança na estrutura populacional, além de mudança no tamanho populacional, de uma maneira cíclica. No primeiro modelo, uma fase panmítica foi alternada com uma fase em que a população é subdividida em demes de igual tamanho, sem migração entre eles. A partir deste modelo, obteve-se uma expressão para a esperança matemática do tempo de coalescência de duas seqüências. Esta expressão permitiu investigar os efeitos possivelmente opostos da redução populacional e do aumento de estrutura durante os períodos glaciais, identificando situações em que os tempos médios de coalescência – e conseqüentemente o polimorfismo – seriam aumentados ou diminuídos.

O segundo modelo foi mais geral, permitindo subdivisão populacional – de acordo com o modelo de finitas ilhas – com migração, em ambas as fases. Os parâmetros de tamanho populacional, taxa de migração e número de demes foram variados entre as fases. A partir deste segundo modelo estudou-se também, além do tempo médio, a distribuição dos tempos de coalescência de duas seqüências. Estudou-se como esta distribuição é alterada pelas mudanças cíclicas em estrutura e tamanho populacional, quando comparada à distribuição esperada no modelo de Kingman (1982a), de uma população panmítica de tamanho constante ("standard coalescent"). Também, a partir de ambos os

modelos estudados, investigou-se a importância da estrutura populacional na manutenção da variabilidade genética durante períodos de redução populacional.

CAPÍTULO 1

EXPECTED COALESCENCE TIMES AND SEGREGATING SITES IN A MODEL OF GLACIAL CYCLES

Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research

Expected coalescence times and segregating sites in a model of glacial cycles

F.F. Jesus[1], J.F. Wilkins[2], V.N. Solferini[1] and J. Wakeley[3]

[1]Departamento de Genética e Evolução, Instituto de Biologia, Unicamp, Campinas, SP, Brasil
[2]Santa Fe Institute, Santa Fe, NM, USA
[3]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA
Corresponding author: F.F. Jesus
E-mail: flaviaf.jesus@gmail.com/ffjesus@unicamp.br

Genet. Mol. Res. 5 (3): 466-474 (2006)
Received April 26, 2006
Accepted June 16, 2006
Published July 31, 2006

ABSTRACT. The climatic fluctuations of the Quaternary have influenced the distribution of numerous plant and animal species. Several species suffer population reduction and fragmentation, becoming restricted to refugia during glacial periods and expanding again during interglacials. The reduction in population size may reduce the effective population size, mean coalescence time and genetic variation, whereas an increased subdivision may have the opposite effect. To investigate these two opposing forces, we proposed a model in which a panmictic and a structured phase alternate, corresponding to interglacial and glacial periods. From this model, we derived an expression for the expected coalescence time and number of segregating sites for a pair of genes. We observed that increasing the number of demes or the duration of the structured phases causes an increase in coalescence time and expected levels of genetic variation. We compared numerical results with the ones expected for a panmictic population of constant size, and showed that the mean number of segregating sites can be greater in our model even

when population size is much smaller in the structured phases. This points to the importance of population structure in the history of species subject to climatic fluctuations, and helps explain the long gene genealogies observed in several organisms.

Key words: Refugia, Coalescent theory, Population structure, Pairwise differences

INTRODUCTION

The climatic fluctuations of the Quaternary, with cycles of glacial and interglacial periods, have affected the distribution of several organisms (Webb and Bartlein, 1992; Burnham and Graham, 1999). Many plant and animal species are subject to population reduction and fragmentation, becoming restricted to refugia during glacial periods and expanding again during the interglacials. This affects both population size and structure, in a cyclic fashion.

Demographic changes may be detectable in DNA sequences. In fact, there have been a great number of studies based on sequence or polymorphism data from species that are thought to have undergone these cyclic changes (e.g., Hugall et al., 2002; Lessa et al., 2003; Petit et al., 2003; for reviews, see Hewitt, 2000, 2004). In several of these, the questions of interest are the existence and number of refugia, as well as their importance for current genetic variation.

From the theoretical side there have been numerous studies on the genetic consequences of population size change (e.g., Wright, 1938; Nei et al., 1975; Slatkin and Hudson, 1991) and on the importance of population structure (for review, see Charlesworth et al., 2003). Also, some authors have included subdivision in models of population expansion or colonization (Austerlitz et al., 1997; Ray et al., 2003; Excoffier, 2004). However, none of them have taken into account cyclic changes in both population size and structure. Such a model is necessary if we are to investigate the genetic consequences of cyclic climatic fluctuations.

One question of interest is whether the demographic changes caused by these climatic fluctuations lead to an increase or a decrease in coalescence times and, consequently, in levels of genetic variation. The reduction in population size is expected to reduce the effective population size, mean coalescence time and overall genetic variation. On the other hand, the increase in population subdivision may cause a stretch of the mean coalescence times and an increase in genetic variation.

In the present study, we investigated coalescence times and genetic variation in a model of repeated changes in population size and structure. We used a coalescent approach (Kingman, 1982a,b; Hudson, 1983; Tajima, 1983), and obtained expressions for the expected coalescence time and the number of segregating sites of two sequences, sampled in the present interglacial period. We considered the effects of the opposing forces of population reduction and increased structure, focusing specifically on what conditions result in an overall increase or decrease in mean coalescence times and genetic variation.

MODEL AND THEORY

In this model the population undergoes cycles composed of two phases: a panmictic phase, corresponding to the interglacial periods, and a structured phase, corresponding to the glacial periods (Figure 1). We assumed that the species is favored by the interglacials, and therefore has a larger and more continuous population during those times, while having a smaller and more fragmented population during the glacial periods.

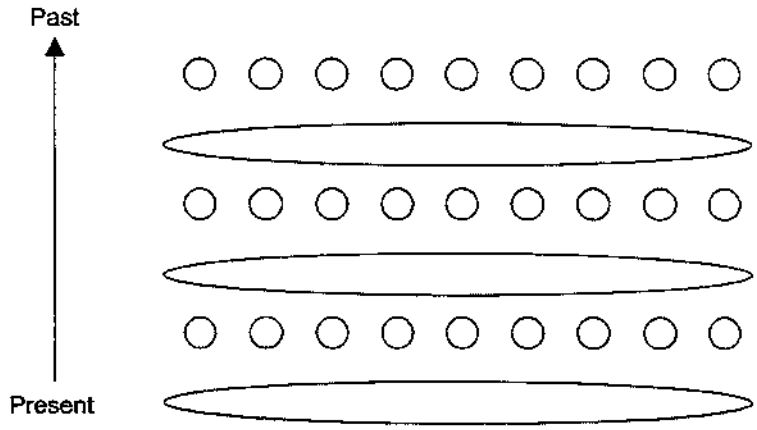

Figure 1. Schematic representation of the model, with alternating panmictic and structured phases corresponding to interglacial and glacial periods, respectively. The current phase is an interglacial.

Reproduction is according to the Wright-Fisher model (Fisher, 1930; Wright, 1931), both in the panmictic phase and within each deme during the structured phase. We considered a haploid organism, but the conclusions can be extended to diploid organisms. We also made the usual coalescent assumption that the population or deme size is much larger than the number of sequences sampled (see Hudson, 1990).

The model has five parameters, which are the population size during the panmictic phases (N_p), the duration in generations of each panmictic phase (t_p), the number of demes in the structured phases (D), the deme size (N_S), and the duration in generations of each structured phase (t_S). We assumed that during the structured phases the population is subdivided into D demes of equal size (N_S), with no migration between them. In the transition from a panmictic to a structured phase, backwards in time, the lineages are randomly distributed among the D demes.

We have derived an expression (see Appendix) for the expected coalescence time (E[coal. time]) of two genes sampled in the present interglacial:

$$E[\text{coal. time}] = \left[\frac{D}{D-(D-1)e^{-t_p/N_p}-e^{-(t_p/N_p+t_S/N_S)}}\right] \cdot\left\{N_p.(1-e^{-t_p/N_p})+e^{-t_p/N_p}[(1-1/D).t_s+(1/D).N_S.(1-e^{-t_S/N_S})]\right\} \quad \text{(Equation 1)}$$

We were also interested in levels of genetic variation, and one measure which is largely used is the number of segregating - or polymorphic - sites in a sample of genes. The expectation of this number is a simple function of the mean coalescence time if we assume no intra-genic recombination and a simple mutation model. Assuming an infinite sites mutation model (Kimura, 1969), the expected number of segregating sites in a sample of two genes is, simply:

$$E[S] = 2\mu.E[\text{coal. time}] \qquad \text{(Equation 2)}$$

where μ is the mutation rate per sequence per generation, and E[coal. time] is obtained from Equation 1. The mean number of segregating sites, in this case, is also the mean number of pairwise differences.

RESULTS AND DISCUSSION

We started by studying the influence of each parameter of the structured phases on coalescence times and genetic variation. Since we are interested in habitat reduction during glacial periods, we focused our attention to cases in which both $N_S < N_P$ and $D.N_S < N_P$, i.e., when both the deme size and the total population size in the structured phases - corresponding to the glacial periods - are smaller than the population size in the panmictic phases.

In Figure 2, we plotted the mean number of segregating sites as function of the time (t_S), deme size (N_S) and number of demes (D) of the structured phases. We assumed a mutation rate of $\mu = 10^{-5}$ and fix the parameters of the panmictic phases as $N_P = 100{,}000$ and $t_P = 10{,}000$. The other parameter values for each plot are given in the figure legend.

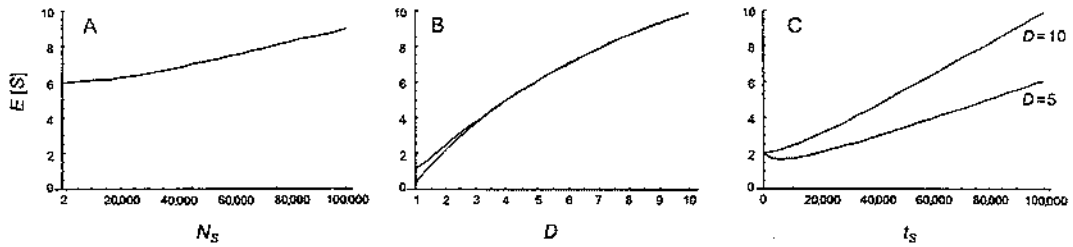

Figure 2. Mean number of segregating sites ($E[S]$) as a function of the following parameters of the structured phase: **A.** size of each deme (N_S), with $t_S = 100{,}000$ and $D = 5$; **B.** Number of demes (D), with $t_S = 100{,}000$ and $N_S = 10{,}000$ (lower line) or $D.N_S = 50{,}000$ (upper line), and **C.** Number of generations (t_S), with $N_S = 10{,}000$ and $D = 5$ (lower line) or $D = 10$ (upper line). In all cases $N_P = 100{,}000$; $t_P = 10{,}000$; $\mu = 10^{-5}$.

As expected, a reduction in population size during the glacial phases will tend to reduce the mean coalescence times and consequently the mean number of segregating sites (Figure 2A). The increase in subdivision, on the other hand, tends to increase the mean coalescence times and the mean number of segregating sites.

In Figure 2B we observe that the mean number of segregating sites increases with the number of demes. We plotted the curve for two different cases: keeping the value of N_S constant or keeping the value of $D.N_S$ constant. In the first case, as the number of demes increases, the total population size ($D.N_S$) also increases. In the second case, the total population size is kept constant, and increasing D means subdividing the population into smaller demes. It is inter-

esting to observe that both curves behave almost identically. Especially, that a greater number of demes, but with smaller size each, will still cause a large increase in coalescence time and genetic variation. We have investigated this with other parameter values, and a similar behavior was observed. This indicates that the number of demes is important in maintaining genetic diversity during periods of population reduction - when migration between them is negligible - even when each deme is of small size.

The duration of the structured phases also influences coalescence time and genetic variation. In Figure 2C we observe that an increase in t_S generally causes an increase in the mean number of segregating sites. In the case where $D = 5$, there is a small range of the plot where the mean number of segregating sites decreases with increasing t_S. However, it increases in most of the plot when $D = 5$, and throughout the plot for larger values of D (shown here for $D = 10$).

Having studied the influence of each parameter of the glacial phases, we now turn to a study of the conditions in which the coalescence times and genetic variation are increased or decreased, in comparison to what is expected for a panmictic population of constant size N_P. With the values chosen ($N_P = 100{,}000$, $\mu = 10^{-5}$), the expected number of segregating sites would be $E[S] = 2$.

In Figure 3 we plotted $E[S]$ as a function of the number of demes (D) and the total population size during the structured phases ($D.N_S$). Figure 3A is a three-dimensional plot, and Figure 3B is a contour-plot with two levels: above and below $E[S] = 2$. We observe that when $D \geq 2$, the mean number of segregating sites is greater than two, that is, greater than what would be expected in the case where the population is always panmictic with constant size N_P. Even when the total population size and the deme size in the glacial phases are quite small (Figure 3B, bottom), still the expected genetic variation is increased if the population is subdivided into two or more demes.

