Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Programa de Pós-Graduação em Genética e Biologia Molecular

Origem do *Homo sapiens* e sua chegada às Américas: uma contribuição da antropologia molecular

Nelson Jurandi Rosa Fagundes

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências

Orientador: Dr. Francisco Mauro Salzano
Co-orientador: Dr. Sandro Luis Bonatto
Dr. Laurent Excoffier

Porto Alegre
Maio de 2007

Universidade Federal do Rio Grande do Sul

Instituto de Biociências

Programa de Pós-Graduação em Genética e Biologia Molecular

Origem do *Homo sapiens* e sua chegada às Américas: uma contribuição da antropologia molecular

Nelson Jurandi Rosa Fagundes

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências

Orientador: Prof. Dr. Francisco Mauro Salzano

Co-orientadores: Prof. Dr. Sandro Luis Bonatto

Participação Especial: Prof. Dr. Laurent Excoffier

Porto Alegre

Maio de 2007

Este trabalho foi realizado nas instalações do Centro de Biologia Genômica e Molecular da Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) e no Laboratório de Genética de Populações Molecular e Computacional (CMPG) do Instituto de Zoologia da Universidade de Berna, subvencionado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela Fundação Nacional Suíça de Ciências (SNSF) e pela Universidade de Berna (UniBe).

For whom, it suddenly occurred to him to wonder, was he writing this diary? For the future, for the unborn. His mind hovered for a moment round the doubtful date on the page, and then fetched up with a bump against the Newspeak word doublethink. For the first time, the magnitude of what he had undertaken came home to him. How could you communicate with the future? It was of its nature impossible. Either the future would resemble the present in which case it would not listen to him, or it would be different from it, and his predicaments would be meaningless.

...

Suddenly, he began writing in sheer panic, only imperfectly aware of what he was setting down. His small but childish handwriting straggled up and down the page, shedding first its capital letters and finally even its full stops.

George Orwell, *1984*

AGRADECIMENTOS

Ao Dr. Francisco M. Salzano, por ter me aceitado como seu aluno, pelo constante incentivo, pela confiança em mim depositada, por partilhar comigo um pouquinho de sua erudição infindável e por ser um cientista exemplar.

Ao Dr. Sandro L. Bonatto, mestre e amigo, por ter sido sempre uma fonte de renovação do meu entusiasmo científico.

Ao Dr. Laurent Excoffier, por aceitar me receber em seu laboratório, fazer o possível para que eu me sentisse à vontade no laboratório e contribuir enormemente para minha formação como cientista.

A este Programa de Pós-Graduação e seus professores que me proporcionaram um ambiente propício onde eu pudesse amadurecer como cientista.

Ao CNPq e a CAPES, por terem me concedido as bolsas de estudo com as quais pude realizar o doutorado.

Aos demais professores do Centro de Biologia Genômica e Molecular, Dr. Maurício R. Bogo e Dr. Eduardo Eizirik, por todo o incentivo, pelas oportunidades de colaboração e pelo grande ambiente de trabalho que me proporcionaram.

Ao Dr. Jomar P. Laurino, pela amizade, apoio constante e oportunidades de colaboração.

Ao Elmo J. A. Cardoso e Ellen Mezzeck, pelo trabalho competente e sua eterna disposição para resolver problemas de última hora.

À Laci Krupahtz, pela infindável disposição em ajudar.

À Rita Schneider, pela ajuda local ou transoceânica sempre eficiente.

Aos colegas do PPGBM com quem durante esses anos pude aprender muito.

Ao Dr. Nicolas Ray, com quem pude aprender muitíssimo, pela amizade, por sua eterna disposição em discutir modelos evolutivos e cenários demográficos e idéias mirabolantes.

Aos demais colegas no CMPG pela acolhida calorosa e por terem me proporcionado um ambiente agradabilíssimo e cientificamente enriquecedor.

À Cladi, pela amizade, companheirismo e sua disposição infindável em ajudar.

À Jaque, pela grande amizade e companheirismo em mais de 10 anos nos trabalhos com Nativos Americanos.

Ao André, Felipe, Ricardo e Ronaldo pela imensa amizade e constante troca de idéias.

À Alice, minha "pupila", e a todos os "ex-pupilos" por me ensinarem muito mais do que imaginam.

A todos aqueles que passaram pelo mitogenoma: "Pessoas que construíram uma lenda..."

A todos os demais colegas, ex-colegas e amigos do Centro de Biologia Genômica e Molecular pela convivência prazerosa e por terem de alguma forma contribuído para minha formação.

Aos meus amigos em geral, que me ajudam a respirar.

À minha avó, Jandyra, e ao Ronaldo, por todo o carinho e incentivo que puderam me dar em parte dessa caminhada.

À minha mãe por todo o apoio há quase 30 anos

À Daisy, simplesmente por existir.

SUMÁRIO

Resumo 7
Abstract 9
Capítulo I: Introdução Geral..... 10
I.1. Origem do *Homo sapiens*..... 11
I.1.1. Fósseis e modelos evolutivos..... 11
I.1.2. Estudos iniciais em Genética – o triunfo do modelo de substituição?..... 13
I.1.3. Seqüenciamento de genes autossômicos: o multirregionalismo de volta à tona..... 14
I.1.4. Estudos recentes de dados genéticos..... 15
I.2. O Povoamento das Américas..... 17
I.2.1. Evidências antropológicas..... 17
I.2.2. Marcadores genéticos: número e idade das ondas migratórias..... 19
I.2.3. Marcadores genéticos: tamanho da população fundadora..... 20
I.3. A escolha dos marcadores genéticos estudados no presente trabalho..... 21
I.4. Objetivos..... 22
Capítulo II: Statistical Evaluation of Alternative Models of Human Evolution …….. 23
Capítulo III: Mitochondrial Genomics and the Peopling of the Americas…… 74
Capítulo IV: Discussão Geral……..…….. 108
Capítulo V: Referências Bibliográficas……. 115
Capítulo VI: Anexos…….. 127
Anexo I: Worldwide genetic variation at the 3'-UTR region of the LDLR gene: possible influence of natural selection…… 129
Anexo II: *Alu* insertion polymorphisms in Native Americans and related Asian populations…….. 142
Anexo III: Mitochondrial DNA and *Alu* insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay……... 162

RESUMO

Desde o final dos anos 80, o estudo de marcadores de DNA tem contribuído enormemente para um melhor entendimento de questões antropológicas. Atualmente, duas questões têm sido debatidas fortemente: a primeira envolve o modelo de origem dos humanos modernos e a possível assimilação de linhagens arcaicas pelas populações de *Homo sapiens* saídas da África. A segunda envolve o tempo e o modo do povoamento das Américas, o último continente a ser colonizado pelos humanos.

Foi realizado um estudo envolvendo o seqüenciamento de 50 locos autossômicos em uma amostra de 12 indígenas americanos, totalizando cerca de 50.000 pares de bases de seqüência por indivíduo. Os dados foram analisados juntamente com seqüências já publicadas oriundas de indivíduos africanos e asiáticos. Usando computação bayesiana aproximada e simulações de coalescência, foi estimada diretamente, pela primeira vez, a probabilidade relativa de três modelos de evolução humana. Um modelo de origem africana com substituição obteve a maior probabilidade relativa (98%), muito superior àquela de modelos de assimilação ou de evolução multirregional. O cenário evolutivo sugerido pode explicar não apenas a ancestralidade recente do DNA mitocondrial e do cromossomo Y, como também a existência de linhagens antigas em locos autossômicos. Quanto ao povoamento das Américas, os dados indicaram a ausência de um efeito de gargalo-de-garrafa forte e sugerem um tempo recente de povoamento, de até 13.350 anos atrás.

Para obter um cenário mais refinado para o povoamento das Américas, foram seqüenciados 53 genomas mitocondriais completos de indivíduos nativos americanos pertencentes aos cinco haplogrupos mitocondriais principais (A-D, X), que foram analisados juntamente com 191 genomas disponíveis em bancos de dados públicos. Os resultados indicaram um povoamento após o último máximo glacial há aproximadamente 18.000 anos atrás, associados a uma forte expansão populacional (100 vezes) com uma entrada no continente pela costa do Pacífico.

Os dois conjuntos de dados sugerem fortemente a ausência de um efeito fundador marcante durante a colonização das Américas e o povoamento do continente após o último máximo glacial. As discrepâncias entre as duas datas obtidas para a entrada nas Américas podem indicar os efeitos de uma segunda migração, mas a explicação mais simples

envolveria fatores como desenho amostral e incertezas vinculadas a alguns dos parâmetros utilizados.

ABSTRACT

Since the end of the 80s, the study of DNA markers has contributed enormously to a better understanding of anthropological questions. Currently, two issues have been hotly debated: the first involves a model for the origin of modern humans, and the possible assimilation of archaic lineages by *Homo sapiens* populations which had left Africa. The second involves the time and mode of the peopling of the Americas, the last continent colonized by humans.

A study involving the sequencing of 50 autosomal loci was performed in a sample of 12 American Indians, totaling about 50,000 base pairs per individual. These data were analyzed in conjunction with previously published sequences from African and Asian individuals. Using approximate Bayesian computation and coalescent simulations, the relative probability of three models of human evolution was assessed for the first time. A model of African origin with replacement received the highest relative probability (98%), much higher than that associated to assimilation or multiregional models. The favored evolutionary scenario can explain not only the recent ancestry for mitochondrial DNA and the Y chromosome, but also the existence of deep lineages in autosomal loci. Regarding the peopling of the Americas, the data indicate the lack of a bottleneck effect and suggest a recent time for its settlement, up to 13,350 years ago.

To obtain a more accurate scenario for the peopling of the Americas, 53 mitochondrial genomes from Native American individuals belonging to the five main mitochondrial haplogroups (A-D, X) were sequenced and were analyzed together with 191 genomes available in public databases. The analysis indicated a peopling of the continent after the last glacial maximum at approximately 18,000 years ago, associated to a strong population expansion (100 fold), and with an entry in the continent through the Pacific coast.

Both datasets strongly suggest the nonexistence of a marked founder effect during the settlement of the Americas and the peopling of the continent after the last glacial maximum. The discrepancies between the two calculated times for the entry in the continent could reflect the effects of a secondary migration, but the simplest explanation would be related to factors such as sampling design and to uncertainties intrinsic to some of the parameters.

CAPÍTULO I

Introdução Geral

I. INTRODUÇÃO GERAL

I.1. Origem do *Homo sapiens*

I.1.1. Fósseis e modelos evolutivos

Sem exageros, pode-se afirmar que a origem de nossa espécie é uma questão que nos tem fascinado desde que existimos, tendo ocupado uma posição central no desenvolvimento de sistemas filosóficos e religiosos. Do ponto de vista científico, o tema vem ocupando paleoantropólogos a partir do momento em que os primeiros fósseis de hominídeos foram encontrados (uma revisão histórica acerca do impacto da descoberta inicial desses fósseis pode ser encontrada em Lewin, 1998). Atualmente, supõe-se que os primeiros *Homo sapiens* anatomicamente modernos teriam surgido na África há cerca de 195.000 anos (McDougall *et al.*, 2005).