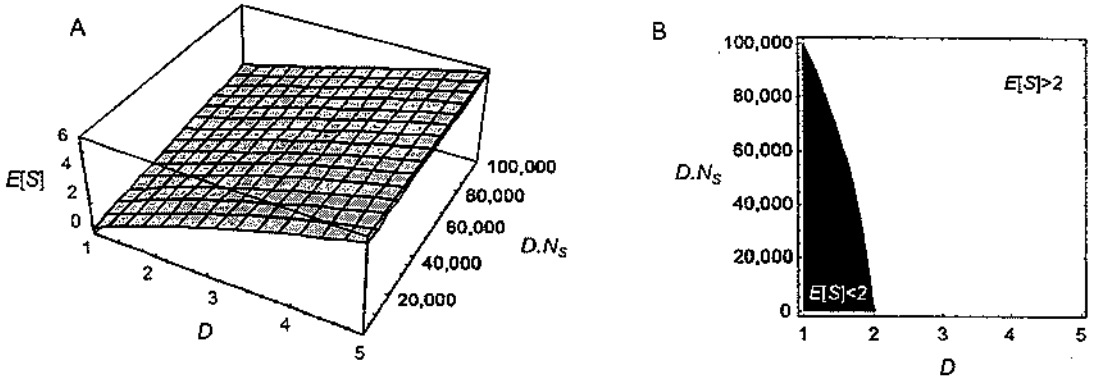

Figure 3. Mean number of segregating sites ($E[S]$) as a function of the number of demes (D) and the total population size ($D.N_S$) during the structured phases, with $N_P = 100{,}000$; $t_P = 10{,}000$; $t_S = 100{,}000$, and $\mu = 10^{-5}$. **A.** Three-dimensional plot; **B.** Contour-plot with the region where $E[S]<2$ in black and $E[S]>2$ in white.

We also plotted the mean number of segregating sites as a function of the duration (t_S) and deme size (N_S) during the structured phases, fixing the number of demes at $D = 2$ (see Figure 4). We observe that, when t_S is large, the mean number of segregating sites is greater than two even with very small values of N_S (Figure 4B, bottom right corner). Paleoclimatic

studies support the notion that the glacial periods lasted much longer (~80,000-90,000 years) than the interglacials (~10,000) (Petit et al., 1999; Jouzel, 2003). Therefore, in species with one generation/year, for instance, such large values of t_s are highly plausible.

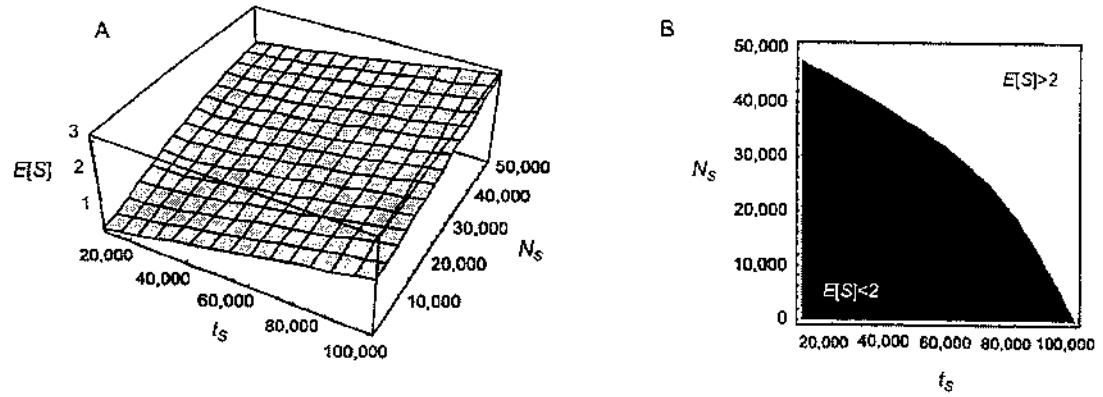

Figure 4. Mean number of segregating sites ($E[S]$) as a function of the time in generations (t_s) and the size of each deme (N_s) during the structured phases, with N_p = 100,000; t_p = 10,000; D = 2, and μ = 10^{-5}. **A.** Three-dimensional plot; **B.** Contour-plot with the region where $E[S]<2$ in black and $E[S]>2$ in white.

Most molecular studies of species which have undergone the cyclic changes of the Quaternary have focused on the genetic consequences of the last cycle only. However, in several species the most recent common ancestor at some loci predates the last glacial period, sometimes tracing back even to the Pliocene (>2 mya) (e.g., Paulo et al., 2001; Hewitt, 2004; Smith and Farrell, 2005). The results obtained here suggest that population subdivision, especially during the long glacial periods, is a possible explanation for these long gene genealogies.

Population subdivision is known to cause an increase in genetic variation at the species level. Based on the island model, for instance, Wright (1951) concluded that a subdivided population is capable of preserving a greater number of alleles than a well-mixed population of the same size. Nei and Takahata (1993), studying mean coalescence times in a finite island-model, found that the effective size of a species can be larger when migration rates are low. In the present study, we have shown that, even when the increase in population structure - by means of subdivision - is accompanied by a simultaneous reduction in population size, it can still cause an increase in expected coalescence time and levels of genetic variation.

The model presented here has some quite unrealistic assumptions, such as the total absence of population subdivision during interglacial periods as well as the absence of migration during the glacial periods. The investigation of a more general model, where these assumptions are relaxed, is the subject of an ongoing study. Nonetheless, the results obtained with the present model are useful in pointing to the importance of population subdivision in the history of species subject to climatic cycles, and offer a potential explanation for some of the observed patterns in molecular data.

ACKNOWLEDGMENTS

We thank all members of the Wakeley Lab, and especially Tee Muirhead, for discussions and helpful comments. J. Wakeley was supported by a Career Award (DEB-0133760)

from the National Science Foundation and F.F. Jesus was the recipient of graduate fellowships from CAPES (PDEE) and FAPESP. Research supported by FAPESP (No. 04/09742-0).

REFERENCES

Austerlitz F, Jung-Muller B, Godelle B and Gouyon P (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theor. Popul. Biol.* 51: 148-164.

Burnham RJ and Graham A (1999). The history of Neotropical vegetation: new developments and status. *Ann. MO Bot. Gard.* 86: 546-589.

Charlesworth B, Charlesworth D and Barton NH (2003). The effects of genetic and geographic structure on neutral variation. *Annu. Rev. Ecol. Evol. Syst.* 34: 99-125.

Excoffier L (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol. Ecol.* 13: 853-864.

Fisher RA (1930). The genetical theory of natural selection. Clarendon, Oxford, England.

Hewitt G (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405: 907-913.

Hewitt GM (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond. B* 359: 183-195.

Hudson RR (1983). Testing the constant rate neutral allele model with protein sequence data. *Evolution* 37: 203-217.

Hudson RR (1990). Gene genealogies and the coalescent process. In: Oxford surveys in evolutionary biology (Futuyma D and Antonovics J, eds.). Vol. 7. Oxford University Press, Oxford, England, 1-44.

Hugall A, Moritz C, Moussalli A and Stanisic J (2002). Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proc. Natl. Acad. Sci. USA* 99: 6112-6117.

Jouzel J (2003). Past climate (the last 400 ka): from geological times to future climate change. *C. R. Geoscience* 335: 509-524.

Kimura M (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61: 893-903.

Kingman JFC (1982a). The coalescent. *Stochastic Processes Applications* 13: 235-248.

Kingman JFC (1982b). On the genealogy of large populations. *J. Appl. Probab.* 19A: 27-43.

Lessa EP, Cook JA and Patton JL (2003). Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proc. Natl. Acad. Sci. USA* 100: 10331-10334.

Nei M and Takahata N (1993). Effective population size, genetic diversity, and coalescence time in subdivided populations. *J. Mol. Evol.* 37: 240-244.

Nei M, Maruyama T and Chakraborty R (1975). The bottleneck effect and genetic variability in populations. *Evolution* 29: 1-10.

Paulo OS, Dias C, Bruford MW, Jordan WC, et al. (2001). The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. *Proc. R. Soc. Lond. B Biol. Sci.* 268: 1625-1630.

Petit JR, Jouzel J, Raynaud D, Barkov NI, et al. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* 399: 429-436.

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, et al. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563-1565.

Ray N, Currat M, Excoffier L (2003). Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.* 20: 76-86.

Slatkin M and Hudson RR (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129: 555-562.

Smith CI and Farrell BD (2005). Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate changes. *Mol. Ecol.* 14: 1025-1044.

Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Webb T and Bartlein PJ (1992). Global changes during the last 3 million years - Climatic controls and biotic responses. *Annu. Rev. Ecol. Syst.* 23: 141-173.

Wright S (1931). Evolution in Mendelian populations. *Genetics* 16: 97-159.

Wright S (1938). Size of population and breeding structure in relation to evolution. *Science* 87: 430-431.

Wright S (1951). The genetical structure of populations. *Ann. Eugen.* 15: 323-354.

APPENDIX

In order to obtain an expression for the mean coalescence time, we made use of the fact that the expectation of a sum of random variables is equal to the sum of the expectations of each random variable. The expected mean coalescence time is the sum of the contributions of each phase, which we will call here X_1, X_2, X_3, etc. We need to obtain the sum of the $E[X_i]$'s.

For the first phase - panmictic - the contribution can be divided into two cases: if coalescence happens already during this phase, and if it does not. If it does not occur, the contribution of this phase for the coalescence time is simply t_P. If it does, the contribution is the number of generations that have passed until the coalescence event. Summing these parts multiplied by their respective probabilities, and simplifying, we obtain:

$$E[X_1] = \tau_p.e^{-\tau_p} + \int_{t=0}^{\tau_p} t.e^{-t}dt = 1 - e^{-\tau_p} \qquad \text{(Equation A1)}$$

where $\tau_P = t_P / N_P$. It is important to note that, in this calculation, we used the continuous-time approximation (see Hudson, 1990), with time measured in units of N_P generations. The contribution of this phase in units of generations is, therefore:

$$E[X_1] = N_p.(1 - e^{-t_p / N_p}) \qquad \text{(Equation A2)}$$

For the second phase (structured), the contribution will be equal to zero if the two lineages have already coalesced during the previous phase. If they have not coalesced (probability equal to $e^{-\tau}$), the two lineages will be randomly distributed among the demes. The probability that they will end up in a single deme is 1/D, and in separate demes 1-1/D. Since we are assuming that there is no migration, two lineages in separate demes have zero probability of coalescing. In this case, the contribution for the mean coalescence time is t_S. And in the case where the lineages end up in the same deme, the contribution is calculated in the same way as for the panmictic phase, using t_S and N_S instead of t_P and N_P. Therefore, the contribution of this phase for the mean coalescence time is:

$$E[X_2] = e^{-t_p / N_p}[(1-1/D).t_S + (1/D).N_S.(1 - e^{-t_S / N_S})] \qquad \text{(Equation A3)}$$

For the third phase, panmictic, we only need the probability that the lineages have not coalesced yet, multiplied by the result of the first phase (Equation A2). The probability that the lineages have not coalesced by the beginning of the third phase is:

$$P[\text{no coal. in 1st or 2nd phase}] = e^{-t_p / N_p}.[1 - (1/D) + (1/D).(e^{-t_S / N_S})] = \\ = e^{-t_p / N_p}.(D - 1 + e^{-t_S / N_S}) / D \qquad \text{(Equation A4)}$$

The contribution of the third phase is then:

$$E[X_3] = [e^{-t_p / N_p}.(D - 1 + e^{-t_S / N_S}) / D].N_p.(1 - e^{-t_p / N_p}) \qquad \text{(Equation A5)}$$

Likewise, the contribution of the fourth phase is:

$$E[X_4]=[e^{-t_p/N_p}.(D-1+e^{-t_S/N_S})/D].e^{-t_p/N_p}[(1-1/D).t_S+(1/D).N_S.(1-e^{-t_S/N_S})] \quad \text{(Equation A6)}$$

For each odd phase, then, the contribution for the mean coalescence time is of the form:

$$E[X_i]=[e^{-t_p/N_p}.(D-1+e^{-t_S/N_S})/D]^{\frac{i-1}{2}}N_p.(1-e^{-t_p/N_p});\ i \text{ odd} \quad \text{(Equation A7)}$$

And in each even phase, the contribution is of the form:

$$E[X_i]=[e^{-t_p/N_p}.(D-1+e^{-t_S/N_S})/D]^{\frac{i-2}{2}} . e^{-t_p/N_p}[(1-1/D).t_S+(1/D).N_S.(1-e^{-t_S/N_S})];\ i \text{ even} \quad \text{(Equation A8)}$$

Combining both, and summing for each phase, we obtain the mean coalescence time:

$$E[\text{coal. time}]=\left\{\sum_{i=0}^{\infty}\left[e^{-t_p/N_p}.(D-1+e^{-t_S/N_S})/D\right]^i\right\} .\left\{N_p.(1-e^{-t_p/N_p})+e^{-t_p/N_p}[(1-1/D).t_S+(1/D).N_S.(1-e^{-t_S/N_S})]\right\} \quad \text{(Equation A9)}$$

The sum present in the previous equation is a geometric series of kind:

$$\sum_{i=0}^{\infty}a^i \quad \text{(Equation A10)}$$

with $|a|<1$, which converges to $1/(1-a)$. This allows one more simplification:

$$E[\text{coal. time}] = \left[\frac{D}{D-(D-1)e^{-t_p/N_p}-e^{-(t_p/N_p+t_S/N_S)}}\right] .\left\{N_p.(1-e^{-t_p/N_p})+e^{-t_p/N_p}[(1-1/D).t_s+(1/D).N_S.(1-e^{-t_S/N_S})]\right\} \quad \text{(Equation A11)}$$

This is then our Equation 1, presented beforehand in the main text.

CAPÍTULO 2

PAIRWISE COALESCENCE TIMES IN A MODEL OF REPEATED GLACIAL CYCLES

Pairwise coalescence times in a model of repeated glacial cycles

Authors:
Flavia F. Jesus[1,3], Jon F. Wilkins[2], Vera N. Solferini[1], John Wakeley[3].

Addresses:
[1] Departamento de Genética e Evolução, Instituto de Biologia, Unicamp, Campinas, SP 13083-970, Brazil.
[2] Santa Fe Institute, Santa Fe, NM 87501 USA
[3] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

Keywords: refugia, Pleistocene, repeated bottlenecks, island model, F_{ST}, pairwise differences.

Correspondence: Flavia F. Jesus
Departamento de Genética e Evolução
IB, Unicamp
CP 6109
Campinas SP, 13083-970
Brazil
Phone: +55 (19) 3788-6239
Fax: +55 (19) 3788-6235
e-mail: ffjesus@unicamp.br
flaviaf.jesus@gmail.com

Running title: *coalescence times and repeated glacial cycles*

Abstract

The climatic cycles of the Quaternary have influenced the population size and structure of numerous species. In many cases gene genealogies have been observed for which the most recent common ancestor predates the last glacial period. This points to the importance of taking multiple cycles into account in theoretical studies. In the present study, we investigate a model of repeated glacial cycles. We model demographic history as two alternating phases corresponding to glacial and interglacial periods. We assume an island model of subdivision in both phases, with cyclic changes in the number of demes, deme size and migration rate. We obtain expressions for the distribution and expectation of coalescence times for a pair of genes, sampled either from the same deme or from separate demes. The distributions observed are quite different from what is expected under a panmictic population of constant size. Reduced deme size during glacial periods produces multiple peaks in the distribution of coalescence times and a reduction in the mean times. Increased population structure (reduced migration rates) increases the mean coalescence times and attenuates the peaks of the distribution. Mean coalescence times also generally increase with the number of demes during glacial periods. Our results help to clarify the genetic consequences of glacial cycles and the role of population structure in maintaining genetic variation. The approach developed here can also be used to study the genetic consequences of cycles with shorter time scales, *e.g.* seasonal cycles.

Introduction

The Quaternary period has been characterized by great climatic changes, with an alternation between glacial and interglacial periods. Paleoclimatic studies indicate that these cyclic fluctuations are related to periodic changes in the Earth's orbit (Jouzel 2003, Petit et al. 1999, Tzedakis et al. 1997). Each complete cycle lasts approximately 100,000 years, with long glacial periods interrupted by shorter (usually less than 10,000 years) interglacial periods, and some smaller fluctuations nested within each period. The present time is an interglacial, and began about 11,000 years ago (Petit et al. 1999). These climatic fluctuations have influenced the distributions of organisms (Burnham & Graham 1999, Webb & Bartlein 1992). Many species have undergone population reduction and fragmentation, becoming restricted to refugia during the glacial periods and expanding again during the interglacial periods. The biotic and abiotic changes associated with these cycles have been most pronounced in temperate and high latitude regions, but have also been significant in the tropics, where the cooler and drier times allowed an expansion of grassland and savannah, with a reduction and possible fragmentation of forest habitats (Hewitt 2004, 2000; Lessa et al. 2003 and references therein).

Changes in population size and structure may be detectable in DNA sequences. In fact, there are a number of studies of genetic diversity in species thought to have undergone these cyclic changes (*e.g.* McLachlan et al. 2005, Galbreath & Cook 2004, Lessa et al. 2003, Petit et al. 2003, Hugall et al. 2002). These studies have focused particularly on the existence and number of refugia, as well as their impact on current genetic variation. While most of the analytic tools available focus only on the genetic consequences of the last cycle, it is not uncommon for the most recent common ancestor at certain loci to predate the last glacial period in some species whose distribution and abundance are known to have been affected by glaciation (Hewitt 2004). For example, the times to common ancestry for mtDNA in two species of flightless beetles (genus *Moneleima*) from desert regions of North America have been estimated at upwards of 1 to 2 Myr (Smith & Farrell 2005). Another example is the Iberian lizard, *Lacerta schreiberi*, in which mtDNA lineages appear to trace back to the Pliocene (> 2 Myr ago) (Paulo et al. 2001). In addition, some phylogenetic studies indicate a pre-Pleistocene origin of several

taxa (e.g. Klicka & Zink 1997, for North American birds; Moritz et al. 2000, for tropical forest faunas).

These studies suggest that many of the species extant today may have experienced multiple glacial cycles, and might carry some genetic signal of this history. For this reason, we have investigated the distribution of coalescence times between pairs of sequences in a model of glacial cycles. The aim is to describe the possible genetic signatures of cyclical changes in population size and structure. The model is a great simplification of the complex histories of most species, but it allows us to assess the effects of glacial cycles quantitatively. The two key features of glacial cycles we wish to understand are habitat reduction and habitat fragmentation during glacial periods. These are expected to affect both the number of individuals the habitat can carry (*i.e.* population size) and the distribution of individuals across the landscape (*i.e.* population structure).

There have been many theoretical studies of the effects of changes in population size, and to a lesser extent of changes in population structure. The effects of bottlenecks have long been of interest in population genetics, and one of the first accounts of the effect of bottlenecks in genetic variation was motivated by cyclic population fluctuations. One of Wright's (1938) well known results is that, when the cycles are of short length – in generations – the inbreeding effective population size is well approximated by the harmonic mean of the sizes over time. Later studies described the effects of a single bottleneck on heterozygosity and number of alleles (Nei et al. 1975, Maruyama & Fuerst 1984, 1985a) under an infinite-alleles model. The effect of periodic bottlenecks was studied by Maruyama & Fuerst (1985b), using a model of fixed cycle length, for which they described the fluctuations of heterozygosity through time.

In the context of coalescent theory (Kingman 1982a,b, Hudson 1983, Tajima 1983) and assuming the infinite sites model of mutation (Kimura 1969), Tajima (1989b) studied the expected number of segregating sites and the mean number of pairwise differences in a population subjected to a sudden change in size or a single bottleneck. Slatkin & Hudson (1991) studied the distribution of pairwise sequence differences in a model of exponential growth, and Rogers and Harpending (1992) considered this distribution in a model of sudden expansion. In addition, computational methods have been developed for the

estimation of population size and growth parameters, using information from the gene genealogies (Griffiths & Tavaré 1994, Kuhner et al. 1998).

The impact of population structure - in the form of subdivision and migration - on genetic diversity has been the focus of numerous studies. Among the models of population subdivision and migration, the island model (Wright 1931) is the best studied one, due in large part to its mathematical simplicity. In this model, the population is subdivided into a number of sub-populations, or demes, all of which exchange migrants with one another. In general, a lower rate of migration between demes results in a reduction of genetic diversity within each deme and an increase in genetic differentiation between demes. Based on the island model, Wright (1951) concluded that a subdivided population is capable of preserving a greater number of alleles than a well-mixed population of the same size. That is, one possible consequence of subdivision is the accumulation of additional genetic diversity.

Wright's formulation of the island model assumed an infinite number of sub-populations, or demes, but it was later modified to include a finite number (Maruyama 1970, Latter 1973). Slatkin (1991) obtained expressions for inbreeding coefficients in terms of pairwise coalescence times for this and other models. Nei & Takahata (1993) derived a formula for an effective population size in the finite island model, using mean coalescence times, and found this effective size to be larger in populations with small migration rates. A thorough review of the literature on the genetics of subdivided populations was recently published by Charlesworth et al. (2003).

A few recent studies have been concerned with models of population expansion or colonization that include population structure. Austerlitz et al. (1997) studied colonization in a linear stepping-stone model, starting from a single deme. They obtain the distribution and expectation of coalescence times for two genes, and also study the evolution of F_{ST} through time. They observe a founder effect that increases with distance from the starting deme, and generally decreases with a higher migration or growth rate. They also study the effect of a bottleneck within a single deme. Of relevance to the work we present here, they observed a multi-modal distribution of coalescence times for a pair of genes, with a peak corresponding to the time of the bottleneck. Austerlitz et al. (2000)

simulated the colonization process in a two-dimensional stepping-stone model. They also studied the impact of overlapping generations and a long juvenile phase in the distribution of genetic variation, in order to explain the observed patterns in genetic data from trees and annual plants. Ray et al. (2003) published a simulation study of a spatial expansion from a single deme in a two-dimensional stepping-stone model, recording several aspects of intra-deme molecular diversity. Excoffier (2004) studied a sudden expansion with subdivision, in an infinite-island model, and obtained analytical expressions for the distribution of the number of differences between a pair of sequences, sampled after the expansion. A very recent study by Ewing and Rodrigo (2006) presents a method for parameter estimation that allows changes in several population parameters through time, including the number of demes. The method is developed for serial samples, that is, samples from different points in time, and it can be used to study colonization – and other processes – when such samples are available.

Motivated by these empirical and theoretical results, we study coalescence times in a model of cyclical changes in population size and structure. In a previous study (Jesus et al. 2006, in press) we investigated a simple model, where a panmictic and a structured phase – with no migration – alternate, and obtained an expression for the expected coalescence time of two genes sampled in the present interglacial. In that model, the increase in mean coalescence time due to population subdivision was capable of outweighing the effect of population reduction during glacial periods. In the present study, we investigate a more general model, allowing for population subdivision and migration in both phases. We are particularly concerned with two questions. First, how does the distribution of the time to coalescence under cyclical changes in population size and structure differ from that under the standard Kingman (1982a) coalescent? Second, what is the role of population structure in maintaining genetic diversity through periods of glaciation? We investigate these questions quantitatively under the assumption that the population is structured according to the finite island model, with cyclical changes. Specifically, we obtain expressions for the distribution of the time to coalescence for two genes sampled from the same deme and from different demes. The expressions are functions of the matrices that describe coalescence during each phase, and are easy to evaluate numerically. We plot these functions over different sets of parameters, reflecting

different biological scenarios. Expressions for the expected coalescence times are also obtained, and we use these to investigate the opposing effects of population reduction and increased structure during glacial periods.

Model and Theory

The model forward in time

We model demographic history as an alternation of two distinct phases which we will call phase *I* and phase *G*, for interglacial and glacial periods, respectively (see Fig. 1 for a representation of the model backwards in time). We are currently experiencing an interglacial period, so the most recent phase is of type *I*. The population may be subdivided during both phases, and we assume that population structure follows a symmetric island model (Wright 1931) with a finite number of demes (*e.g.*, Latter 1973). The number of demes, deme size and migration rate can differ between the two phases, permitting examination of the joint effects of cyclic change in both population size and structure.

Each phase lasts several generations, and transitions between phases occur in a single generation. One complete cycle is made up of $t_I + t_G$ generations: $t_I - 1$ for phase *I*, $t_G - 1$ for phase *G*, and one generation for each one of the transitions. Each generation includes a reproduction step followed by a migration step. Generations are non-overlapping with reproduction occurring within demes according to the Wright-Fisher model (Fisher 1930, Wight 1931), with offspring sampled from an infinite "gamete" pool. We consider here a haploid organism, but the conclusions can be extended to diploid organisms in the usual way. In addition to the times, t_I and t_G, the model has six parameters, which are the number of demes (D), the deme size (N), and the migration fraction (m), with subscripts I and G to denote the phase. Migration is modeled as follows: a fraction m of the individuals in a deme are drawn from a migrant pool, to which all demes contribute equally. Looking backwards in time, each ancestral lineage "migrates" with probability m to a randomly chosen deme, which could be the deme from which it was derived (probability $1/D$). This kind of migration produces the same results as the case when

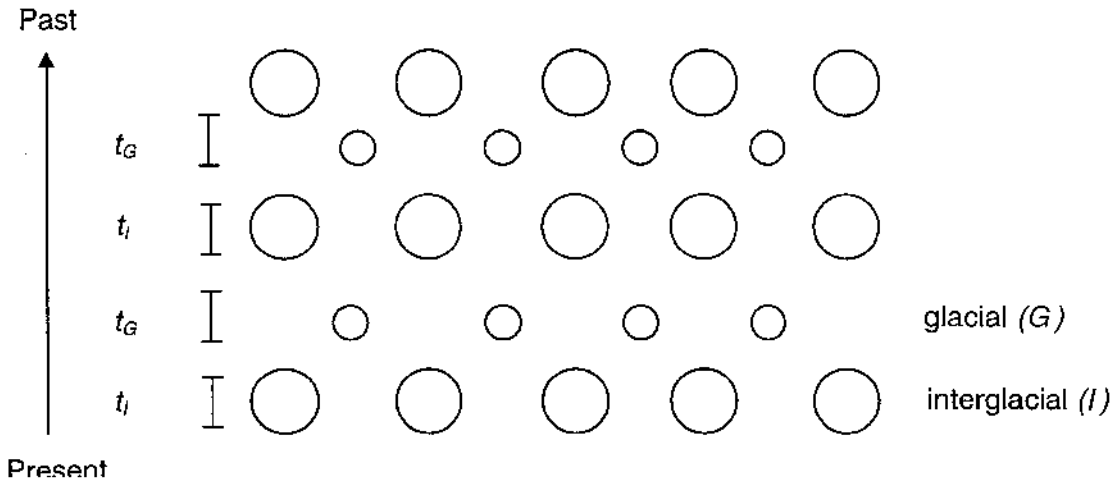

Figure 1. Schematic representation of the model, with two alternating phases: *I* (interglacial) and *G* (glacial). The parameters t_I and t_G represent the duration of each phase, in generations. Although the demes are aligned in this figure, an island model of population subdivision is assumed (see text).

lineages must migrate to a different deme if our m is replaced by $m(D\text{-}1)/D$ (Takahata & Nei 1984).