O gradual acúmulo de crânios de hominídeos fósseis permitiu que fossem elaborados os primeiros cenários sobre a origem do homem. Durante a década de 40, Franz Weidenreich, reconhecendo continuidades morfológicas regionais nas coleções de fósseis que analisara, sugeriu o surgimento do *H. sapiens* independentemente na África, Ásia e Europa a partir das populações regionais de hominídeos arcaicos, no que ficou conhecido como o "modelo do candelabro" (Weidenreich, 1946). Neste modelo, que não implicava em fluxo gênico significativo entre os grandes grupos continentais humanos, esses teriam uma história longa de isolamento; desde cerca de 1 milhão de anos atrás. Embora a formulação original de Weidenreich não tivesse conotações racistas aparentes (Lewin, 1998), a idéia de que os grandes grupos humanos pudessem ter histórias independentes bastante profundas gerou hipóteses nas quais esses grupos teriam atingido o "nível *H. sapiens*" em diferentes pontos no tempo (Coon, 1962), de modo que o modelo do candelabro acabou sendo adaptado para justificar e exacerbar as diferenças entre os grandes grupos humanos.

Num outro extremo estava o cenário defendido por Louis S. B. Leakey, elaborado na década de 60 a partir das descobertas de fósseis na África sub-Saariana feitos por seu grupo (p. ex. Leakey, 1966, Leakey e Goodall, 1969). Leakey sugeria uma origem Africana para o *H. sapiens* há cerca de 130.000 anos e via o *H. erectus* asiático como uma linhagem que teria sido extinta sem contribuir para a formação de nossa espécie. Esta hipótese formou a base do modelo de origem africana recente, ou modelo de substituição

africana, defendido por diversos pesquisadores atualmente (p. ex. Lahr e Foley, 1998; Stringer, 2002; Mellars, 2005)

A descoberta de novos fósseis e a re-interpretação dos fósseis pré-existentes, porém, fez surgir novas hipóteses acerca do surgimento dos humanos modernos. Thorne e Wolpoff (1992), por exemplo, adaptaram o modelo do candelabro sugerindo um intenso fluxo gênico entre os grandes grupos continentais desde as primeiras incursões de hominídeos fora da África (cerca de 2 milhões de anos atrás), de modo a haver uma conexão contínua entre os grupos humanos que mediava as modificações morfológicas vistas no registro fóssil, no que se tornou conhecido como modelo de evolução multirregional. Além disso, uma nova classe de modelos, intermediários entre os dois cenários extremos apresentados acima, começava a surgir. Alguns proponentes de uma origem única africana passaram a sugerir que embora a maioria das populações arcaicas houvesse desaparecido por substituição, parte delas deveria ter se hibridizado com os humanos modernos (Bräuer, 1992; Zilhão, 2006). Em consonância com este modelo, possíveis híbridos entre Neandertais (*H. neanderthalensis*) e humanos modernos teriam sido descobertos em Portugal e na Romênia (Duarte *et al.*, 1999; Rougier *et al.*, 2007), embora haja controvérsias quanto à interpretação desses esqueletos (Tattersall e Schwartz, 1999; Stringer, 2002). Possíveis híbridos entre hominídeos arcaicos e modernos são igualmente previstos por uma variação do modelo de hibridização, conhecida como modelo de assimilação (Smith, 1992; Trinkaus, 2007). Nela, as populações de humanos modernos que saíram da África entre 50.000 e 100.000 anos atrás assimilaram linhagens gênicas e características morfológicas das populações arcaicas locais através de extenso fluxo gênico. É interessante ressaltar que alguns proponentes do modelo de evolução multirregional parecem estar se movendo em direção a modelos de assimilação (Hawks *et al.*, 2000; Wolpoff *et al.*, 2001).

Embora tanto Aiello (1993) quanto Lewin (1998) e Stringer (2002) reconheçam a existência de quatro grandes modelos de evolução humana em debate atualmente (origem africana com substituição; origem africana com substituição e hibridização, assimilação e evolução multirregional), cabe ressaltar que as diferenças entre os modelos de hibridização e assimilação podem ser sutis. Ambos pressupõem (diferentemente do modelo multirregional) uma saída significativa de populações da África após o surgimento do conjunto de características responsáveis pela "modernidade" do *H. sapiens*, e hibridização

entre modernos e populações "arcaicas" regionais, tipicamente representadas pelos neandertais na Europa e o *H. erectus* na Ásia. A grande diferença entre eles, portanto, relaciona-se mais ao grau e duração da hibridização. Enquanto o modelo de substituição com hibridização pressupõe um grau pequeno de hibridização com rápida transição entre arcaísmo e modernismo, o modelo de assimilação pressupõe um processo demorado, de transição gradual, no qual foi facilitada a assimilação de caracteres arcaicos. Não deixa de ser curioso a semelhança entre tais modelos, se lembrarmos que ambos derivaram de modelos opostos (multirregional *vs.* origem africana com substituição).

I.1.2. Estudos iniciais em Genética – o triunfo do modelo de substituição?

Os primeiros marcadores genéticos usados para o estudo da diversidade humana foram os polimorfismos protéicos. Consistentemente, estudos utilizando esses marcadores revelaram que a maior parte da variabilidade genética é encontrada dentro dos grandes grupos continentais humanos, e não entre eles (p. ex. Nei e Roychoudhury, 1982; Cavalli-Sforza *et al.*, 1994), sugerindo que a sua separação é relativamente pequena, favorecendo o modelo de substituição. Além disso, árvores populacionais utilizando freqüências alélicas revelaram que a maior distância separava populações africanas das não-africanas, novamente em maior consonância com as expectativas do modelo de origem africana e substituição (p. ex. Cavalli-Sforza *et al.*, 1994).

A partir do final da década de 80 começaram a ser utilizados marcadores de DNA. Wainscoat *et al.* (1986) replicaram a maior distância entre africanos e não-africanos através de polimorfismos no tamanho de fragmentos de restrição (RFLPs) próximos ao gene da β-globina. No ano seguinte, Cann *et al.* (1987), estudando RFLPs no DNA mitocondrial (mtDNA) sugeriram que a coalescência desse loco ocorrera mais provavelmente na África há 200.000 anos atrás, e que a diversidade genética nesse continente era maior do que a encontrada em populações não-africanas. Os achados iniciais desses estudos pioneiros foram então replicados em alguma medida por outros marcadores de DNA, como a região hipervariável do mtDNA (Vigilant *et al.*, 1991), diversos RFLPs espalhados ao longo do genoma humano (Barbujani *et al.*, 1997), diversos locos de microssatélites (STRs) (Bowcock *et al.*, 1994; Jorde *et al.*, 1995; Barbujani *et al.*, 1997; Jorde *et al.*, 1997), inserções *Alu* (Batzer *et al.*, 1994), além da contrapartida biológica do mtDNA, o cromossomo Y (Hammer *et al.*, 1995). Finalmente, a publicação da seqüência

da região controladora do mtDNA de vários neandertais sugeriu fortemente que estes últimos não devem ter contribuído para o conjunto gênico dos humanos modernos (Krings *et al.*, 1997; 2000; Ovchinnikov *et al.*, 2000).

Ao final da primeira década de estudos evolutivos humanos com marcadores de DNA, tinha-se a sensação de que praticamente todos eles sugeriam fortemente um cenário de origem recente africana seguida de uma substituição total das populações arcaicas.

I.1.3. Seqüenciamento de genes autossômicos: o multirregionalismo de volta à tona

No final da década de 90, estudos baseados no seqüenciamento de locos nucleares passaram a ser amplamente utilizados, inicialmente sob o formato de estudos de um único gene ou região gênica e análise de estatísticas populacionais e da árvore gênica resultante. O primeiro desses estudos a ter grande impacto científico foi o de Harding *et al.* (1997), que seqüenciaram aproximadamente 3kb do gene da β-globina. Esses autores surpreenderam-se por não replicar os resultados obtidos por marcadores genéticos uniparentais. O estudo apontava maior variabilidade genética na África, mas isso foi interpretado em favor de um maior tamanho populacional efetivo nesse continente, e não de uma maior antiguidade. Além disso, os autores sugeriram que o fato da genealogia dos haplótipos presentes na Ásia ser de aproximadamente 200.000 anos não era facilmente compatível com as predições do modelo de origem africana e substituição.

A esse estudo, seguiram-se outros relatos de genes cujo padrão de variação supostamente não poderia ser explicado facilmente pelo modelo de origem africana com substituição (p. ex. Zhao *et al.*, 2000; Yu *et al.*, 2001), muito embora a maioria dos estudos de variação em seqüências nucleares apontasse uma melhor concordância com esse modelo (p. ex. Kaessmann *et al.*, 1999; Zhao *et al.*, 2000; Yu *et al.*, 2001). Excoffier (2002) sugeriu que os padrões discordantes apresentados pelos marcadores uniparentais e autossômicos poderia ser causada por pressão de seleção balanceadora sobre esses últimos, sem que isso implicasse que o modelo de origem africana com substituição estivesse equivocado. Takahata *et al.* (2001) usaram algumas predições dos modelos de origem africana com substituição e de evolução multirregional para testar qual deles melhor se ajustava aos dados de variação de seqüências de DNA. Usando dados de marcadores uniparentais, autossômicos e do cromossomo Y, os autores concluíram que o modelo de origem africana com substituição era amplamente favorecido. Entretanto, Templeton

(2002), igualmente baseado num conjunto de dados que incluía diversos marcadores, usou uma abordagem baseada em *nested clade analysis* (NCA) para sugerir um cenário complexo, onde várias saídas da áfrica com a assimilação de linhagens antigas pelos novos migrantes formariam o conjunto gênico das populações atuais.

I.1.4. Estudos recentes de dados genéticos

Nos últimos anos, o volume de geração de dados genéticos cresceu enormemente. Em relação aos marcadores uniparentais, estudos utilizando o genoma mitocondrial completo não mudaram em quase nada o cenário geral apresentado pelas pesquisas anteriores com esse marcador (Ingman *et al.*, 2000), da mesma forma que a investigação de diversos polimorfismos do tipo SNP no cromossomo Y (Underhill *et al.*, 2000). A ausência de cruzamento entre neandertais e humanos modernos foi também reafirmada por uma re-análise da diversidade mitocondrial desses neandertais (Currat e Excoffier, 2004).

Em relação aos estudos com seqüências autossômicas, análises de seqüências multi-locos (Frissé *et al.*, 2001; Yu *et al.*, 2002; Akey *et al.*, 2004; Voight *et al.*, 2005) têm consistentemente mostrado maior diversidade genética na África, possivelmente devido a um maior tamanho populacional efetivo histórico e/ou à existência de uma subdivisão populacional antiga na África; bem como uma variabilidade reduzida fora dela, sugerindo que as populações não-africanas teriam passado por algum evento do tipo gargalo-de-garrafa que poderia estar associado à saída da África pelos primeiros humanos modernos. Dessa forma, esses dados favorecem fortemente o modelo de origem africana com substituição (mas ver também Templeton, 2005 para um contraponto).

Os resultados obtidos com o estudo de seqüências autossômicas vêm sendo replicados pelo estudo com marcadores do tipo SNPs (polimorfismo de um único nucleotídeo). Esses marcadores permitem também que o nível de desequilíbrio de ligação entre regiões genômicas seja avaliado juntamente com o espectro de freqüência alélica dos polimorfismos genotipados. Desde o estudo de Reich *et al.* (2001), que analisaram 274 SNPs, incluindo os trabalhos de Gabriel *et al.* (2002) e Marth *et al.* (2003; 2004) baseados em ~4.000 e 500.000 SNPs, até o estudo de cerca de 1 milhão de SNPs realizado pelo consórcio HapMap (The International Hapmap Consortium, 2005), os resultados principais mostram que os blocos de desequilíbrio de ligação são significativamente menores em africanos, sugerindo para as populações de fora desse continente uma história marcada por

um evento de gargalo-de-garrafa com posterior crescimento populacional. Schaffner *et al.* (2006) recentemente usaram um conjunto de 4.000 SNPs para estimar diversos parâmetros demográficos de interesse em um modelo de origem africana com substituição.

Grandes conjuntos de dados também vêm sendo gerados para marcadores do tipo STR. Uma bateria de 377 STRs usada originalmente para estimar o grau de estruturação genética presente em populações humanas em nível mundial (Rosemberg *et al.*, 2002) foi analisada por Zhivotovsky *et al.* (2003), que estudaram o padrão e o tempo de divergência entre grandes grupos continentais e encontraram resultados em concordância com a hipótese de origem africana e substituição. O mesmo conjunto de dados foi analisado por Ray *et al.* (2005), que usaram uma análise Bayesiana para estimar a probabilidade posterior de diferentes origens geográficas para a espécie humana. Estes autores encontraram uma maior probabilidade associada a uma possível origem no leste da África. Embora uma origem africana não seja incompatível com modelos alternativos de evolução humana (ver item 1.1), esses autores concluíram que o modelo de origem africana e substituição era o que melhor se ajustava aos dados. Um apoio independente a este modelo veio da análise de cerca de 750 STRs realizada por Ramachandran *et al.* (2005), que sugeriram que um modelo simples de isolamento por distância a partir da África ajustava-se surpreendentemente bem aos dados observados.

Muito embora a grande maioria dos estudos genéticos recentes apóie o modelo de origem africana com substituição, têm sido publicados relatos pontuais, mas de grande impacto, de conjuntos de dados que se ajustam melhor a modelos alternativos. Investigações de genes específicos cujo padrão de variabilidade aparentemente não se adequava às predições do modelo de origem africana com substituição, normalmente devido a um longo tempo de coalescência (TMRCA) ou a um perfil de freqüência haplotípica onde o haplótipo ancestral era mais freqüente fora da África, levaram Garrigan *et al.* (2005a,b), Hayakawa *et al.* (2006) e Evans *et al.* (2006) a sugerir a assimilação de linhagens arcaicas pelas populações modernas de *H. sapiens*. Alguns estudos de locos múltiplos, entretanto, também sugeriram a assimilação de linhagens ancestrais como cenário que melhor explicava os padrões de variabilidade encontrados, sendo que a taxa de assimilação poderia afetar entre 5% (Plagnol e Wall, 2006), ou até 80% do genoma de nossa espécie (Eswaran *et al.*, 2005). Finalmente, a publicação de dados preliminares acerca do genoma do Homem de Neandertal gerou resultados conflitantes. Enquanto um

grupo (Noonan *et al.*, 2006) obteve uma estimativa de máxima verossimilhança de zero (com intervalo de confiança entre 0 e 20%) para uma possível contribuição dos neandertais para o conjunto gênico de nossa espécie, outra equipe (Green *et al.*, 2006) sugeriu um cenário de algum fluxo gênico envolvendo principalmente *H. sapiens* masculinos.

Os estudos já realizados sobre as origens dos humanos modernos, sempre basearam a escolha do seu modelo evolutivo favorito em um determinado conjunto de previsões acerca do padrão de diversidade genética que poderia ser refutado ou corroborado. Jamais, porém, tentou-se comparar diretamente modelos evolutivos alternativos para estimar a probabilidade relativa de cada modelo.

I.2. O Povoamento das Américas

Durante a dispersão do *H. sapiens* moderno após seu surgimento, o continente americano foi o último a ser povoado. Os povos que se espalharam por toda sua extensão desenvolveram uma grande variedade de culturas adaptadas ao ambiente específico de cada tribo. Passados ~500 anos da chegada dos colonizadores europeus, estima-se uma redução populacional de 95% dos nativos americanos (Cavalli-Sforza *et al.*, 1994), extinguindo grande parte desta diversidade. Atualmente, enquanto vários grupos sobrevivem em relativo isolamento, outros foram incorporados (geneticamente, inclusive) à sociedade colonial como escravos ou peões. A contribuição cultural destes grupos aborígenes foi marcante na gênese de uma "cultura colonial" miscigenada, em oposição à cultura metropolitana européia (Kern, 1998). Por tratar-se de um tema multidisciplinar, lingüistas, arqueólogos, antropólogos físicos e geneticistas têm formulado hipóteses e modelos sobre o povoamento das Américas (revisão em Salzano, 2007).

I.2.1. Evidências antropológicas

Um dos modelos mais influentes sobre o povoamento das Américas é o de Greenberg *et al.* (1986). Utilizando evidências fundamentalmente lingüísticas e de morfologia dental e tendo certo respaldo de marcadores genéticos de grupos sangüíneos e protéicos, estes autores separaram os nativos americanos em Ameríndios, Na-Denes e Esquimó-Aleutas. Cada grupo corresponderia a uma migração distinta, sendo os Ameríndios os mais antigos (>11.000 anos atrás), seguidos pelos Na-Denes (9.000 anos atrás) e finalmente pelos Esquimós e Aleutas (4.000 anos atrás). Embora a análise

lingüística de Greenberg seja muito criticada (ver Bolnik *et al.*, 2004), sua importância histórica é inegável.

Outro ponto controverso é se o complexo arqueológico Clovis, localizado no centro dos EUA e datado até recentemente em 11.500 anos atrás representaria os restos líticos de caçadores de grandes animais de pradaria, possivelmente os primeiros habitantes do Novo Mundo (Steele e Powell, 1993) conforme predito pelo modelo de Greenberg *et al.* (1986). A descoberta de novos sítios começou a mudar esta visão (ver Meltzer, 1993; Roosevelt *et al.*, 1996). Alguns autores sugeriram que os primeiros habitantes do continente seriam caçadores-coletores florestais que teriam penetrado no continente ~25.000 anos atrás seguindo uma rota costeira (Rogers *et al.*, 1992; Prous, 1995). Assim, não existiriam sítios arqueológicos desta antigüidade porque estes teriam sido submersos com a elevação do nível do mar ao final do Pleistoceno. A aceitação da data de 14.600 anos atrás para o sítio chileno de Monte Verde (Meltzer, 1997) e a revisão nas datas do sítio de Clovis situando-o a cerca de 13.000 anos atrás, contemporâneo a outros sítios nas Américas do Norte e do Sul (Waters e Stafford Jr, 2007) passaram a favorecer fortemente uma entrada nas Américas anterior a Clovis.

Recentemente, a antropologia física tem provocado um grande debate sobre os modelos de povoamento das Américas a partir da identificação de morfologias proto-mongoloides (p. ex. Neves e Hubbe, 2005; Neves *et al.*, 2005; 2007) nos esqueletos mais antigos já identificados no continente. Dentre as hipóteses mais influentes está a de Neves *et al.* (1999), na qual teria havido uma migração antiga que teria trazido para a América indivíduos de morfologia bastante distinta do padrão mongolóide atual. A seguir, uma nova migração de indivíduos de morfologia mongolóide teria ocorrido no início do Holoceno, e esses grupos, dotados de melhor tecnologia ou mais bem adaptados às condições locais teriam causado a total extinção dos grupos antigos, de modo similar à suposta substituição de humanos arcaicos por modernos (ver item 1.1). Uma hipótese alternativa sugere que subseqüente à chegada dos povos de morfologia protomongolóide ao continente, uma nova "migração" teria ocorrido a partir da entrada de povos mongolóides no extremo norte do continente. Segundo esse modelo, a morfologia dos povos nativos americanos, embora sempre muito diversa, teria se transformado a partir de um evento de troca genética entre mongolóides asiáticos e protomongolóides (González-José *et al.*, submetido). Um ponto que ainda necessita ser estudado em mais detalhes é o quanto de fluxo gênico seria

necessário para promover a mudança morfológica e qual o papel que variáveis ambientais poderiam ter na promoção da diferenciação morfológica dentro do continente (Bernal *et al.*, 2006; Sardi *et al.*, 2006).

I.2.2. Marcadores genéticos: número e idade das ondas migratórias

Em relação ao DNA mitocondrial (mtDNA), Schurr *et al.* (1990), utilizando marcadores RFLP foram os primeiros a identificar uma ancestralidade asiática para os ameríndios modernos pela existência de quatro grandes haplogrupos. Uma aparente diferenciação entre Ameríndios e Na-Dene/Escaleutas levou Torroni *et al.* (1992) a proporem um modelo de duas migrações para a origem dos Ameríndios, sendo a idade da primeira migração ~30.000 anos. Starikovskaya *et al.* (1998) sugeriram posteriormente que a migração mais recente corresponderia à cultura Clovis. Um cenário alternativo foi inicialmente apresentado por Merriwether *et al.* (1995) e Bonatto e Salzano (1997a,b), que propuseram uma única onda migratória para a colonização da Beríngia, a ponte de terra que ligava a Ásia à América, e também uma entrada antiga no continente (~30.000 anos atrás) com a posterior diferenciação entre os grupos (Ameríndios x Na-Denes + Esquimós) sendo esta devida ao isolamento geográfico durante o ápice do período glacial. Recentemente, Silva *et al*. (2003) seqüenciaram 8kb do mtDNA e obtiveram um intervalo de confiança bastante estreito para a colonização da Beríngia, entre 18.600 e 23.400 anos atrás, claramente no Pleistoceno.

Um ponto ainda obscuro refere-se ao haplogrupo X, de diversidade genética aparentemente mais limitada que os demais, raro tanto na Europa quanto na Ásia Central e nas Américas, onde é restrito à América do Norte (Schurr, 2004; Dornelles *et al.*, 2005). Brown *et al.* (1998) sugeriram que este haplogrupo representava uma migração independente, talvez via Europa. A existência do haplogrupo X em baixas freqüências na Europa, e uma suposta similaridade entre as tecnologias líticas de Clovis e dos Solutreanos levou à hipótese de que ambas as culturas eram de fato ligadas diretamente (Stanford e Bradley, 2002)

Já em relação ao cromossomo Y, os primeiros trabalhos publicados apontavam para um forte efeito fundador no cromossomo Y, sugerindo que as Américas teriam sido povoadas por portadores de um conjunto bastante limitado de seqüências (p. ex. Pena *et al.*, 1995; Santos *et al.*, 1996; Lell *et al.*, 1997; Bianchi *et al.*, 1998; Santos *et al.*, 1999),

em um modelo de migração única no final do Pleistoceno (~22.700 anos atrás), em um forte paralelo com os dados de mtDNA obtidos por Bonatto e Salzano (1997a,b).

O estudo de mais marcadores dentro do cromossomo Y vem possibilitando a descoberta de novos haplótipos fundadores, e Karafet *et al.* (1999) e Lell *et al*. (2002) passaram a sugerir a existência de duas migrações, sendo a migração mais antiga (20-30.000 anos atrás) originada na região centro-sul da Sibéria, enquanto a outra (7.000-9.000 anos atrás) seria originária do leste siberiano, próximo ao mar de Okhotsk. Novos dados têm apoiado o modelo de duas migrações distintas, bem como uma entrada mais recente no continente americano, provavelmente próxima a 13.500 anos atrás (Bortolini *et al.*, 2003; Seielstad *et al.*, 2003). Entretanto, outros estudos têm mostrado conclusões discrepantes, sugerindo a existência de uma única migração (Tarazona-Santos e Santos 2002; Zegura *et al.,* 2004) que teria ocorrido entre 10.100 e 17.200 anos atrás (Zegura et al., 2004).

Até o momento, a única tentativa de datar a entrada no continente americano usando também dados de seqüências autossômicas (e do cromossomo X) é a de Hey (2005). Curiosamente, a análise apresentada por esse autor revela que os dados favorecem ou idades próximas a 44.000 anos ou a 7.000 anos, em ambos os casos fora do intervalo sugerido pelos estudos de marcadores uniparentais.