Since we are interested in habitat reduction during glacial periods, we restrict our attention to cases in which $N_G < N_I$ and $N_G D_G < N_I D_I$, *i.e.* when the deme size and the total population size are smaller during glacial periods. With regard to fragmentation, we are interested in two cases: (1) $D_G > D_I$, so that fragmentation occurs during the glacial periods resulting in an increased number of demes; and (2) $D_I > D_G$, so that the glacial periods cause a reduction in the number of demes by extinction. The details of how the transitions between phases occur will depend on whether we are in case (1) or in case (2). These details, presented below, were chosen for simplicity and also to take into account the increased chance of coalescence that occurs when there is a transition to a smaller number of demes. However, the overall results do not seem to be very sensitive to these details: we investigated a number of other ways these transitions might happen and the results were similar (not shown).

In case (1), where the number of demes is greater during the glacial periods, the transition from interglacial to glacial involves subdivision of each deme into d new demes. Thus, we have $D_G = dD_I$. For simplicity, we assume that this subdivision occurs via the infinite gamete pool during the reproduction step. Specifically, each deme in phase I forms an infinite gamete pool, and from this pool d demes are formed, each one by a random sample of N_G individuals/gametes. In the transition from glacial to interglacial (G to I), each set of d demes that originated from a single deme in the previous phase is fused into a single deme again, also via the gamete pool. The goal here is to model the effect of the demes expanding after the glacial period and thus becoming fused. Specifically, each deme forms an infinite gamete pool, and from each set of d pools one of the D_I demes is formed.

In case (2), where the number of demes is smaller during glacial periods, we allow a fraction of interglacial demes to go extinct in the transition to the glacial phase. In the transition from I to G, D_G of the D_I demes are selected at random to persist. These demes reproduce normally. The remaining D_I–D_G demes go extinct. In the transition from G to I, each of the D_G demes reproduces to become one of the D_I demes in the next phase. For

each of the remaining D_I–D_G demes, one of the D_G demes is chosen at random as a parent, and the new deme is formed by sampling with replacement from that parent. Therefore, some of the D_G demes will give rise to only one deme, and some to more than one. It is important to note that, in the special case when each parameter value is equal for both phases ($D_I = D_G$, $N_I = N_G$, $m_I = m_G$), our model reduces to a standard finite island model.

Coalescence probabilities

We have studied the distribution of coalescence times for a pair genes sampled either from a single deme or from two different demes. During each phase, the ancestry of the sample can be followed back in time using a discrete-time Markov chain. The three possible states for the sampled lineages are: (1) being in the same deme, (2) being in separate demes or (3) having coalesced. The one-generation transition matrix for an interglacial phase is

$$\mathbf{P}_I = \begin{pmatrix} \left(\alpha_I + \frac{1-\alpha_I}{D_I}\right)\left(1-\frac{1}{N_I}\right) & (1-\alpha_I)\left(1-\frac{1}{D_I}\right) & \left(\alpha_I + \frac{1-\alpha_I}{D_I}\right)\left(\frac{1}{N_I}\right) \\ \left(\frac{1-\alpha_I}{D_I}\right)\left(1-\frac{1}{N_I}\right) & 1-\frac{1-\alpha_I}{D_I} & \frac{1-\alpha_I}{D_I N_I} \\ 0 & 0 & 1 \end{pmatrix}, \qquad (1)$$

where $\alpha_I=(1-m_I)^2$ is the probability that neither of the lineages is a migrant. Each of the entries $(\mathbf{P}_I)_{ij}$ of the matrix is the probability that two lineages, now in state i, were in state j in the previous generation. For instance, the probability that two lineages sampled from the same deme coalesce in one generation is $(\mathbf{P}_I)_{13}$. This matrix is equivalent to Equation (2) in Wakeley (2004), where it is also explained.

The one-generation transition matrix for each glacial phase has the same form as the one for interglacial, with the appropriate substitutions of parameters from type I to type G

$(N_G,\ \alpha_G=(1-m_G)^2,\ D_G)$. It is important to note that we are not assuming any particular values for the parameters, which can take on arbitrary values, from small to large. In particular, the migration rate can take any value in the range from zero to one

If our model consisted of only one phase – type *I* for instance – we would already have all the elements in order to obtain the coalescence time probabilities. For a sample taken from one deme ('within'), the probability of coalescence at exactly generation t in the past would be simply

$$f_w(t) = (\mathbf{P}_I^{\,t})_{13} - (\mathbf{P}_I^{\,t-1})_{13}. \tag{2}$$

The term $(\mathbf{P}_I^{\,t})_{13}$ is the probability that the system is in state (3), that is, the lineages are already coalesced, in generation t in the past, given that they are sampled from the same deme (state (1)). For a sample obtained from two separate demes ('between'), the probability of coalescence at generation t would be

$$f_b(t) = (\mathbf{P}_I^{\,t})_{23} - (\mathbf{P}_I^{\,t-1})_{23}. \tag{3}$$

These functions are valid in the present model, but only for the first type *I* phase, that is, only for $t < t_I$. For earlier times ($t \geq t_I$), other expressions are necessary. Our model assumes two alternating phases, and the solution requires explicit consideration of the transitions between them.

In order to do this, we define two other matrices: one for a complete type *I* phase and one for a complete type *G* phase, including their transition generations. These are, respectively,

$$\Pi_I = \mathbf{P}_I^{\,t_I-1}\mathbf{T}_{IG} \tag{4}$$

and

$$\Pi_G = \mathbf{P}_G^{\,t_G-1}\mathbf{T}_{GI}, \tag{5}$$

where $\mathbf{T}_{IG}$ and $\mathbf{T}_{GI}$ are the matrices for transitions between phases, described in the Appendix (A1-A6).

The matrix Π_I accounts for a total of t_I generations, and the matrix Π_G for a total of t_G generations. We also define the matrix for one complete cycle,

$$\Pi = \Pi_I \Pi_G, \tag{6}$$

which accounts for a total of $t_I + t_G$ generations.

The coalescence probability in a given generation t depends on the type of generation. There are four possible types: (1) within a type I phase; (2) during a transition generation from I to G; (3) within a type G phase or (4) during a transition generation from G to I. The part of t that can be accounted for by whole cycles is

$$\varphi(t) = \left\lfloor \frac{t}{t_I + t_G} \right\rfloor, \tag{7}$$

in which the notation $\lfloor x \rfloor$ means the integer part of x. Thus, $\varphi(t)$ is the number of completed cycles from the current generation back to generation t in the past. By definition, $t \geq (t_I + t_G)\varphi(t)$. In order to know which case above a given t represents, we need to know the remainder

$$\tau(t) = t - (t_I + t_G)\varphi(t), \tag{8}$$

which is the part of t that cannot be accounted for in whole cycles.

The distribution of coalescence times can be expressed in terms of these quantities and the matrices defined above. These are separated in cases, according to the type of generation. For a pair of genes sampled from a single deme ('within'), the probability of coalescence at exactly t generations in the past is given by

$$f_w(t) = \begin{cases} \left(\Pi^{\varphi(t)}\mathbf{P}_I^{\tau(t)}\right)_{13} - \left(\Pi^{\varphi(t)}\mathbf{P}_I^{\tau(t)-1}\right)_{13} & \text{if } 1 \le \tau(t) < t_I \\ \left(\Pi^{\varphi(t)}\Pi_I\right)_{13} - \left(\Pi^{\varphi(t)}\mathbf{P}_I^{t_I-1}\right)_{13} & \text{if } \tau(t) = t_I \\ \left(\Pi^{\varphi(t)}\Pi_I\mathbf{P}_G^{\tau(t)-t_I}\right)_{13} - \left(\Pi^{\varphi(t)}\Pi_I\mathbf{P}_G^{\tau(t)-t_I-1}\right)_{13} & \text{if } \tau(t) > t_I \\ \left(\Pi^{\varphi(t)}\right)_{13} - \left(\Pi^{\varphi(t)}\Pi_I\mathbf{P}_G^{t_G-1}\right)_{13} & \text{if } \tau(t) = 0. \end{cases} \quad (9)$$

For a pair of genes sampled from two different demes ('between'), the probability of coalescence at exactly t generations in the past is given by

$$f_b(t) = \begin{cases} \left(\Pi^{\varphi(t)}\mathbf{P}_I^{\tau(t)}\right)_{23} - \left(\Pi^{\varphi(t)}\mathbf{P}_I^{\tau(t)-1}\right)_{23} & \text{if } 1 \le \tau(t) < t_I \\ \left(\Pi^{\varphi(t)}\Pi_I\right)_{23} - \left(\Pi^{\varphi(t)}\mathbf{P}_I^{t_I-1}\right)_{23} & \text{if } \tau(t) = t_I \\ \left(\Pi^{\varphi(t)}\Pi_I\mathbf{P}_G^{\tau(t)-t_I}\right)_{23} - \left(\Pi^{\varphi(t)}\Pi_I\mathbf{P}_G^{\tau(t)-t_I-1}\right)_{23} & \text{if } \tau(t) > t_I \\ \left(\Pi^{\varphi(t)}\right)_{23} - \left(\Pi^{\varphi(t)}\Pi_I\mathbf{P}_G^{t_G-1}\right)_{23} & \text{if } \tau(t) = 0. \end{cases} \quad (10)$$

These expressions depend on the population parameters (N_I, m_I, D_I, N_G, m_G and D_G) and the time of each complete phase (t_I and t_G). It is often desirable to write expressions such as these as explicit functions of the demographic parameters, in order to gain some

intuition about the influence of each one. However, this is not practical for this model, due to the complexity and length of the resulting equations. Instead, we have written a computer routine to obtain values for the above functions and plot graphs from them, by substituting the parameter values in the matrices and using the above equations to obtain the distribution numerically. The routine to perform this calculation was written in Mathematica (Wolfram Research 2001) and is available from the authors upon request.

The functions above are the probability mass functions of coalescence time. The cumulative distributions are given by the first term on the right hand side of the equation for each case. Expressions for the mean coalescence time and expected value of F_{ST} are given in the Appendix.

Results

The effect of population reduction

To isolate the effect of population reduction during glacial periods, we fixed all the other parameters as being equal in the two phase types, allowing a difference only in the population sizes such that $N_G < N_I$. We assumed a panmictic population in both phases, so that this version of the model is equivalent to other models of repeated bottlenecks. The size reduction in the glacial phases causes an increase in the coalescence probabilities during those phases, relative to the ones during the interglacial phases. This increase causes peaks and sharp discontinuities in the plots, in contrast with the strictly decreasing geometric or exponential curve expected for a panmictic population of constant size. This effect is more pronounced for greater reductions (when $N_G << N_I$). Figure 2A shows the coalescence probability distribution for the case $N_I = 10\ N_G$. Only one plot is shown since, in a panmictic population, sampling within or between 'demes' yields the same results. Note that we have used $D_G = D_I = 10$, but with both migration rates equal to one, so that the populations behave as panmictic ones of sizes $N_I D_I = 10{,}000$ and $N_G D_G = 1{,}000$. The mean coalescence time for this case – assuming $t_I = t_G = 100$ – is 1,859 generations, which is quite close to the harmonic mean approximation of 1,818.

We also studied the effect of the same population reduction when both phase types are structured. In these calculations we chose values for the parameters such that an

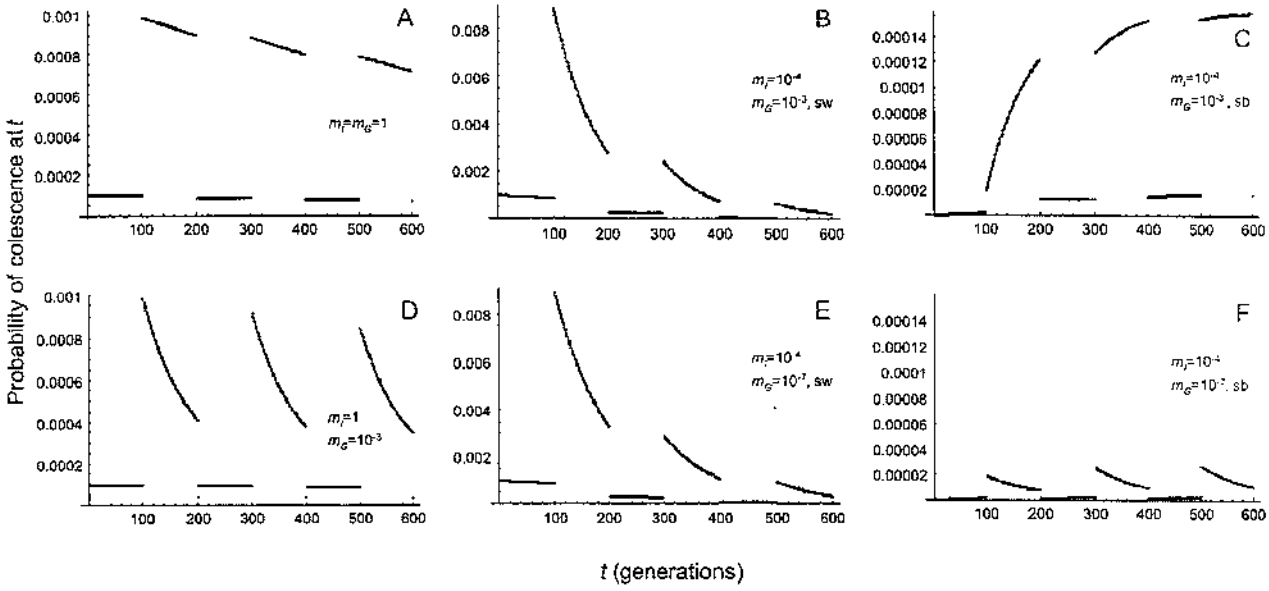

Figure 2. Probabilities of coalescence, with parameters N_I=1000; N_G=100; D_I=D_G=10 and t_I=t_G=100; 'sw' and 'sb' stand for sampling within and sampling between demes, respectively. Migration rates are given above. (A) Both phases panmictic; (B) & (C) both phases structured ($N_I m_I$=$N_G m_G$=0.1): (B) sampling within one deme; (C) between demes; (D) phases *I* panmictic and *G* structured ($N_G m_G$=0.1); (E) & (F) both phases structured, with *G* much more structured ($N_I m_I$=0.1; $N_G m_G$=10^{-5}): (E) sampling within one deme; (F) between demes.