I.2.3. Marcadores genéticos: tamanho da população fundadora

Apesar dos avanços no estudo do Povoamento das Américas possibilitados pelo estudo de marcadores genéticos, pouco se avançou em relação a uma estimativa consistente do tamanho da população fundadora e de que processos demográficos teriam ocorrido nessa população subseqüentemente ao povoamento. Em relação ao mtDNA, embora alguns estudos tenham identificado gargalos-de-garrafa e expansões populacionais neste marcador (p. ex., Bonatto e Salzano, 1997b; Silva *et al*., 2003), nenhum destes trabalhos conseguiu mensurar a magnitude do efeito gargalo-de-garrafa ou da expansão populacional com grande confiança, e apenas Bonatto e Salzano (1997b) apresentaram alguns dados numéricos, embora imprecisos, sobre estes possíveis eventos. A ausência de dados quantitativos sobre os eventos demográficos durante o povoamento das Américas repete-se nos estudos baseados no cromossomo Y, mesmo naqueles mais recentes, que parecem ligar as linhagens americanas a poucos haplótipos fundadores (Lell *et al.*, 2002; Bortolini *et al.*, 2003; Seielstad *et al.*, 2003; Zegura et al., 2004). Esta situação também

aparece nos estudos de polimorfismos nucleares (protéicos, de RFLP e de inserções *Alu*), que tendem a rejeitar um efeito gargalo-de-garrafa (p. ex. Kidd *et al.*, 1991; Callegari-Jaques *et al*., 1993; Nowick *et al*., 1998; Heller *et al.,* 2004; Mateus Pereira *et al.*, 2005; Battilana *et al.*, 2007), ou sugerir um efeito moderado (Fagundes *et al.*, 2005 – ver Anexo I; Battilana *et al.*, 2006 – ver Anexo II) embora não existam dados quantitativos.

Recentemente, Hey (2005) fez uma análise bastante refinada usando dados de vários locos publicados e concluiu que o continente americano poderia ter sido povoado por uma população muito pequena, de até 80 indivíduos. Porém, os dados reunidos por ele compreendiam trabalhos que utilizaram estratégias de amostragem muito diferentes e não se sabe até que ponto o modelo de análise (que já assumia de antemão algum grau de gargalo-de-garrafa) poderia ter influenciado os resultados. Até o momento, nenhum trabalho foi feito em populações nativas americanas utilizando o seqüenciamento de múltiplos locos nucleares em uma mesma amostra.

I.3. A escolha dos marcadores genéticos estudados no presente trabalho

Dois conjuntos de dados serão utilizados na presente tese. O primeiro (ver Cap. 2) é um conjunto de seqüências de 50 locos autossômicos previamente caracterizados por Yu *et al.* (2002) em uma amostra de indivíduos africanos, asiáticos e europeus. Como foi ressaltado no item anterior, até o momento, nenhum trabalho foi feito em populações nativas americanas utilizando o seqüenciamento de múltiplos locos nucleares em uma mesma amostra. A genealogia de uma única unidade de recombinação, seja autossômica, do mtDNA, ou do cromossomo Y, fornece apenas a história daquela região, que pode ser completamente diferente da história demográfica da população em estudo (Knowles e Maddison, 2002; Marjoram e Tavare, 2006).

A utilização de dados provindos de estudos de seqüenciamento possui ainda algumas vantagens sobre dados do tipo SNP. Atualmente, milhões de SNPs estão disponíveis para genotipagem automatizada. Entretanto, o processo de descoberta, validação e seleção de SNPs é enviesado contra polimorfismos raros (Clark *et al.*, 2005). Embora teoricamente seja possível corrigir esse tipo de viés se é sabido como os marcadores foram selecionados (Brumfield *et al.*, 2003), estudos recentes mostram que mesmo quando o processo de seleção de SNPs é conhecido, as estimativas de diversidade permanecem enviesadas mesmo após correção (Clark *et al.*, 2005). Por outro lado, estudos

de seqüenciamento não possuem nenhum tipo de viés na estimativa dos parâmetros de diversidade constituindo-se, assim, em excelentes marcadores para estudos populacionais.

O segundo conjunto de dados utilizado na presente tese é composto de 53 genomas mitocondriais inéditos de nativos americanos, que somados a outros já publicados formaram um conjunto de 244 genomas mitocondriais de nativos americanos incluindo os 5 principais haplogrupos mitocondriais presentes no continente (ver item 2.2). O genoma mitocondrial tem uma taxa evolutiva maior em relação ao DNA nuclear, o que o torna um marcador adequado para datar eventos recentes (Brown *et al.*, 1979). Os estudos iniciais com mtDNA humano concentraram-se na região controladora, cuja evolução é ainda mais rápida. Porém, os sítios da região controladora exibem uma heterogeneidade de taxa evolutiva muito grande, e a distância entre humanos e chimpanzés não pode ser estimada de modo trivial, dificultando a definição de uma escala de tempo para a evolução dessa região (Wakeley, 1993; Tamura e Nei, 1993). Recentemente, diversos estudos passaram a utilizar o seqüenciamento de toda a região codificante do mtDNA, que possui uma taxa evolutiva não tão rápida, uma heterogeneidade de taxas menos extrema e, portanto, presta-se melhor à datação de eventos demográficos ou divergências entre haplogrupos (Torroni *et al.*, 2006). Apesar de existirem seqüências de mtDNA completas de haplótipos nativos americanos em bancos de dados públicos, ainda não foi realizada uma análise compreensiva dessas seqüências. Igualmente importante é o fato de que até o momento poucas dessas seqüências pertencem ao haplogrupo X, fazendo-se necessário que sejam gerados mais dados. A análise de mtDNA completos já possibilitou uma série de avanços no entendimento da filogenia dos haplogrupos mitocondriais, além de fornecer datações mais precisas para eventos importantes como a coalescência do mtDNA humano e a saída dos humanos modernos da África (para uma breve revisão, ver Torroni *et al.*, 2006).

I.4. Objetivos

Os objetivos da presente tese são:

1. Avaliar de maneira estatística, através de simulação sofisticada e de um grande número (50) de marcadores genéticos autossômicos, hipóteses alternativas sobre como ocorreu a diferenciação continental dos grupos humanos modernos; e

2. Através da análise de 244 genomas mitocondriais completos re-examinar a época e a maneira como deve ter ocorrido a colonização das Américas.

CAPÍTULO II

Statistical Evaluation of Alternative Models of Human Evolution

Manuscrito submetido à PLoS Genetics

Statistical Evaluation of Alternative Models of Human Evolution

Nelson J. R. Fagundes[1,2,3], Nicolas Ray[3], Samuel Neuenschwander[3,4], Mark Beaumont[5], Francisco M. Salzano[2], Sandro L. Bonatto[1*] & Laurent Excoffier[3*]

[1] Centro de Biologia Genômica e Molecular, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), 90619-900 Porto Alegre, RS, Brazil.

[2] Departamento de Genética, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil.

[3]Computational and Molecular Population Genetics (CMPG), Zoological Institute, University of Bern, CH-3012 Bern, Switzerland.

[4]Dept. of Ecology and Evolution, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland.

[5]School of Animal & Microbial Sciences, University of Reading, RG6 6AJ, Reading, United Kingdom.

*To whom correspondence should be addressed.

E-mail: slbonatto@pucrs.br; laurent.excoffier@zoo.unibe.ch

An appropriate model of recent human evolution is not only important to understand our own history, but it is necessary to disentangle the effects of demography and selection on genome diversity. While most genetic data support the view that our species originated recently in Africa, it is still unclear if it completely replaced former members of the *Homo* genus, or if some interbreeding occurred during its range expansion. Several scenarios of modern human evolution have been proposed on the basis of molecular and palaeontological data, but their likelihood has never been statistically assessed. Using DNA data from 50 nuclear loci sequenced in African, Asian and American samples, we show here by extensive simulations that a simple African Replacement model with exponential growth has a much higher probability (98%) than alternative multiregional evolution or assimilation scenarios. A Bayesian analysis of the data under this best supported model points to an origin of our species ~145 thousands years ago (Kya), an exit out-of-Africa ~54 Kya, and a recent colonization of the Americas 9.5 Kya. We also find that the African replacement model can not only explain the shallow ancestry of mtDNA or Y-chromosomes, but also the occurrence of deep lineages at some autosomal loci, which has been formerly interpreted as a sign of interbreeding with *H. erectus*.

Introduction

Recent international efforts have produced a large amount of genetic data [1] to identify loci involved in complex diseases or genomic regions with unusual patterns of polymorphism that could be indicative of recent selective events [2]. However, because past demographic events are likely to have greatly affected current patterns of genetic diversity, genetic data are difficult to interpret without a general demographic model that

can explain neutral variability [3]. A global scenario of human evolution is also important to understand our origins and how and when human populations have colonized the globe, a question that has fascinated physical and molecular anthropologists over the past decades [4].

Many scenarios of human evolution have been proposed based on palaeontological, archaeological, or genetic data [5,6], and their fit to various aspects of our genetic diversity has been investigated [3,7-9]. The current debate over recent human evolution can be simplified by considering the alternative scenarios shown in Fig. 1 [5]. The African replacement scenarios (Fig. 1A), which posits a single and recent African origin for all modern humans, are mainly supported by mitochondrial DNA (mtDNA) and Y chromosome polymorphisms [4], by the current lack of Neandertal mtDNA genes in modern humans [10], and by gradients of nuclear genetic diversity from Africa towards the Americas [4,11]. Recent examination of nuclear DNA has however revealed patterns of polymorphism that were judged incompatible with a pure African Replacement scenario [7,12-16]. For instance, the presence of very old lineages in Africa and Asia raised claims for some degree of interbreeding between modern and archaic *Homo* forms [13,14,16]. Such interbreeding can occur under Assimilation scenarios (Fig. 1B), where modern humans migrating out-of-Africa would have hybridized with local *H. erectus* and incorporated old lineages [15,17], or under Multiregional scenarios (Fig. 1C), where migrants would have been continuously exchanged between Africa and Asia, leading to a synchronized emergence of modern anatomy.

Previous approaches to understand human evolution using genetic data have not attempted to compare directly alternative scenarios within a global statistical framework, and the posterior probability of the models presented above has never been evaluated. In

principle, alternative models can be directly compared if their likelihood can be computed. Even though these likelihoods can now be computed for relatively simple demographic scenarios involving a few parameters [18], the likelihood function of complex demographic scenarios may be very difficult or even impossible to solve analytically [19]. In this paper, we overcome this problem by taking an Approximate Bayesian Computation (ABC) approach [20] to compare models and estimate the parameters of interest. The ABC approach is a convenient way of dealing with such situations because it is possible to compare the probability of obtaining the observed data (or summary statistics computed from them) under alternative scenarios, marginal to (i.e. irrespective of) the parameter values. Complex models can thus be compared even though they depend on many parameters, the true values of which are very uncertain.

Results

We first evaluated the posterior probabilities of different models within each class of the three scenarios considered here, which are the African replacement, multiregional evolution and assimilation scenarios (see Fig 1. and Material and methods section for further information on the models). Under the African replacement and Assimilation scenarios, models with exponential growth (AFREG and ASEG) were found to have the largest posterior probabilities (0.997 for AFREG, and 0.979 for ASEG, Fig. 1A and B), suggesting that both the emergence of modern humans in Africa and their spread into other continents are better modeled as a gradual rather than an instantaneous process. Among the Multiregional evolution models, the MREBIG model (Fig. 1C) is clearly favored with a posterior probability of 0.62. This model implements a bottleneck in Africa with an instantaneous recovery, a recent population growth in Asia, and it allows for different migration rates between Africa and Asia at different periods.