equivalent level of structure was maintained in both phases (as would be characterized by measures such as F_{ST}). Both the number of demes and the number of migrants per deme per generation were kept constant ($D_G = D_I = 10$; $N_I\, m_I = N_G\, m_G = 0.1$). The effect of population reduction is also seen here (Figs. 2B & 2C), with discontinuities and peaks in the distribution of coalescence times. The shapes of the curves are different due to structure and to the sampling scheme (also note the difference in scale between the plots). For instance, in Fig. 2B, the peak of the most recent glacial phase is much higher than in the panmictic case (Fig. 2A). This occurs because, with the low migration rate, the lineages are mostly restricted to one deme of size $N_G = 100$ in the glacial phases, whereas in Fig. 2A, because of the very high migration, they are part of one larger population of size $N_G\, D_G = 1000$. The curves in Fig. 2B descend much more quickly than in Fig. 2A also because of the low migration rate, which restricts the lineages in the demes. In Fig. 2C, the probabilities of coalescence in any generation are very low, due to the sampling between demes associated with the low migration. Also, the curves ascend in the beginning, which is typical of plots for sampling between demes. The mean coalescence time is 1,851 generations for Fig. 2B and 10,938 generations for Fig. 2C. Slatkin (1991) derived an expression for F_{ST} in terms of mean coalescence times for pairs of samples from the same deme and pairs drawn at random from the entire population. According to this equation (see Appendix), these mean coalescence times correspond to an F_{ST} value of 0.81.

The effect of increased structure

In the previous section, we considered two cases where the difference between the phases is restricted to the population size: one panmictic (Fig. 2A) and one with similar population structure in both phases (Figs. 2B & 2C). In both cases we observed that the reduction in population size assumed for the glacial phases causes discontinuities and peaks in the distribution of coalescence times. Here, we consider the effects of an increased population structure during the glacial phases (Figs. 2D – 2F). We use the same parameters of the cases described above (with $N_I = 10\, N_G$; $D_I = D_G = 10$; and $m_I = 1$ or 10^{-4}) changing only the migration rate of the glacial phases (m_G). We are still assuming a

population reduction during the glacial periods, as this appears to be a pattern common to many species.

We considered two scenarios. In both cases the level of migration was reduced in the glacial periods ($N_G\ m_G < N_I\ m_I$). The first case (Fig. 2D) is a modification of the panmictic one presented in Fig. 2A. The migration rate in the interglacial phases was kept very high, so that the population is still effectively panmictic in these phases; and a low migration rate was introduced in the glacial phases ($N_G\ m_G = 0.1$) (Fig. 2D). The other case (Figs. 2E & 2F) is a modification of the structured case presented in Figs. 2B & 2C. Migration is limited in both phases, but now it is more severely limited during the glacial periods ($N_I\ m_I = 0.1$; $N_G\ m_G = 10^{-5}$; see Figs. 2E & 2F).

Comparing Fig. 2D with Fig. 2A we note that the structure introduced during the glacial phases causes a faster decrease in the coalescence probabilities during these periods. This increases the mean coalescence time from 1,859 (Fig. 2A) to 2,617 generations (Fig. 2D).

The decrease in coalescence probabilities is also evident when we compare Fig. 2F with Fig. 2C. In these figures, both phases are structured, and the samples are drawn from separate demes. The reduction in migration increases the mean coalescence time by almost a factor of ten (101,798 generations in Fig. 2F versus 10,938 in Fig. 2C). It also changes the shape of the curves during the glacial phases: they turn into decreasing functions, because the coalescence probabilities are dominated by the proportion of lineages that are in a single deme by the end of the first interglacial (backwards in time), since the migration rate is very restricted.

For a pair of samples taken from the same deme, we also see an increase in the mean coalescence time (2,788 generations in Fig. 2E versus 1,851 in Fig. 2B), although it is much less pronounced than in the between-demes case. These mean coalescence times correspond to an F_{ST} value of 0.97 (Figs. 2E & 2F), increased from 0.81 in Figs. 2B & 2C.

Number of demes

In order to investigate the effect of changes in number of demes, we studied two scenarios. In both cases we have assumed population structure in both phases, with $N_I m_I = 0.1$ and two possible values for $N_G m_G$: 0.1 or 10^{-5}. First, we consider a case where the number of demes is reduced during the glacial periods ($D_G = 5$, $D_I = 10$; see Fig. 3). Then we consider the opposite case, where the number of demes is increased in the glacial periods ($D_G = 20$, $D_I = 10$; see Fig. 4).

We observe no dramatic qualitative effect on the distribution of coalescence probabilities. However, the mean coalescence times are strongly affected. As would be expected, they are smaller in Fig. 3 than in Fig. 2 (where $D_G = D_I = 10$). For the case where $N_G m_G = 0.1$, the mean times are 382 generations when sampling is done within a deme (Fig. 3A) and 1,266 generations for sampling between demes (Fig. 3B). These compare with 1,851 and 10,938 in the case presented in Fig. 2B and Fig. 2C, respectively. For the case where $N_G m_G = 10^{-5}$, the mean times are 255 (Fig. 3C) and 1,328 generations (Fig. 3D), for sampling within and between demes, respectively, which compare with 2,788 and 101,798 generations in Fig. 2E and Fig. 2F. The corresponding F_{ST} values are 0.68 ($N_G m_G = 0.1$; Figs. 3A & 3B) and 0.79 ($N_G m_G = 10^{-5}$; Figs. 3C & 3D).

The transitions contribute significantly to the reduction in the mean times. When the transition – backwards in time – occurs between a phase with more demes to one with fewer demes, there is an increased probability that separate lineages are joined in the same deme, increasing their probability of coalescence (see Appendix). In the case of Fig. 3, this happens at each transition from an interglacial to a glacial phase. We verified that, when the sample is taken from one deme, this mostly affects the tail of the distribution (not shown). When the sample is taken from separate demes, however, the first transition is already quite important, increasing the probability that the lineages meet. In the cases shown in Fig. 2, the rate at which lineages moved into the same deme depended only on the migration rate.

It is interesting to note that the mean time is smaller (255 generations) in Fig. 3C than in Fig. 3A (382 generations), where migration is higher. That is, increasing the structure during the glacial phases caused a decrease in the mean coalescence time in this case, in

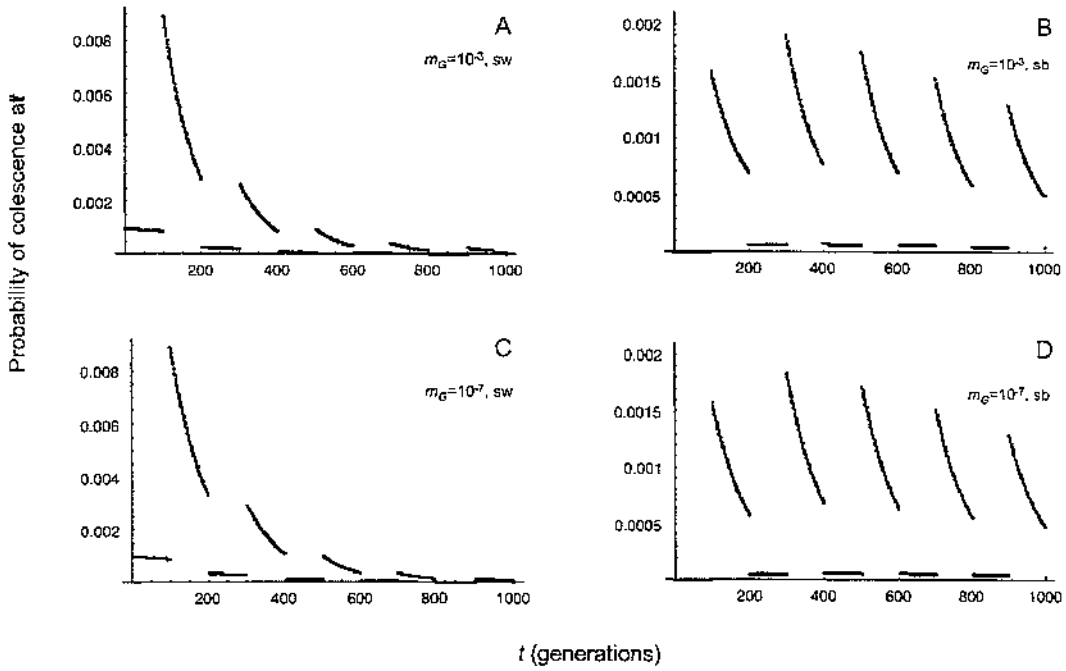

Figure 3. Probabilities of coalescence times, with D_G<D_I. Parameters: N_I=1000; N_G=100; D_I=10; D_G=5; m_I=10^{-4}; t_I=t_G=100; 'sw' and 'sb' stand for sampling within and sampling between demes, respectively. (A) & (B) m_G=10^{-3}: (A) sampling within one deme; (B) between demes; (C) & (D) m_B=10^{-7}: (C) sampling within one deme; (D) between demes.

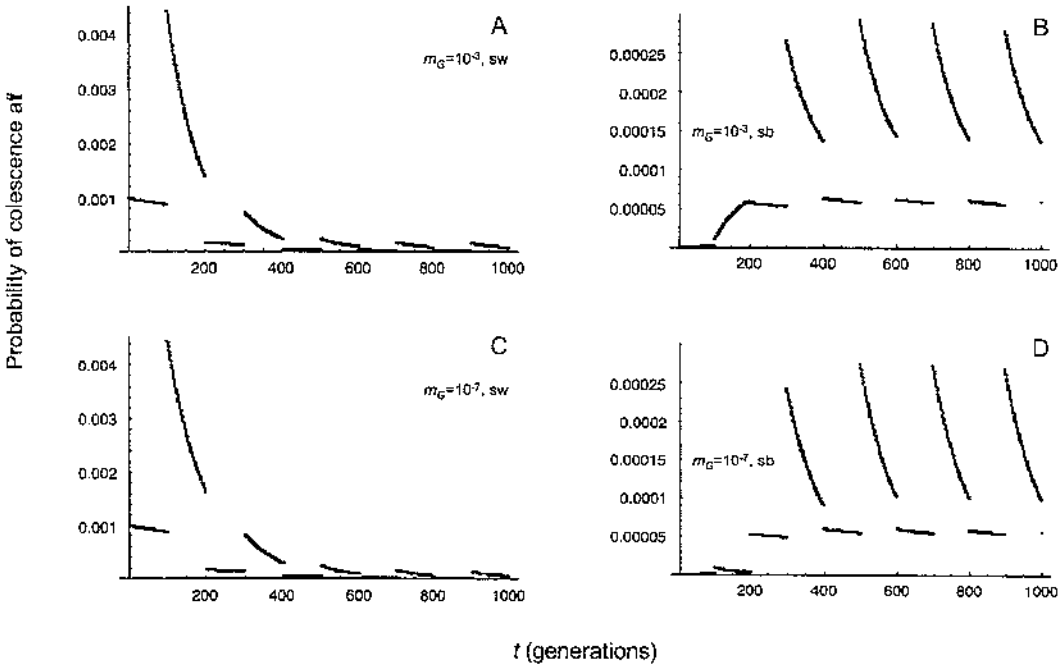

Figure 4. Probabilities of coalescence times, with D_G>D_I. Parameters: N_I=1000; N_G=100; D_I=10; D_G=20; m_I=10^{-4}; t_I=t_G=100; 'sw' and 'sb' stand for sampling within and sampling between demes, respectively. (A) & (B) m_G=10^{-3}: (A) sampling within one deme; (B) between demes; (C) & (D) m_B=10^{-7}: (C) sampling within one deme; (D) between demes.

contrast with what was seen in Fig. 2. This occurs here because the lineages are mostly trapped in one deme, and when they migrate during a glacial phase they still have a high probability of meeting, due to the low number of demes.

We now turn to the case where the number of demes increases during the glacial periods ($D_G = 20$, $D_I = 10$; see Fig. 4). When the samples are taken from one deme (Fig. 4A & 4C), the shape of the distribution is very similar to what was observed in previous cases (e.g. Fig. 2B & 2E, Fig. 3A & 3C) The scale is different, however, with the coalescence probabilities during the glacial phases smaller than in those cases. This is because the transition from interglacial to glacial phase – backwards in time – involves deme fission ($D_G > D_I$). Therefore, two lineages in the same deme during an interglacial may be split into separate demes during the transition to a glacial phase (see Appendix). This is also reflected on the mean times, which are very high: 4,397 generations in Fig. 4A and 4,717 generations in Fig. 4C

The transition from glacial to interglacial, however, entails a chance of joining separate lineages into one deme. This plays an important role in the distributions of coalescence times for samples from separate demes. We observe in Fig. 4B and Fig. 4D that it takes a few phases for the highest peaks to occur. Specifically, they occur during the glacial phases, when the demes are small, but only after the first glacial-interglacial transition has passed. The mean times are 7,310 generations (Fig. 4B) and 8,381 (Fig. 4D). These are smaller than what was seen in Fig. 2C and Fig. 2F, precisely because of this increased probability that lineages will be brought together into a single deme during the glacial-interglacial transitions (backwards in time).

In Figs. 4A – 4B, the number of migrants per deme per generation is $N_G m_G = 0.1$, and the F_{ST} value based on mean coalescence times is 0.37. In Figs. 4C – 4D, the number of migrants is $N_G m_G = 10^{-5}$ and the corresponding F_{ST} value is slightly increased to 0.41.

In general, studying the cases where the number of demes differs between the phases, we note that the increase of structure – through reduced migration – tends to cause an increase of the mean coalescence times, as in the previous cases, where the number of demes is constant. An exception, however, is shown in Fig. 3C, for samples taken from a

single deme, and a reduced number of demes during the glacial phase. This will be further discussed in the section about the mean times.

Number of generations

So far we have looked into cases where the duration of both phases is the same in number of generations (t_I=t_G=100). However, paleoclimatic studies support a much longer duration of the glacial periods (~80,000 – 90,000 years) in comparison to the interglacials (~ 10,000) (Petit et al. 1999). Therefore, we now present cases where the duration of the glacial phases is nine times that of the interglacial phases (t_G=900; t_I=100; see Figure 5). For comparison, we have kept the other parameter values equal to those used for the calculations presented in Figure 2.

We wish to highlight two aspects of these results. The first is that the discontinuity of the graphs is still observed in most cases, although it is quite reduced in Fig. 5B and Fig. 5E – where both phases are structured and sampling is within demes. This pattern also appears when $D_G \neq D_I$ (not shown).

The second aspect is regarding the mean times. In the cases where migration is equivalent in both phases (Figs. 5A – 5C), the mean times are 1,149, 1,162 and 6,630 generations, respectively. These are all smaller than their corresponding values in Fig. 2 (1,859; 1,851 and 10,938 generations). However, in the cases where migration is more restricted during the glacial phases, the mean times are much higher than their corresponding values in Figure 2: 4,148, 8,684 and 499,521 generations in Figs. 4D – 4F against 2,616, 2,788 and 101,798 generations in Figs. 2D – 2F. This will be further discussed in the next section. The values of F_{ST} are 0.81 when $N_G m_G = 0.1$ (Figs. 5B & 5C) and 0.98 when $N_G m_G = 10^{-5}$ (Figs. 5E & 5F), which are very similar to the corresponding values in Figure 2 (0.81 and 0.97, respectively).

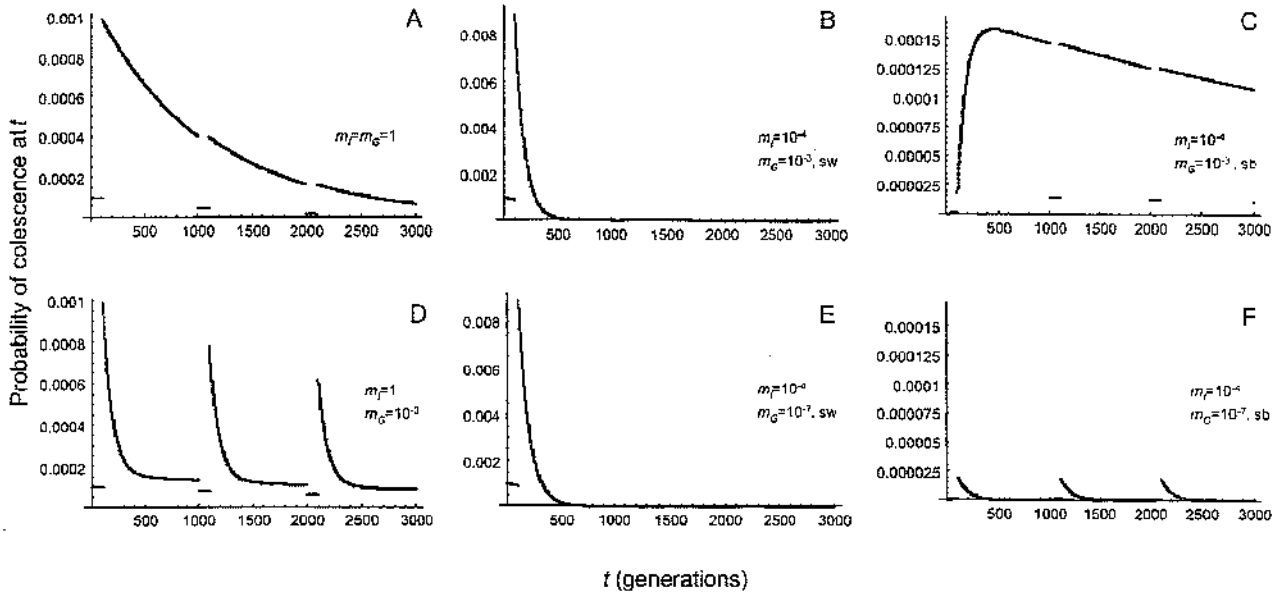

Figure 5. Probabilities of coalescence, with $t_G=9t_I$. Parameters: $N_I=1000$; $N_G=100$; $D_I=D_G=10$; $t_I=100$ and $t_G=900$; 'sw' and 'sb' stand for sampling within and sampling between demes, respectively. Migration rates are given above. (A) Both phases panmictic; (B) & (C) both phases structured ($N_Im_I= N_Gm_G=0.1$): (B) sampling within one deme; (C) between demes; (D) phases *I* panmictic and *G* structured ($N_Gm_G=0.1$); (E) & (F) both phases structured, with *G* much more structured ($N_Im_I=0.1$; $N_Gm_G=10^{-5}$): (E) sampling within one deme; (F) between demes.

Mean coalescence times

Species levels of genetic variation are directly correlated with coalescence times, since longer times entail more opportunities for mutations to occur. In order to understand how the different parameters of our model affect the mean coalescence times – and consequently levels of genetic variation – we plotted the mean times as functions of the glacial phase parameters (Fig. 6). We fixed the parameters of the interglacial phases for all cases at N_I = 1,000; D_I = 10; $m_I = 10^{-4}$ and t_I = 100. First we examine the effect of changing deme size (Fig. 6A). The behavior of the curve is straightforward: a smaller deme size in the glacial phases leads to a smaller mean coalescence time. This reduction occurs regardless of the sampling scheme, and it is robust to changes in the other parameters (not shown). The curves are not linear though, and the mean coalescence time is limited by the duration of the glacial phase. That is, as N_G tends to infinity, the contribution of each glacial phase tends to t_G, being limited by it. Unfortunately the actual function that generates the curve is too complex to allow further discussion of its behavior.

The second parameter studied was the migration rate. In both cases shown, the behavior of the curve for sampling between demes is the same: a more restricted migration leads to an increase in the mean coalescence times, until the curves reach a plateau (Fig. 6B – 6C). This is also true for cases with other parameter values (not shown). However, the behavior of the curve for sampling within demes depends on the number of demes: when $D_G < D_I$, restricting migration leads to a decrease of the mean coalescence times (Fig. 6B). On the other hand, when $D_G = D_I$ (not shown) or $D_G > D_I$ (Fig. 6C), the restriction in migration leads to an increase in the mean times. It is interesting to note that, in any case, the mean coalescence time for a pair of genes sampled within demes is not independent of migration rate, as it is in the standard island model (Strobeck 1987, Hey 1991).

The behavior of the mean times as function of the number of demes is somewhat more complex. First, when sampling is done 'within demes', an increase in the number of demes during the glacial phases leads to an increase of the mean coalescence times (Fig. 6D). This is also true for sampling 'between demes', when $D_G \leq D_I$. However, when $D_G >$

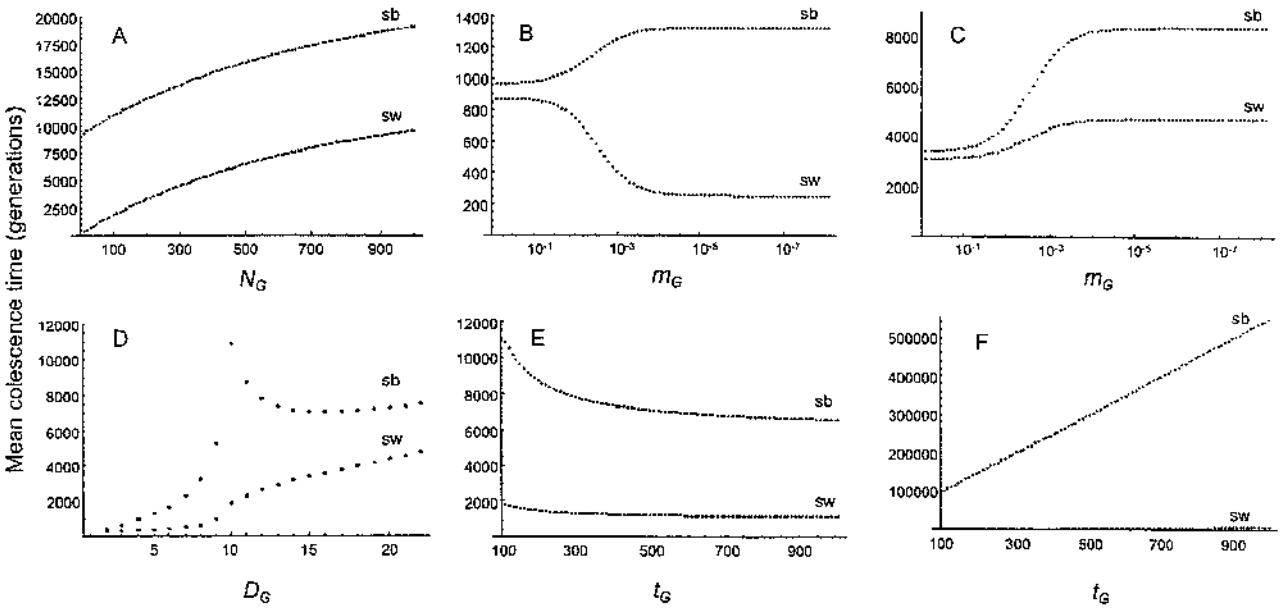

Figure 6. Mean coalescence times as functions of different parameters. In all cases N_I=1000; D_I=10; m_I=10^{-4} and t_I=100; 'sw' and 'sb' stand for sampling within and sampling between demes, respectively. (A) Mean time as function of N_G, with D_G=10, m_G=10^{-3}, t_G=100; (B) Mean time as function of m_G, with N_G=100, D_G=5, t_G=100; (C) Mean time as function of m_G, with N_G=100, D_G=20, t_G=100; (D) Mean time as function of D_G, with N_G=100, m_G=10^{-3}, t_G=100; (E) Mean time as function of t_G, with N_G=100, D_G=10, m_G=10^{-3}; (F) Mean time as function of t_G, with N_G=100, D_G=10, m_G=10^{-7}.

D_I there is a decrease in the mean coalescence times for a certain range of the plot ($10 \leq D_G \leq 16$ with these parameter values), with an increase after that (see Figure 6D). We observe that, with these parameters, the highest value for the mean coalescence time of two genes sampled 'between demes' occurs when $D_G = D_I$. Transition generations increase the probability that lineages will be brought into the same deme only when $D_G \neq D_I$ (see Appendix). Nevertheless, with some other combinations of parameters this effect disappears, and the mean coalescence times become strictly increasing as functions of D_G, independent of the range (not shown).

When the mean coalescence times are plotted as functions of the duration of the glacial phases, t_G, we observe two possible behaviors. The first one is a decrease of the mean times with an increasing t_G. This behavior can be seen in Fig. 6E ($D_G = D_I$ and $N_G\, m_G = N_I\, m_I = 0.1$). In this case, the level of population structure is equivalent for both glacial and interglacial phases, but the deme size is smaller in the glacial phases. Therefore, it is expected that increasing the time spent in each glacial phase will reduce the mean coalescence times. The second behavior is an increase of the mean times with an increasing t_G, as in Fig. 6F ($D_G = D_I$ and $N_G\, m_G = 10^{-5}$, with $N_I\, m_I = 0.1$). In this case, the greater level of population structure in the glacial phases outweighs the reduction in population size, thus leading to an increase in the mean times with an increasing t_G.

Discussion

We have studied a model of alternation between glacial and interglacial cycles, with changes in both population size and structure. One version of the model consisted of variation in population size only, being a model of repeated bottlenecks. We observed that this repeated reduction in size causes multiple peaks and discontinuities in the distribution of coalescence times. To the best of our knowledge, this is a new result However, it is a logical extension of the results of Austerlitz et al. (1997), who studied a model of a panmictic population suffering a single bottleneck and a model of two demes, where one deme grows and colonizes the second one. In both cases, they observed that even if the sample is taken hundreds of generations after the bottleneck or colonization event, the distribution of coalescence times can be multi-modal, with one peak

corresponding to the time of the event. In our model we show that multiple events, in this case bottlenecks, can produce multiple peaks in the distribution – corresponding to each of the bottlenecks. This suggests that a signal of multiple bottlenecks might be preserved in genetic data, as opposed to only one signal corresponding to the most recent event. In principle, it could be observed as multiple peaks in the distribution of the number of pairwise differences at a large number of independent loci.