We then compared the best model of each scenario. We find that the African Replacement model with exponential growth is clearly outcompeting the others, with a posterior probability of 0.978 (Figure 1). The best Multiregional evolution and Assimilation models have much lower posterior probabilities of 0.022 and $5x10^{-5}$, respectively. These results clearly show that neutral nuclear sequence data give a significant support to a recent origin of modern humans without any interbreeding with archaic *Homo* forms, at least in Asia.

The Bayesian estimates of the demographic parameters of the overall best African replacement model (AFREG, Table 1, Fig. 2) suggest a scenario of human evolution where an archaic African population of about ~14,500 effective individuals gave rise to modern humans around ~145 thousand year ago (Kya) after a bottleneck involving ~550 effective individuals. The Out-of-Africa migration, initially involving only ~300 effective individuals would have occurred some ~54 Kya, and the Americas would have been colonized only around ~10 Kya by ~500 individuals.

Discussion

In order to check the power of our model choice procedure, we simulated 1,000 data sets under the best African replacement (AFREG) and multiregional (MREBIG) models. For data simulated under the African replacement scenario, we found that the AFREG model had a higher probability than the multiregional model in ~90% of the cases (see Supporting Information Fig. S2), while for data simulated under the multiregional scenario, the MREBIG model had a higher probability than the African replacement model in only ~67% of the cases. While these results could suggest that a true multiregional model would be wrongly interpreted as an African replacement model in ~33% of the cases, we note that misidentified AFREG models never exceed posterior probabilities of

0.94. Thus, posterior probabilities for the African replacement model as high as that observed (0.98) is only found when it is the correct model, suggesting that our results are not due to some artifact in the model selection procedure.

The demographic and time estimates (Table 1) are in overall good agreement with those obtained previously from fossil or genetic data. The date for the emergence of modern humans is indeed well consistent with palaeontological record suggesting dates of 130-200 Kya [5,21], and with previous genetic estimates [120-160 Kya, 22]. The size of the archaic modern human population is also close to recent estimates of "long-term" or ancestral size for modern humans of around 10,000-15,000 individuals [3,7]. The size and timing of the exit out of Africa are in excellent agreement with recent molecular and archaeological studies suggesting that this migration resulted in a limited number of lineages having left Africa only around 55- 65 Kya [6,11,23]. Finally, the estimates for the current effective continental population sizes show a net decrease from Africa to America compatible with a series of spatial expansions and founder effects during the colonization of the world [4,11].

Our estimated date for the colonization of the Americas is very recent, although the upper limit of the 95% HPD (Highest Posterior Density) is close to the dates of the oldest archaeological sites of ~14 Kya [24]. This result thus nevertheless suggests a late, post-glacial maximum, colonization of the Americas, which is in better agreement with the estimates of ~14 Kya based on the Y-chromosome [25] than on those of ~30 Kya based on mtDNA control region [26]. This young colonization time, in agreement with a recent study [27], could indicate a discontinuity between the ancestors of our sampled individuals and the earliest settlers of the Americas. Alternatively, a too young settlement time could result from the sole sampling of Central and South American individuals, and it is known

from the study of mtDNA and the Y-chromosome that some rare alleles (haplogroups) can only be found in North America [28]. Therefore, the inclusion of northern Native Americans could lead to increased genetic diversity and colonization time estimates. We also note that the estimated founder population size for America is about 6 times larger than that recently proposed by Hey [27], who suggested that less than 80 effective individuals would have colonized the Americas. However, a moderate bottleneck for the settlement of the New World is in agreement with recent results from nuclear loci [29] and with previous mtDNA studies [26]. Differences in sampling design and marker choice between studies could explain this discrepancy: while our study is based on a homogeneous set of 50 nuclear loci genotyped in the same individuals, the former study [27]used a mixture of nuclear, mtDNA and Y chromosomes markers assessed on different individuals drawn from various locations.

When comparing the values of homologous parameters estimated under different models (Supporting Information Table S5), we find that the ASEG model clearly converges to the AFREG model as shown by the very small proportion (<0.5%, Supporting Information Table S5) of archaic lineages that would have introgressed in non-African populations under this model. This suggests that even a small archaic contribution to the modern non-African gene pool results in larger discrepancies between simulated and observed data, at odds with previous results [7,15]. In contrast to the ASEG model, the Multiregional evolution model most compatible with our dataset (MREBIG), however, does not show any convergence towards an African Replacement model, or towards previous implementations of a multiregional model [e.g. 8, 30], where a large archaic African population would send more migrants to Asia than the reverse. The median estimates of the MREBIG model (Supporting Information Table S5) rather suggest small

archaic population sizes in both continents (less than 1,000 effective individuals), and recent migration rates between continents being larger than between the two archaic populations.

Since the occurrence of deep lineages in modern humans has been sometimes taken as evidence against replacement models [e.g. 13,16], we have computed the empirical distribution of the times to the Most Recent Common Ancestors (TMRCAs) for the best model under each of the three scenarios (Fig. 3A, Supporting Information Table S6). We see that the Multiregional model has the narrowest and shortest distribution due to the small estimated archaic population size that promotes coalescent events as soon as archaic Asian lineages are brought back (looking backward in time) to Africa ~ 800 Kya. On the other hand, very old TMRCAs exceeding several millions of years can be readily obtained under the African replacement models, since the larger ancestral size in Africa prevents rapid coalescence of lineages having passed through the speciation bottleneck. When computing continent-specific TMRCAs under the overall best African replacement model (AFREG) (Fig. 3C, Supporting Information Table S8), we see that very ancient TMRCAs are not restricted to African samples, but that they are also found for Asian and Amerindian samples. Our results therefore question the hypothesis that very old TMRCAs should be taken as evidence for interbreeding events between modern humans and individuals of other *Homo* species [13]. Unexpectedly, we find that ~10% of autosomal loci should have TMRCAs younger than 140 Kya, which seems to contradict empirical observations [13] even though we cannot exclude the possibility that loci with little or no variation are under-represented in the published literature. However, in keeping with these results, we note that 8% of our loci (4 out of 50) are entirely monomorphic in the three samples, and therefore indicative of a shallow ancestry.

While our models were fitted to autosomal DNA, they should also be able to explain the observed features of mtDNA and Y chromosome polymorphism, such as their more recent TMRCAs. We have therefore simulated TMRCAs for these uniparentally inherited loci using effective sizes four times smaller than for nuclear loci. We find (Fig. 3B, Supporting Information Table S7) that the African replacement and Assimilation models are fully compatible with TMRCAs smaller than 250 Kya such as those found with mtDNA or Y chromosome data [e.g. 4], while TMRCAs are found mostly larger than 400 Kya for the best multiregional model, which seems therefore fully incompatible with these uniparentally inherited markers.

We thus show in this paper that it is possible to estimate the posterior probability of various models of human evolution under an approximate Bayesian computation framework relying on massive computer simulations. While the analysis of 30 individuals for 50 unascertained and neutral DNA sequence loci is still challenging for small laboratories, it is reassuring that this dataset seems sufficient to obtain unequivocal results. It suggests that complex evolutionary models could also be tested in non-model organisms. While we considered a variety of alternative scenarios, we did not specifically attempt to design models of human evolution that would maximize the fit between observed and simulated data. However, these very simple models certainly capture the basic differences between proposed alternative scenarios of human evolution [see e.g. 5]. More elaborated models incorporating intra-continental population subdivisions, long-distance dispersal, or spatially explicit information [8] could certainly be implemented, and the current model choice framework could be used to evaluate their respective merits. An analysis of genome-wide resequencing data [e.g. 31] or of STR data performed on population samples from various continents [32] would be helpful to confirm our results and would possibly

allow one to refine our estimates. However, these much larger data sets would be more challenging to study and would require much more computer power than that used in our study, which already exceeded 10 CPU-months of computations on a Linux cluster.

In conclusion, while our best supported model (African replacement with exponential growth) certainly does not represent the complete history of modern humans, we show here that it is much better supported by a random set of neutral loci than any other models involving interbreeding with other *Homo* species. We certainly cannot fully exclude that any interbreeding ever occurred between modern and archaic humans, and that any favorably selected *H. erectus* genes could have spread into modern humans [see e.g. 17]. However, our results clearly suggest that our modern gene pool has a recent and predominant African origin, and they therefore offer a neutral demographic scenario that could be used to detect ancient admixture for specific gene regions. Moreover, the best African replacement model explains key features of other data sets, such as recent TMRCAs for mtDNA or Y chromosome loci, as well as occasional deep lineages of nuclear loci, previously thought to be indicative of balancing selection or interbreeding with *H. erectus* or Neandertals [7,13]. The demographic parameters of this model should reveal useful to improve our ability to detect loci involved in complex diseases or in past adaptive events, by providing better null distributions of various statistics used in genome scans or linkage disequilibrium mapping studies.

Material and Methods

Samples, loci and laboratory methods. For this study, we sequenced 50 independent autosomal loci for about 500 bp each, providing a total of about 25,000 bp information for each individual (see Supporting information Table S1). These 50 loci were first characterized by Chen and Li [33], and further studied in human and chimpanzee

populations [34,35]. They were selected after a preliminary screen of the human genome because they lie in intergenic regions located at least 5 kb away from known or putative functional element, and because they do not contain repetitive elements [33]. Additionally, each of these nuclear sequences are short enough (approximately 500bp) so that they can be considered as non-recombining segments. Because these data have been generated through DNA sequencing, they are not likely to be affected by ascertainment bias.

In order to complement a first data set consisting of 10 African, and 10 Asian individuals previously analyzed by Yu *et al*. [34], we sequenced here 12 Native American individuals, each affiliated to a different tribe. All individuals came from Central and South American populations which belong to the Amerind linguistic phyla [36]. The populations sampled and their linguistic classifications [following Greenberg 36] are: Aché (Equatorial, Kariri-Tupi), Arara (Macro-Carib, Carib), Bribri (Chibchan, Talamanca), Guatuso (Chibchan, Rama), Guaymi (Chibchan, Guaymi), Kuben-Kran-Kegn (Macro-Ge, Cayapo), Lengua (Macro-Panoan, Lengua), Quechua (Andean, Quechua), Tiryio (Macro-Carib, Carib), Waiwai (Macro-Carib, Carib), Xavante (Macro-Ge, Ge-Kaingang) and Zoró (Equatorial, Kariri-Tupi). This "scattered" sampling scheme was used to replicate the sampling strategy of Yu *et al*. [34], who studied a single individual by ethnic group in order to get a general picture of the genetic diversity at the continental level with a limited sample size. To our knowledge, this is the largest multilocus sequence dataset available for Native Americans in which the same panel of individuals has been studied for all loci.

For each locus, we performed PCR amplification using primers and conditions described in Yu *et al.* [34], except for loci T1469, T151, T812, T1386, and T864, for which we designed new amplification primers, whose sequence is available upon request. Sequencing was performed at the Centro de Biologia Genomica e Molecular, PUCRS, in a

MegaBACE1000 system (GE Healthcare) using reagents and protocols recommended by the manufacturer. Individual reads were assembled in the PhredPhrap package [37], together with a reference sequence containing the known variants for each locus. All assemblies were visually inspected using Consed [38], and all possible heterozygous sites have been re-checked using a new PCR product as a template for sequencing. Mutation rates at all loci were estimated after gametic phase estimation and comparison with chimpanzee sequences, as explained in Supporting Information Table S1.