We considered the effect of population structure on this pattern. We expected that higher levels of structure during glacial phases could oppose the effect of the reduction in population size, and cause lineages to persist through several cycles. This was generally the case. Increasing the level of population structure by restricting migration in the glacial phases affected the distribution of coalescence times, usually reducing the signal of repeated bottlenecks (*e.g.* Fig. 2F). Also, this increase in population structure caused an increase in the mean coalescence times in almost all cases studied, including several cases where the samples are taken within demes (*e.g.* Figs. 2E, 5E, 6C). We have also investigated the effect of changing the number of demes in the glacial phases. In general, we have observed that a greater number of demes leads to larger expected coalescence times (Fig. 6D), and would consequently lead to greater levels of genetic variation. These results extend the ones observed previously (Jesus et al. 2006, in press), in a simpler model, and help explain the long gene genealogies found in several species (*e.g.* Smith & Farrell 2005, Hewitt 2004, Paulo et al. 2001), where the most recent common ancestor predates the last glacial period.

The effect of the longer times during the glacial phases was also studied. We observed that the repeated peaks were still present in most cases, corresponding to the repeated demographic changes. However, there were two cases in which only the first peak persisted, corresponding to the most recent demographic change (Figs. 5E, 5F). The signal corresponding to earlier transitions was dramatically reduced. Generation time should also play an important part here. If there are too many generations, relative to population size, in one given phase, there will be a smaller probability that sampled genes persist through several phases without coalescing. One could ask when we should expect a signal of more than one event to be present in the distribution of coalescence times, as opposed to only one signal corresponding to the most recent event.

In a simple case, of one panmictic population suffering repeated bottlenecks, we can look at this explicitly. The probability that two genes sampled in the present have not coalesced by the end of the most recent glacial phase is given by

$$\mathrm{P(no\ coal.)} = \left(1-\frac{1}{N_I}\right)^{t_I}\left(1-\frac{1}{N_G}\right)^{t_G} \approx e^{-\left(\frac{t_I}{N_I}+\frac{t_G}{N_G}\right)}.$$

For this probability to be greater than, say, five percent, we would need that

$$\frac{t_I}{N_I}+\frac{t_G}{N_G} < -\ln(0.05).$$

This could be calculated that simply for other values, in this panmictic version of our model. But for cases where population structure is important, it is necessary to take into account the number of demes and the migration rate, using the cumulative distribution functions (first terms in Eq. 9 and Eq. 10). The same numerical approach used earlier could be applied for this calculation, and a routine can be obtained from the authors upon request.

Another important topic of interest is how the patterns observed here should be reflected in actual genetic data. The mean coalescence times obtained can be directly converted into expectations of levels of genetic variation if an infinite sites model (Kimura 1969), for instance, is assumed. The expected number of segregating sites is a simple function of the expected coalescence time and the mutation rate per sequence per generation

$$E[S] = 2\mu E[T_{coal}].$$

Obtaining the distribution of the number of segregating sites is more complex, and will be the subject of a future study. However, we expect that if mutation rates are high, the signals observed in the distribution of coalescence times should be preserved in the distribution of segregating sites for a sample of size two (n=2). This was observed by Austerlitz et al. (1997), in their two-deme colonization model. It is important to point out that the distribution of segregating sites for n=2 is not the usual "distribution of pairwise differences" or "mismatch distribution". Both in Aulsterlitz et al. (1997) and in the present case, the kind of data that would be necessary to search for possible signals would be multi-locus data, with a sample of size two. Further studies, possibly by simulation, might be able to make use of larger sample sizes, but certainly multi-locus data will be essential in any demographic inference of this sort.

Our model was designed to help clarify the genetic consequences of glacial cycles, but it may also be useful in other cases where there are cyclic demographic changes taking place. Many species suffer such changes in shorter time scales, for example, the seasonal bottlenecks suffered by insects in high latitude and temperate regions. The framework used here could be applied to help understand the genetic consequences of these seasonal changes as well.

Acknowledgements

We thank Aluísio Pinheiro for valuable suggestions regarding the routine for numerical calculations; Craig Moritz and other organizers and participants of the Workshop on Phylogeography and Phylogenetics held in 2005 at the Mathematical Biosciences Institute (MBI), Ohio State University, for exciting discussions and suggestions; Tee Muirhead and all members of the Wakeley Lab for helpful discussions and comments. JW was supported by a Career Award (DEB-0133760) from the National Science Foundation and FFJ received graduate fellowships from CAPES (PDEE) and FAPESP. This study was supported by FAPESP (04/09742-0).

References

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P (1997) Evolution of coalescence times, genetic diversity and structure during colonization. Theoretical Population Biology, 51, 148-164.

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. Genetics, 154, 1309-1321.

Burnham RJ, Graham A (1999) The history of neotropical vegetation: new developments and status. Annals of the Missouri Botanical Garden, 86, 546-589.

Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic structure on neutral variation. Annual Review of Ecology, Evolution, and Systematics, 34, 99-125.

Ewing G, Rodrigo A (2006) Coalescent-based estimation of population paraveters when the number of demes changes over time. Molecular Biology and Evolution, 23, 988-996.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Molecular Ecology, 13, 853-864.

Fisher RA (1930) The Genetical Theory of Natural Selection. Claredon, Oxford.

Galbreath KE, Cook JA (2004) Genetic consequences of Pleistocene glaciations for the tundra vole (Microtus oeconomus) in Beringia. Molecular Ecology, 13, 135-148.

Griffiths RC, Tavaré S (1994) Sampling theory for neutral alleles in a varying environment. Philosophical Transactions of the Royal Society of London Series B Biological Science, 344, 403-410.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature, 405, 907-913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London Series B Biological Science, 359, 183-195.

Hey J (1991) A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theoretical Population Biology, 39, 30-48.

Hudson RR (1983) Testing the constant-rate neutral allele model with protein sequence data. Evolution, 37, 203-217.

Hugall A, Moritz C, Moussalli A, Stanisic J (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail Gnarosophia bellendenkerensis (Brazier 1875). Proceedings of the National Academy of Sciences USA, 99, 6112-6117.

Jesus FF, Wilkins JF, Solferini VN, Wakeley J (2006) Expected coalescence times and segregating sites in a model of glacial cycles. Genetics and Molecular Research (in press).

Jouzel J (2003) Past climate (the last 400 ka): from geological times to future climate change. Comptes Rendus Geoscience, 335, 509-524.

Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics, 61, 893-903.

Kingman JFC (1982a) The coalescent. Stochastic Processes and their Applications, 13, 235-248.

Kingman JFC (1982b) On the genealogy of large populations. Journal of Applied Probability A, 19, 27-43.

Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. Science, 277, 1666-1669.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. Genetics, 149, 429-434.

Latter BDH (1973) The island model of population differentiation: a general solution. Genetics, 73, 147-157.

Lessa EP, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. Proceedings of the National Academy of Sciences USA, 100, 10331-10334.

Maruyama T (1970) Effective number of alleles in a subdivided population. Theoretical Population Biology, 1, 273-306.

Maruyama T, Fuerst PA (1984) Population bottlenecks and nonequilibrium models in population genetics I. Allele numbers when populations evolve from zero variability. Genetics, 108, 745-763.

Maruyama T, Fuerst PA (1985a) Population bottlenecks and nonequilibrium models in population genetics II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics, 111, 675-689.

Maruyama T, Fuerst PA (1985b) Population bottlenecks and nonequilibrium models in population genetics III. Genic homozygosity in populations which experience periodic bottlenecks. Genetics, 111, 691-703.

McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. Ecology, 86, 2088-2098.

Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: an integrated molecular approach. Annual Review of Ecology and Systematics, 31, 533-563.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution, 29, 1-10.

Nei M, Takahata N (1993) Effective population size, genetic diversity, and coalescence time in subdivided populations. Journal of Molecular Evolution, 37, 240-244.

Paulo OS, Dias C, Bruford MW, Jordan WC & Nichols RA (2001) The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard. Proceedings of the Royal Society of London Series B Biological Sciences, 268, 1625-1630.

Petit JR, Jouzel J, Raynaud D, Barkov et al. (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature, 399, 429-436.

Petit RJ, Aguinagalde I, de Beaulieu JL et al. (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. Science, 300, 1563-1565.

Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. Molecular Biology and Evolution, 20, 76-86.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution, 9, 552-569.

Slatkin M (1991) Inbreeding coefficients and coalescence times. Genetical Research, 58, 167-175.

Slatkin M & Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics, 129, 555-562.

Smith CI & Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles, Moneilema gigas and Moneilema armatum, in response to Pleistocene climate changes. Molecular Ecology, 14, 1025-1044.

Strobeck C (1987) Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics, 117, 149-153.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics, 105, 437-460.

Tajima F (1989a) DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-population model. Genetics, 123, 229-240.

Tajima F (1989b) The effect of change in population size on DNA polymorphism. Genetics, 123, 597-601.

Takahata N, Nei M (1984) F_{ST} and G_{ST} statistics in the finite island model. Genetics, 107,501-504.

Takahata N (1991) Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. Genetics, 129, 585-595.

Tzedakis PC, Andrieu V, deBeaulieu JL et al. (1997) Comparison of terrestrial and marine records of changing climate of the last 500,000 years. Earth and Planetary Science Letters, 150, 171-176.

Wakeley J (2004) Recent trends in population genetics: More data! More math! Simple models? Journal of Heredity, 95, 397-405.

Webb T, Bartlein PJ (1992) Global Changes During the Last 3 Million Years - Climatic Controls and Biotic Responses. Annual Review of Ecology and Systematics, 23, 141-173.

Wolfram Research, Inc. (2001) Mathematica, Version 4.1, Champaign, IL.

Wright S (1931) Evolution in Mendelian populations. Genetics, 16, 97-159.

Wright S (1938) Size of population and breeding structure in relation to evolution. Science, 87, 430-431.

Wright S (1951) The genetical structure of populations. Annals of Eugenics, 15, 323-354.

Appendix

Transition generations

Going backwards in time, a transition generation is composed first by a migration step, then a step that takes into account a possible change in number of demes, and then a reproduction step. Each step can be represented by a matrix, such that

$$\mathbf{T}_{IG} = \mathbf{M}_I \mathbf{C}_{IG} \mathbf{R}_G \tag{A1}$$

and

$$\mathbf{T}_{GI} = \mathbf{M}_G \mathbf{C}_{GI} \mathbf{R}_I . \tag{A2}$$

Each migration matrix $\mathbf{M}_I$ or $\mathbf{M}_G$ has the general form

$$\mathbf{M} = \begin{pmatrix} \alpha + \frac{1-\alpha}{D} & (1-\alpha)\left(1-\frac{1}{D}\right) & 0 \\ \frac{1-\alpha}{D} & 1-\frac{1-\alpha}{D} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{A3}$$

with the appropriate subscripts (I or G) for the parameters D and $\alpha=(1-m)^2$.

The reproduction matrices $\mathbf{R}_I$ and $\mathbf{R}_G$ have the general form

$$\mathbf{R}=\begin{pmatrix} 1-\frac{1}{N} & 0 & \frac{1}{N} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{A4}$$

with the appropriate subscript (I or G) for the parameter N.

The matrix for the change in number of demes for a transition from I to G – backwards in time – is

$$\mathbf{C}_{IG}=\begin{cases} \begin{pmatrix} 1 & 0 & 0 \\ (1-\beta_{IG})\left(\frac{1}{D_G}\right) & \beta_{IG}+(1-\beta_{IG})\left(1-\frac{1}{D_G}\right) & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{if } D_I > D_G \\ \begin{pmatrix} \frac{D_I}{D_G} & 1-\frac{D_I}{D_G} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{if } D_I \le D_G, \end{cases} \tag{A5}$$

where

$$\beta_{IG}=\frac{\binom{D_G}{2}}{\binom{D_I}{2}}=\frac{D_G(D_G-1)}{D_I(D_I-1)}.$$

The matrix for the change in number of demes for a transition from G to I – backwards in time – is

$$\mathbf{C}_{GI} = \begin{cases} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{if } D_I > D_G \\ \begin{pmatrix} 1 & 0 & 0 \\ \left(\dfrac{D_G - D_I}{D_I(D_G - 1)}\right) & 1 - \left(\dfrac{D_G - D_I}{D_I(D_G - 1)}\right) & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{if } D_I \leq D_G. \end{cases} \tag{A6}$$

Mean coalescence times and F_{ST}

The mean time to coalescence can be obtained by summing the contributions of each phase. Let X_I be the contribution of each type I phase, and X_G the contribution of each type G phase. Those depend on the state of the system at the immediately 'preceding' generation – backwards in time. If the two lineages have already coalesced, then the contribution to coalescence time is zero. If the system is on state (1), that is, both lineages in one deme but not coalesced ('within' demes), then the contribution to the expected coalescence time is

$$\varepsilon_{Iw} = E[X_I \mid w] = \sum_{k=0}^{t_I - 1} \left[\mathbf{1} - (\mathbf{P}_I)^k{}_{13}\right], \tag{A7}$$

for a type I phase, and

$$\varepsilon_{Gw} = E[X_G \mid w] = \sum_{k=0}^{t_G - 1} \left[\mathbf{1} - (\mathbf{P}_G)^k{}_{13}\right], \tag{A8}$$

for a type G phase.

And if the system is on state (2), that is, with the lineages in two separate demes ('between' demes), the contributions of type *I* and type *G* phases, are, respectively,

$$\varepsilon_{Ib} = E[X_I \mid b] = \sum_{k=0}^{t_I - 1} \left[1 - (\mathbf{P}_I)^k{}_{23} \right], \tag{A9}$$

$$\varepsilon_{Gb} = E[X_G \mid b] = \sum_{k=0}^{t_G - 1} \left[1 - (\mathbf{P}_G)^k{}_{23} \right]. \tag{A10}$$

In order to get to the equations above we use the fact that, if the two lineages have not coalesced by a certain generation, the contribution of the 'following' generation – backwards in time – to the mean coalescence time is equal to one. Therefore we only need to calculate the probability that the lineages have not coalesced by a given generation, conditioned on the initial state of the phase. The sum of these probabilities is the contribution of the whole phase to the mean coalescence time.

If the sample is originally taken from one deme on a type *I* phase, then its expected coalescence time is

$$\begin{aligned} E[T_w] = {} & \varepsilon_{Iw} + (\Pi_I)_{11}\varepsilon_{Gw} + (\Pi_I)_{12}\varepsilon_{Gb} + \\ & + (\Pi^1)_{11}\varepsilon_{Iw} + (\Pi^1)_{11}(\Pi_I)_{11}\varepsilon_{Gw} + (\Pi^1)_{11}(\Pi_I)_{12}\varepsilon_{Gb} + \\ & + (\Pi^1)_{12}\varepsilon_{Ib} + (\Pi^1)_{12}(\Pi_I)_{21}\varepsilon_{Gw} + (\Pi^1)_{12}(\Pi_I)_{22}\varepsilon_{Gb} + \\ & + (\Pi^2)_{11}\varepsilon_{Iw} + (\Pi^2)_{11}(\Pi_I)_{11}\varepsilon_{Gw} + (\Pi^2)_{11}(\Pi_I)_{12}\varepsilon_{Gb} + \\ & + (\Pi^2)_{12}\varepsilon_{Ib} + (\Pi^2)_{12}(\Pi_I)_{21}\varepsilon_{Gw} + (\Pi^2)_{12}(\Pi_I)_{22}\varepsilon_{Gb} \cdots \end{aligned} \tag{A11}$$

Simplifying, we obtain

$$E[T_w]=\left[\sum_{k=0}^{\infty}\left(\Pi^k\right)_{11}\right]\left[\varepsilon_{Iw}+\left(\Pi_I\right)_{11}\varepsilon_{Gw}+\left(\Pi_I\right)_{12}\varepsilon_{Gb}\right]+$$

$$\left[\sum_{k=0}^{\infty}\left(\Pi^k\right)_{12}\right]\left[\varepsilon_{Ib}+\left(\Pi_I\right)_{21}\varepsilon_{Gw}+\left(\Pi_I\right)_{22}\varepsilon_{Gb}\right]; \qquad \text{(A12)}$$

Similarly, if the sample is obtained from two separate demes, the mean coalescence time is

$$E[T_b]=\left[\sum_{k=0}^{\infty}\left(\Pi^k\right)_{21}\right]\left[\varepsilon_{Iw}+\left(\Pi_I\right)_{11}\varepsilon_{Gw}+\left(\Pi_I\right)_{12}\varepsilon_{Gb}\right]+$$

$$\left[\sum_{k=0}^{\infty}\left(\Pi^k\right)_{22}\right]\left[\varepsilon_{Ib}+\left(\Pi_I\right)_{21}\varepsilon_{Gw}+\left(\Pi_I\right)_{22}\varepsilon_{Gb}\right], \qquad \text{(A13)}$$

The infinite sums present above in A12 and A13 can be expressed in terms of the eigenvalues and eigenvectors of Π. They then become reduced to geometric series of form $\sum_{i=0}^{\infty}a^i$, which converge to $1/(1-a)$ when $|a|<1$. We first use the fact that Π^k can be written in spectral form as

$$\Pi^k=\lambda_0^k\mathbf{r}_0\mathbf{l}_0'+\lambda_1^k\mathbf{r}_1\mathbf{l}_1'+\lambda_2^k\mathbf{r}_2\mathbf{l}_2',$$

where the λ_i's, $\mathbf{r}_i$'s e $\mathbf{l}_i$'s are the eigenvalues and right and left eigenvectors of Π, respectively, normalized so that $\mathbf{l}_i'\mathbf{r}_i=1$. For the first sum in A12, for instance, we have

$$\sum_{k=0}^{\infty}\left(\Pi^{k}\right)_{11}=\sum_{k=0}^{\infty}\left[\sum_{i=1}^{2}\lambda_{i}^{k}\left(\mathbf{r}_{i}\mathbf{l}_{i}^{'}\right)_{11}\right]=\sum_{i=1}^{2}\left[\sum_{k=0}^{\infty}\lambda_{i}^{k}\left(\mathbf{r}_{i}\mathbf{l}_{i}^{'}\right)_{11}\right]=\sum_{i=1}^{2}\left[\left(\mathbf{r}_{i}\mathbf{l}_{i}^{'}\right)_{11}\sum_{k=0}^{\infty}\lambda_{i}^{k}\right]=$$

$$=\sum_{i=1}^{2}\left[\left(\mathbf{r}_{i}\mathbf{l}_{i}^{'}\right)_{11}\left(\frac{1}{1-\lambda_{i}}\right)\right], \tag{A14}$$

since both λ_1 and λ_2 have absolute values smaller than one. The first eigenvalue, λ_0 has value 1, but the associated matrix $\mathbf{r}_0\mathbf{l}_0^{'}$ is

$$\mathbf{r}_{0}\mathbf{l}_{0}^{'}=\begin{pmatrix}0 & 0 & 1\\ 0 & 0 & 1\\ 0 & 0 & 1\end{pmatrix}, \tag{A15}$$

and so all the entries that we are interested in (see A12 & A13) have value zero, and do not contribute to any of the sums.

A routine to calculate the mean coalescence times was written using Mathematica (Wolfram Research 2001) and is available from the authors upon request.

We calculate F_{ST} as a function of the mean coalescence times, as given by Slatkin (1991, Eq. 8) assuming small mutation rates. In our notation,

$$F_{ST}=\frac{E[T_m]-E[T_w]}{E[T_m]}, \tag{A16}$$

where $E[T_w]$ is equal to the mean coalescence time for two lineages sampled from the same deme and $E[T_m]$ is the mean coalescence time for two lineages sampled from the set of all demes, given by

$$E[T_m]=\frac{1}{D_I}E[T_w]+\frac{D_I-1}{D_I}E[T_b]. \tag{A17}$$

CONCLUSÕES GERAIS

CONCLUSÕES GERAIS

A partir do estudo dos dois modelos teóricos desenvolvidos neste trabalho, algumas inferências podem ser feitas:

- A redução de tamanho populacional durante os períodos glaciais leva a uma redução dos tempos médios de coalescência. Já o aumento de subdivisão, em geral, tem efeito contrário.

- Há indícios de que o número de sub-populações é importante na manutenção da variabilidade genética durante os períodos de redução populacional, mesmo quando cada sub-população é muito pequena.

- A estruturação é capaz de superar o efeito da redução populacional durante os períodos glaciais.

- As distribuições de tempos de coalescência com alternância de parâmetros populacionais são muito diferentes da distribuição esperada para uma população panmítica de tamanho constante ("standard coalescent").

- A alternância de tamanhos populacionais produz descontinuidades e picos múltiplos na distribuição dos tempos de coalescência.

- O aumento da estrutura populacional concomitante à redução populacional, em geral, atenua os picos da distribuição dos tempos de coalescência.

- Há indícios de que múltiplos gargalos podem deixar sinais em dados genéticos. Este sinal seria observado como múltiplos picos na distribuição de diferenças par-a-par, obtida a partir de um grande número de *loci* independentes.

- Há indícios de que a estrutura populacional gerada por subdivisão e migração tem um papel importante na manutenção da variabilidade genética.

- As longas genealogias gênicas observadas em muitas espécies animais e vegetais poderiam ser explicadas por uma grande subdivisão populacional durante os períodos glaciais.

Referências Bibliográficas

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P (1997) Evolution of coalescence times, genetic diversity and structure during colonization. Theoretical Population Biology, 51, 148-164.

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. Genetics, 154, 1309-1321.

Burnham RJ, Graham A (1999) The history of neotropical vegetation: new developments and status. Annals of the Missouri Botanical Garden, 86, 546-589.

Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic structure on neutral variation. Annual Review of Ecology, Evolution, and Systematics, 34, 99-125.

Ewing G, Rodrigo A (2006) Coalescent-based estimation of population paraveters when the number of demes changes over time. Molecular Biology and Evolution, 23, 988-996.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Molecular Ecology, 13, 853-864.

Fisher RA (1930) The Genetical Theory of Natural Selection. Clarendon, Oxford.

Galbreath KE, Cook JA (2004) Genetic consequences of Pleistocene glaciations for the tundra vole (Microtus oeconomus) in Beringia. Molecular Ecology, 13, 135-148.

Griffiths RC, Tavaré S (1994) Sampling theory for neutral alleles in a varying environment. Philosophical Transactions of the Royal Society of London Series B Biological Science, 344, 403-410.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature, 405, 907-913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London Series B Biological Science, 359, 183-195.

Hey J (1991) A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theoretical Population Biology, 39, 30-48.

Hey J, Machado CA (2003) The study of structured populations - New hope for a difficult and divided science. Nature Reviews Genetics, 4, 535-543.

Hudson RR (1983) Testing the constant-rate neutral allele model with protein sequence data. Evolution, 37, 203-217.

Hudson RR (1990). Gene genealogies and the coalescent process. In: Oxford Surveys in Evolutionary Biology Volume 7 (Futuyma D and Antonovics J, eds.). Oxford University Press, Oxford, pp. 1-44.

Hugall A, Moritz C, Moussalli A, Stanisic J (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail Gnarosophia bellendenkerensis (Brazier 1875). Proceedings of the National Academy of Sciences USA, 99, 6112-6117.

Jouzel J (2003) Past climate (the last 400 ka): from geological times to future climate change. Comptes Rendus Geoscience, 335, 509-524.

Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics, 61, 893-903.

Kingman JFC (1982a) The coalescent. Stochastic Processes and their Applications, 13, 235-248.

Kingman JFC (1982b) On the genealogy of large populations. Journal of Applied Probability A, 19, 27-43.

Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. Science, 277, 1666-1669.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. Genetics, 149, 429-434.

Latter BDH (1973) The island model of population differentiation: a general solution. Genetics, 73, 147-157.

Lessa EP, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. Proceedings of the National Academy of Sciences USA, 100, 10331-10334.

Maruyama T (1970) Effective number of alleles in a subdivided population. Theoretical Population Biology, 1, 273-306.

Maruyama T, Fuerst PA (1984) Population bottlenecks and nonequilibrium models in population genetics I. Allele numbers when populations evolve from zero variability. Genetics, 108, 745-763.

Maruyama T, Fuerst PA (1985a) Population bottlenecks and nonequilibrium models in population genetics II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics, 111, 675-689.

Maruyama T, Fuerst PA (1985b) Population bottlenecks and nonequilibrium models in population genetics III. Genic homozygosity in populations which experience periodic bottlenecks. Genetics, 111, 691-703.

McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. Ecology, 86, 2088-2098.

Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: an integrated molecular approach. Annual Review of Ecology and Systematics, 31, 533-563.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution, 29, 1-10.

Nei M, Takahata N (1993) Effective population size, genetic diversity, and coalescence time in subdivided populations. Journal of Molecular Evolution, 37, 240-244.

Paulo OS, Dias C, Bruford MW, Jordan WC & Nichols RA (2001) The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence

from the phylogeography of an Iberian lizard. Proceedings of the Royal Society of London Series B Biological Sciences, 268, 1625-1630.

Petit JR, Jouzel J, Raynaud D, Barkov et al. (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature, 399, 429-436.

Petit RJ, Aguinagalde I, de Beaulieu JL et al. (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. Science, 300, 1563-1565.

Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. Molecular Biology and Evolution, 20, 76-86.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution, 9, 552-569.

Slatkin M (1991) Inbreeding coefficients and coalescence times. Genetical Research, 58, 167-175.

Slatkin M & Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics, 129, 555-562.

Smith CI & Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles, Moneilema gigas and Moneilema armatum, in response to Pleistocene climate changes. Molecular Ecology, 14, 1025-1044.

Strobeck C (1987) Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics, 117, 149-153.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics, 105, 437-460.

Tajima F (1989a) DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-population model. Genetics, 123, 229-240.

Tajima F (1989b) The effect of change in population size on DNA polymorphism. Genetics, 123, 597-601.

Takahata N (1991) Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. Genetics, 129, 585-595.

Takahata N, Nei M (1984) F_{ST} and G_{ST} statistics in the finite island model. Genetics, 107,501-504.

Tzedakis PC, Andrieu V, deBeaulieu JL et al. (1997) Comparison of terrestrial and marine records of changing climate of the last 500,000 years. Earth and Planetary Science Letters, 150, 171-176.

Wakeley J (1998) Segregating sites in Wright's island model. Theoretical Population Biology, 53, 166-174.

Wakeley J (1999) Nonequilibrium migration in human history. Genetics, 153, 1863-1871.

Wakeley J (2001) The coalescent in an island model of population subdivision with variation among demes. Theoretical Population Biology, 59, 133-144.

Wakeley J (2004) Recent trends in population genetics: More data! More math! Simple models? Journal of Heredity, 95, 397-405.

Webb T, Bartlein PJ (1992) Global Changes During the Last 3 Million Years - Climatic Controls and Biotic Responses. Annual Review of Ecology and Systematics, 23, 141-173.

Wolfram Research, Inc. (2001) Mathematica, Version 4.1, Champaign, IL.

Wright S (1931) Evolution in Mendelian populations. Genetics, 16, 97-159.

Wright S (1938) Size of population and breeding structure in relation to evolution. Science, 87, 430-431.

Wright S (1951) The genetical structure of populations. Annals of Eugenics, 15, 323-354.