Note that two East Indian individuals were excluded from the Asian sample, since the Indian sub-continent has been recently colonized by Indo-Europeans, to which West-Asians are genetically most similar [e.g. 39]. Also, while 10 European individuals were also sequenced for the same 50 loci [34], we did not incorporate them in the present study for the following reasons. First, it appears that the colonisation of Europe by modern humans has been quite complex, with a delay compared to the colonisation of Asia, and several migration waves from the Near-East whose contribution to the present European gene pool is difficult to assess [see e.g. 40]. Therefore, due to the uncertainty about this settlement history, the modelling of Europe's colonization would require the introduction of many additional parameters in our simulations, which would become overly complex.

Tested evolutionary scenarios. We modelled three different sets of scenarios constructed to capture most of the current debate concerning modern human evolution (see e.g. refs. [5,41] for a general account on different models of human evolution). Because there is still some uncertainty on the exact details of past human demography, we chose to evaluate several alternative models within each class of scenarios. For example, previous attempts of fitting molecular data to the African Replacement scenario have used different

demographic growth models (instantaneous, exponential, linear, or logistic) [3,7-9,22], but it is still unclear if one of these models has better properties than others.

A general representation of the models contrasted in this study is shown in Fig. 1, and a detailed schematic representation is shown in Supporting information Fig. S1, where we list the parameters of all models. The African replacement models (Fig. S1A) are simulated with instantaneous (AFRIG) or exponential (AFREG) growth after bottlenecks. Looking forward in time, both models start with an ancestral (archaic) population in Africa which passes through a bottleneck and gives rise to a population of modern humans. After the bottleneck, the population is allowed to grow to its current size, either instantaneously or exponentially, depending on the model. Following this event, a migration occurs from Africa to Asia, and finally from Asia to the Americas. In both cases after a few generations the founding population is allowed to expand to its current size.

Multiregional evolution, in which the transition towards modern morphology occurs simultaneously due to ongoing gene flow between continents was simulated as shown in Fig. S1C. We simulated four different models that differ in the way population sizes change over time and whether population growth has been instantaneous or exponential. Forward in time, all models start with an archaic African population that moves out-of-Africa in an event that attempts to model the peopling of Asia by *Homo erectus*. Since then, and up to the present, Africa and Asia exchange migrants. Another major migration event only takes place from Asia to the Americas. In model MRE1S, African and Asian population sizes and migration rates are held constant over the whole simulated period. In model MRE2S, there is a transition between an “archaic” and “modern” population size that occurs independently first in Africa and then in Asia, with new migration rates occurring after the demographic transition in Africa. The remaining models implement a

bottleneck in Africa during the emergence of modern humans: in model MREBIG all populations grow instantaneously, while in model MREBEG all “modern” populations grow exponentially.

Finally, the African origin with assimilation (Supporting information Fig. S1B) is a “hybrid” model that includes an early dispersal of *H. erectus* out-of-Africa, but it differs from MRE in two major aspects: there is no migration between continents and a fraction of ”modern” Asian lineages have originated recently from Africa, like in the African replacement model. However, another fraction of the “modern” Asian lineages come from the archaic Asian population. The ASIG and ASEG models differ by implementing instantaneous or exponential growth, respectively, after the bottlenecks associated to the founding of each continent by “modern” humans. These scenarios have been adapted from the models reviewed in Stringer [41]. The prior distributions of the parameters of the eight tested models are described in Supporting information Table S2 (see next section).

Approximate Bayesian Computations. Parameter estimation and model evaluation were done under an approximate Bayesian computation (ABC) framework [20], implemented in a number of programs developed by us (M.B, L. E., S. N. and N. R). The different steps of the ABC parameter estimation procedure are described in detail elsewhere [20,42], but we briefly outline them below. For each model, we first perform a large number of genetic simulations based on a pre-defined demographic history, using the program SIMCOAL ver. 2 [43]. Some or all parameters that define the model (*e.g.* population sizes, migration rates, timing of the demographic events, mutation rates) are considered as random variables for which some prior distribution must be defined, as shown in Supporting information Table S2. For each simulation, the parameter values are drawn from their prior distributions defining a demographic history that is used to build a

specific input file for the SIMCOAL program. SIMCOAL then performs coalescent-based [44] simulations to generate the genetic diversity of samples, with the same number of gene copies and loci than those observed. Summary statistics (S_{sim}) identical to those computed on the observed data (S_{obs}) are then calculated for the simulated dataset. As in any coalescent approach, our simulations were performed considering haploid individuals and with time scaled in generations. Following Beaumont *et al.* [20], a Euclidean distance δ is calculated between normalized S_{sim} and S_{obs} for each simulated dataset.

Prior distributions. The prior distributions of the parameters of all eight models are shown in Supporting information Table S2. We used uniform priors for parameters with a search space made up of discrete values or of continuous values with one order-of-magnitude between the smallest and largest values, or for parameters where no prior information is available (e.g. the timing of all events, bottleneck population sizes, duration of the bottlenecks, and the fraction of the archaic chromosomes to invade the "modern" Asian population). We used a log-uniform distribution for parameters with a larger search space, such as the current population size, migration rates, and the "archaic" population sizes in Africa and Asia. This strategy implies that the sampling for these parameters is denser for smaller values, which seems reasonable since most studies suggest population sizes of a few thousand individuals [30], or low recurrent migration between continents [45].

Summary statistics. Summary statistics of genetic diversity were calculated using program Arlequin ver 3.1 [46]. The following summary statistics were computed: total and per population number of segregating sites (S), nucleotide diversity (π) for each population, Tajima's D [47] for each population, total and pairwise F_{ST}'s [48]. Since there is some uncertainty associated to the phasing procedure, we only used summary statistics

that do not depend on phase information. Summary statistics calculated for the 50 loci are reported in Supporting information Table S3.

Framework for model choice. The posterior probability of each model was estimated by an approach developed by one of us [49] (http://sapc34.reading.ac.uk/~mab/stuff/ABC_distrib.zip), which is based on a weighted multinomial logistic regression procedure. This is an extension of ordinary logistic regression to more than two categories. Logistic regression gives the probability that a categorical variable takes one of two states as a function of the explanatory variables. For the ABC procedure, the different models are coded as categorical variables and the method then directly estimates the posterior probability of each model, conditional on the observed summary statistics. The regression of summary statistics on the models is carried out on the 5,000 retained simulations with smallest δ for all models pooled together, and Epanechnikov kernel-based weights are assigned to each simulation [49]. This procedure has been shown [49] to substantially improve on a previous method [50,51] for selecting models using ABC. It should be noted that the posterior probabilities of particular models may depend on the choices made for the prior distributions of the parameters within each model. However, because we have examined different models within each set of scenarios, our conclusions are likely to be robust to most reasonable specifications of the priors. Model selection within each set of scenarios was based on two million simulations for each model. We performed three additional million simulations for each of the best African Replacement, Multiregional and Assimilation models, to obtain their posterior probability based on a total of 5 millions simulations for each model. Overall, simulations took the equivalent of about 10 CPU-months.

Parameter estimation. For the best model within each set of scenarios, we retained the 5,000 simulations with smallest associated Euclidean distance δ computed on a total of 5 million simulations. Then posterior distributions of the parameters are obtained via a locally-weighted multivariate regression (see [20] for more details). Parameters (x) were transformed as $y = \log[\tan(x)^{-1}]$ before regression to prevent estimations to exceed distribution limits [52]. We performed a small study on the accuracy of several possible point estimators for the parameters (i.e. mean, median, mode, regression coefficient), from which we concluded that the median had overall the best properties (see Supporting information Table S4). This point estimator is therefore reported in Table 1, whereas Table S5 lists the median and the mode of the parameter posterior distributions estimated under the three best models.

TMRCA simulations. We generated for each model the expected distribution of the time to the most common ancestor (TMRCA) by performing 5,000 simulations of 50 loci, using as fixed parameter values the median estimates obtained under our ABC approach, which is a reasonably good point in the parameter space. We generated in the same way the distribution of TMRCAs for uniparentally inherited markers by dividing the population sizes by four since the effective size for these markers is four times less than for nuclear loci.

Acknowledgments

We are grateful to the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul and to Institutos do Milênio and Apoio a Núcleos de Excelência Programs for extra support, and to EGEE2 European project for providing access to its computing grid infrastructure and for user assistance. Thanks to Pierre Berthier for computational services,

to Johan Montagnat and Cladinara Sarturi for technical help, to Wen-Hsiung Li for the primers, and to Kim Hill, A. Magdalena Hurtado, Ramiro Barrantes, Francisco R. Carnese, and Eduardo Tarazona-Santos for sample donations, as well as to all individuals who, by contributing their own samples, made this study possible. We are also grateful to Montgomery Slatkin and Paul Mellars for their comments on a previous version of this manuscript.

Funding. This work has been supported by grants from: the University of Bern to NJRF; Swiss National Foundation No. 3100A0-112072 to LE; Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brazil) No. 477780/2003-2 to SLB. NJRF was partially supported by a CAPES (Brazil) scholarship No. 3624-05-6.

References

1. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, et al. (2005) A haplotype map of the human genome. Nature 437: 1299-1320.

2. Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM (2006) Genomic signatures of positive selection in humans and the limits of outlier approaches. Genome Res 16: 980-989.

3. Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, et al. (2005) Calibrating a coalescent simulation of human genome sequence variation. Genome Res 15: 1576-1583.

4. Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. Nat Genet 33 Suppl: 266-275.

5. Stringer C (2002) Modern human origins: progress and prospects. Phil Trans Roy Soc B 357: 563.

6. Mellars P (2006) Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. PNAS 103: 9381-9386.

7. Plagnol V, Wall JD (2006) Possible ancestral structure in human populations. PLoS Genet 2: e105.

8. Ray N, Currat M, Berthier P, Excoffier L (2005) Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. Genome Res 15: 1161-1167.

9. Marth G, Schuler G, Yeh R, Davenport R, Agarwala R, et al. (2003) Sequence variations in the public human genome data reflect a bottlenecked population history. PNAS 100: 376-381.

10. Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, et al. (1997) Neandertal DNA sequences and the origin of modern humans. Cell 90: 19-30.

11. Liu H, Prugnolle F, Manica A, Balloux F (2006) A Geographically Explicit Genetic Model of Worldwide Human-Settlement History. Am J Hum Genet 79: 230-237.

12. Templeton A (2002) Out of Africa again and again. Nature 416: 45-51.

13. Garrigan D, Hammer MF (2006) Reconstructing human origins in the genomic era. Nat Rev Genet 7: 669.

14. Templeton A (2005) Haplotype Trees and Modern Human Origins. American Journal of Physical Anthropology 128: 33-59.

15. Eswaran V, Harpending H, Rogers AR (2005) Genomics refutes an exclusively African origin of humans. Journal of Human Evolution 49: 1.

16. Hayakawa T, Aki I, Varki A, Satta Y, Takahata N (2006) Fixation of the Human-Specific CMP-N-Acetylneuraminic Acid Hydroxylase Pseudogene and Implications of Haplotype Diversity for Human Evolution. Genetics 172: 1139-1146.

17. Evans PD, Mekel-Bobrov N, Vallender EJ, Hudson RR, Lahn BT (2006) Evidence that the adaptive allele of the brain size gene microcephalin introgressed into Homo sapiens from an archaic Homo lineage. Proc Natl Acad Sci U S A 103: 18178-18183.

18. Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc Natl Acad Sci U S A 104: 2785-2790.

19. Marjoram P, Molitor J, Plagnol V, Tavare S (2003) Markov chain Monte Carlo without likelihoods. Proc Natl Acad Sci U S A 100: 15324-15328.

20. Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162: 2025-2035.

21. McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature 433: 733.

22. Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, et al. (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. PNAS 102: 18508-18513.

23. Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, et al. (2005) Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. Science 308: 1034-1036.

24. Dillehay T (1997) Monte Verde: A Late Pleistocene Settlement in Chile II: The Archaeological Context and Interpretation. Washington D. C.: Smithsonian Institution Press. 1071 p.

25. Bortolini M-C, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, et al. (2003) Y-Chromosome Evidence for Differing Ancient Demographic Histories in the Americas. Am J Hum Genet 73: 524-539.

26. Bonatto SL, Salzano FM (1997) Diversity and Age of the Four Major mtDNA Haplogroups, and Their Implications for the Peopling of the New World. Am J Hum Genet 61: 1413-1423.

27. Hey J (2005) On the Number of New World Founders: A Population Genetic Portrait of the Peopling of the Americas. PLoS Biology 3: e193.

28. Schurr TG (2004) The peopling of the New World: Perspectives from molecular anthropology. Annual Review Of Anthropology 33: 551-583.

29. Battilana J, Fagundes NJ, Heller AH, Goldani A, Freitas LB, et al. (2006) Alu insertion polymorphisms in Native Americans and related Asian populations. Ann Hum Biol 33: 142-160.

30. Takahata N, Lee SH, Satta Y (2001) Testing multiregionality of modern human origins. Mol Biol Evol 18: 172-183.

31. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306: 636-640.

32. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci U S A 102: 15942-15947.

33. Chen FC, Li WH (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 68: 444-456.

34. Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, et al. (2002) Larger genetic differences within africans than between Africans and Eurasians. Genetics 161: 269-274.

35. Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, et al. (2003) Low nucleotide diversity in chimpanzees and bonobos. Genetics 164: 1511-1518.

36. Greenberg JH (1987) Language in the Americas. Stanford: Stanford University Press. 438 p.

37. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8: 175-185.

38. Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8: 195-202.

39. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. (2002) Genetic structure of human populations. Science 298: 2381-2385.

40. Currat M, Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. Proc Biol Sci 272: 679-688.

41. Stringer C (2001) Modern Human Origins - Distinguishing the Models. Afr Arch Rev 18: 67-75.

42. Excoffier L, Estoup A, Cornuet J-M (2005) Bayesian Analysis of an Admixture Model With Mutations and Arbitrarily Linked Markers. Genetics 169: 1727-1738.

43. Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. Bioinformatics 20: 2485-2487.

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

45. Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, et al. (2005) Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet 1: e70.

46. Excoffier L, Laval G, Schneider S (2005) Arlequin3.0: an integrated software package for population genetics data analysis. EBO 1: 47-50.

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

48. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

49. Beaumont MA (2007) Joint determination of topology, divergence time, and immigration in population trees. In: Matsumura S, Forster P, Renfrew C, editors. Simulations, Genetics and Human Prehistory - A Focus on Islands. Cambridge: McDonald Institute Monographs.

50. Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, et al. (2005) Multiple transatlantic introductions of the western corn rootworm. Science 310: 992.

51. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16: 1791-1798.

52. Hamilton G, Stoneking M, Excoffier L (2005) Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. PNAS 102: 7476-7480.

Table 1. Demographic and historical parameters estimated under the favored African Replacement model with exponential growth (AFREG).

Parameters[1]	**Median**[2]	**95% HPD**[3]
Speciation time for modern human (years)	144,825	103,975 – 191,025
Exit out of Africa (years)	54,225	40,200 – 75,500
Colonization of the Americas (years)	9,350	7,600 – 13,375
Size of archaic African population	14,649	7,753 – 22,815
Bottleneck size during speciation	536	101 – 1,533
Bottleneck size when leaving Africa	317	57 – 917
Bottleneck size when leaving Asia	476	74 – 1,332
Current African population size	143,318	9,809 – 557,985
Current Asian population size	22,386	1,443 – 75,799
Current American population size	7,724	784 – 19,267

[1] Population sizes are given in effective number of diploid individuals.

[2] Median value of the marginal posterior density.

[3] 95% Highest Posterior Density interval.

The estimates were calibrated by assuming a human-chimpanzee divergence of 6 million years and a generation time of 25 years.

Figure legends

Figure 1. Alternative scenarios of human evolution. **A**. African Replacement model: it assumes that modern humans originated in Africa and colonized the rest of the world by completely replacing *H. erectus* in Asia. Population growth is modeled as an instantaneous demographic expansion having occurred right after each bottleneck (AFRIG), or by a continuous exponential demographic expansion (AFREG); **B**. Assimilation model: similar to the African Replacement model in **A**, but allowing for some archaic Asian lineages to have entered the modern gene pool by hybridization. As in **A**, population growth is modeled either as instantaneous (ASIG) or exponential (ASEG); **C**. Multiregional evolution model: it assumes that the transition between previous *Homo* and modern humans occurred simultaneously in Africa and Asia due to ongoing gene flow between these continents [5]. Alternative models under this scenario include one where archaic and modern population sizes are the same in each continent (MRE1S), one that allows an instantaneous transition between archaic and modern population sizes (MRE2S), and two that assume a bottleneck in Africa followed by either instantaneous (MREBIG) or exponential (MREBEG) growth . For all scenarios, the dark colors represent modern human populations, while lighter colors represent archaic populations. AF: Africa; AS: Asia; AM: Americas. A more detailed description of these scenarios is provided in the Material and Methods section, and in Supporting Information Fig. S1. The posterior probability of different models within each major scenario is given below each model. The posterior probabilities of the best model selected under each scenario are reported within boxes.

Figure 2. Posterior (thick line) and prior (thin line) distributions of the estimated parameters of the AFREG model. Given their low R^2 values (<0.01, see Table S5), the duration of the bottlenecks were considered as nuisance parameters. Parameter labels (x axis) correspond to those shown in Supporting Information Fig. S1 and are described in the text.

Figure 3. Empirical TMRCA distribution obtained by simulation under different models. Parameter values were set to the median of the estimated marginal posterior distributions. Each distribution combines a mirrored estimated density surface in grey with a standard boxplot representation. Boxplots display the median of the distribution as a white dot, the interquartile range (IQR, 25%-75%) as a thick line, and the region of ± 1.5 IQR as a thin line ending with vertical whiskers. To facilitate the comparison among models, all distributions (apart those from MREBIG model) were cut after the 99th percentile (full distributions are available in Tables S5, S6 and S7). **A.** Autosomal loci **B.** mtDNA and Y-chromosome. For these markers, simulations were performed by using estimates of effective sizes four times smaller than those obtained for autosomal loci, to reflect the smaller population sizes of these uniparentaly inherited markers. **C**. Autosomal loci under the best model (AFREG) where only the samples of each of the three regions are considered.

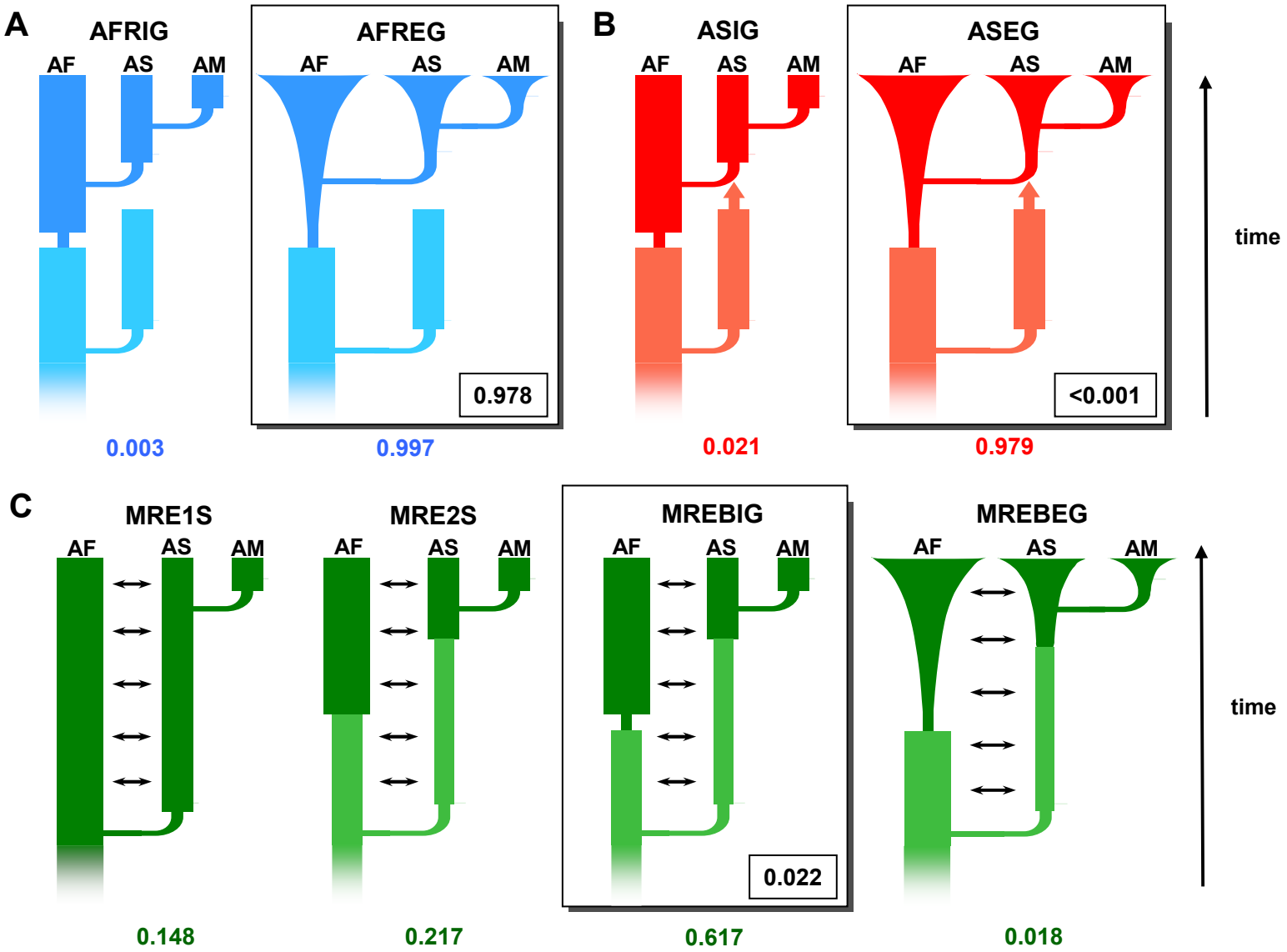

A
AFRIG
AF
AS
AM
0.003
AFREG
AF
AS
AM
0.978
0.997
B
ASIG
AF
AS
AM
0.021
ASEG
AF
AS
AM
<0.001
0.979
time
C
MRE1S
AF
AS
AM
0.148
MRE2S
AF
AS
AM
0.217
MREBIG
AF
AS
AM
0.022
0.617
MREBEG
AF
AS
AM
0.018
time

Figure 1

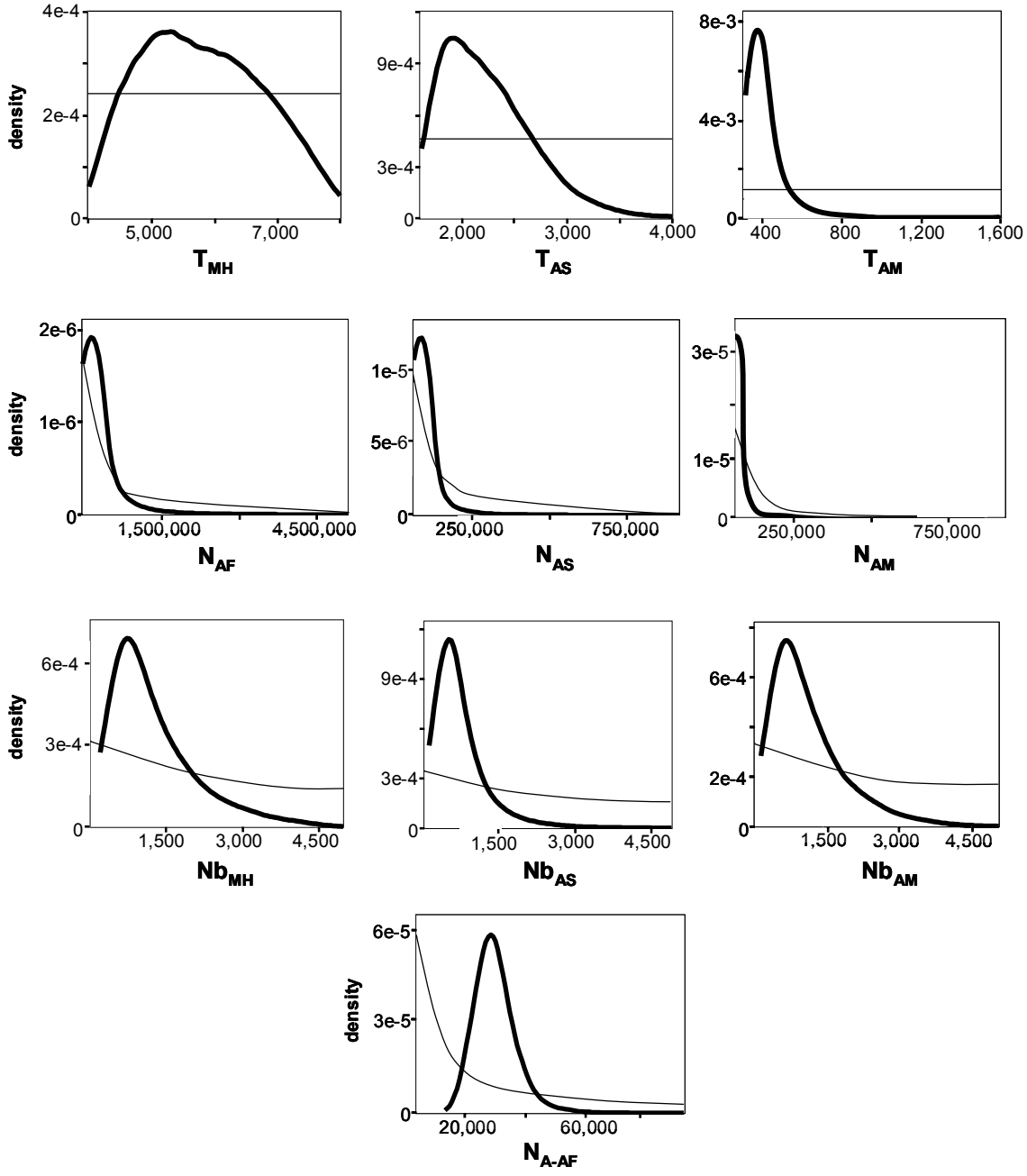

density
4e-4
2e-4
0
5,000
7,000
T_{MH}
9e-4
3e-4
0
2,000
3,000
4,000
T_{AS}
8e-3
4e-3
0
400
800
1,200
1,600
T_{AM}
2e-6
1e-6
0
1,500,000
4,500,000
N_{AF}
1e-5
5e-6
0
250,000
750,000
N_{AS}
3e-5
1e-5
0
250,000
750,000
N_{AM}
6e-4
3e-4
0
1,500
3,000
4,500
Nb_{MH}
9e-4
3e-4
0
1,500
3,000
4,500
Nb_{AS}
6e-4
2e-4
0
1,500
3,000
4,500
Nb_{AM}
6e-5
3e-5
0
20,000
60,000
$N_{A\text{-}AF}$

Figure 2.

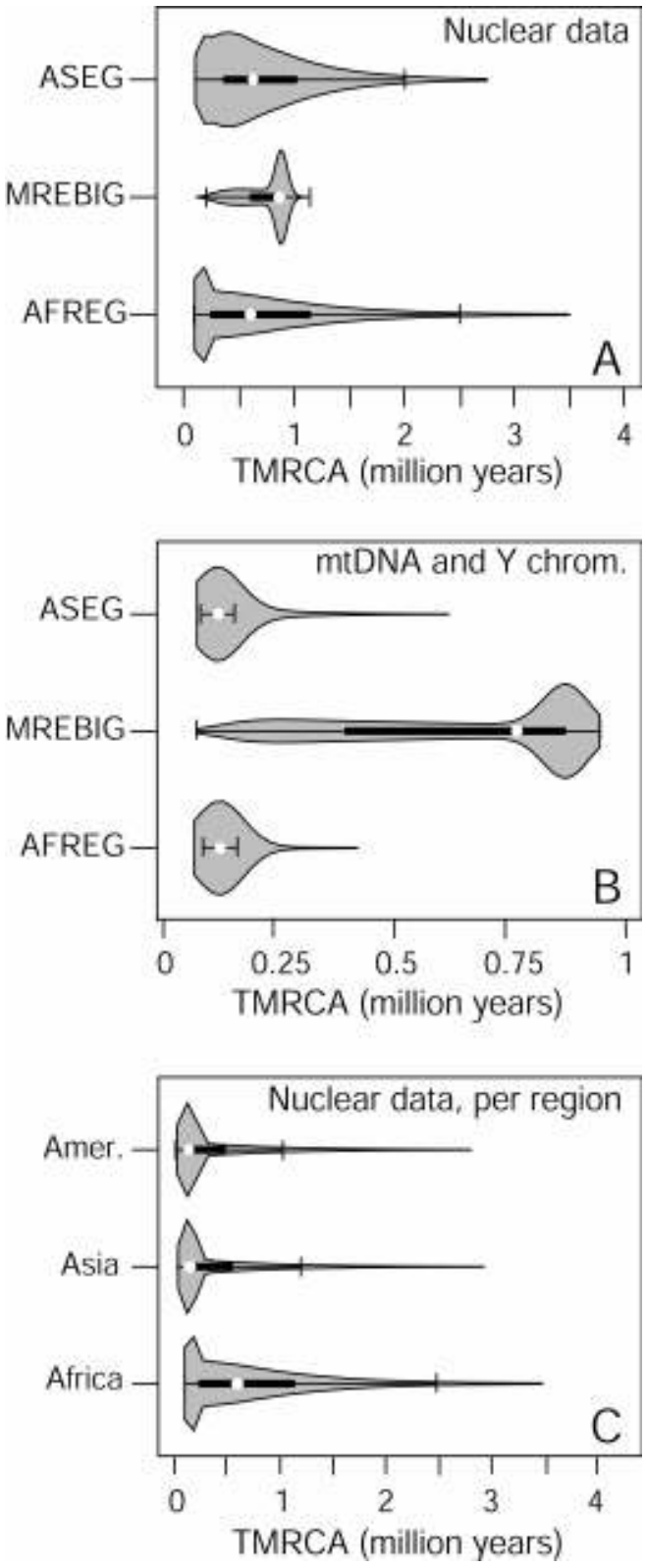

Nuclear data
ASEG
MREBIG
AFREG
A
0
1
2
3
4
TMRCA (million years)
mtDNA and Y chrom.
ASEG
MREBIG
AFREG
B
0
0.25
0.5
0.75
1
TMRCA (million years)
Nuclear data, per region
Amer.
Asia
Africa
C
0
1
2
3
4
TMRCA (million years)

Figure 3

Supporting Information

2 Figures

8 Tables

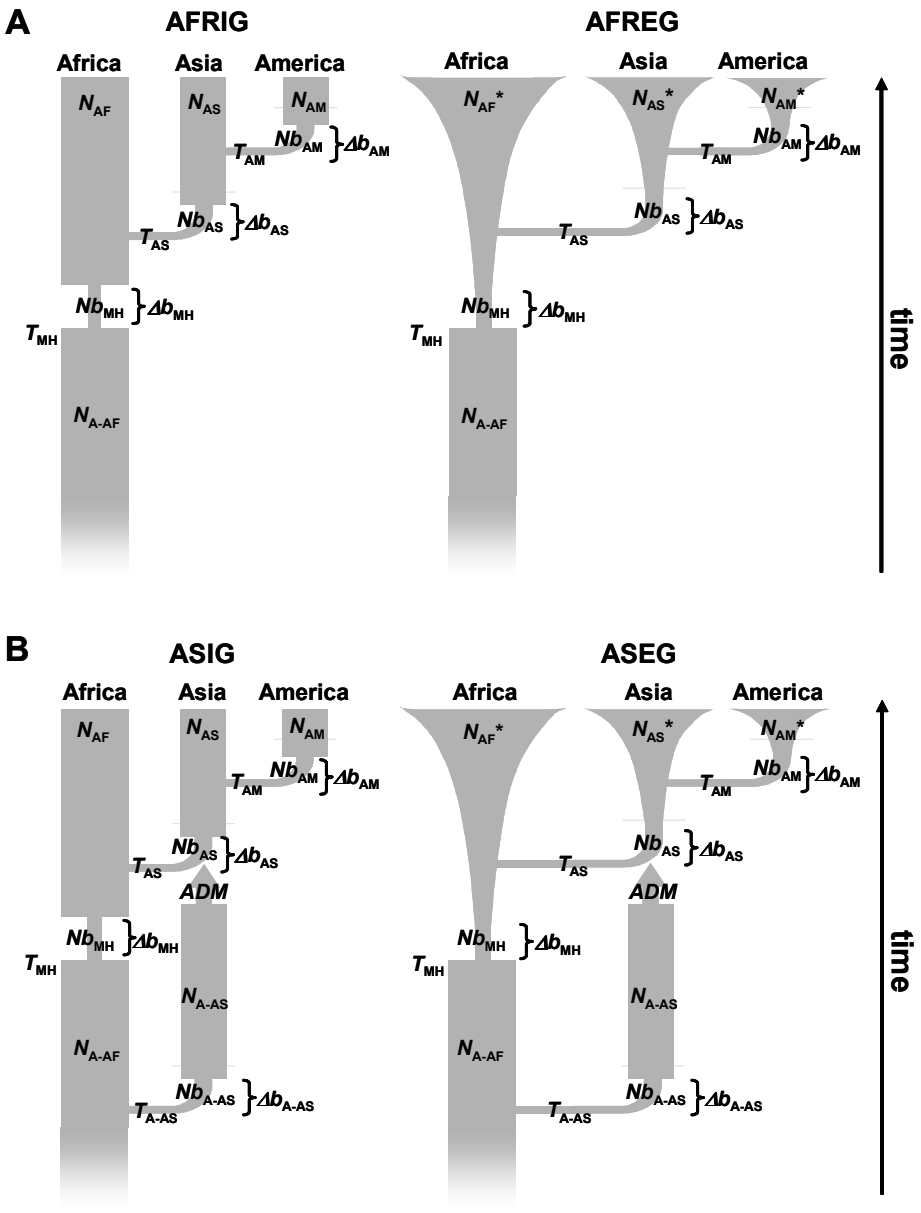


Figure S1. Graphical presentation of the eight alternative models of human evolution tested in our study, and their associated sets of parameters. In parameter acronyms, "N" represents population sizes, "T" represents the timing of some events, "M" represents migration rates, and "Δb" is for the duration of a bottleneck period. A. African replacement models with instantaneous (AFRIG) or exponential growth (AFREG). B. African origin with assimilation models with instantaneous (ASIG) or exponential growth (ASEG). (figure legend continues on the next page)

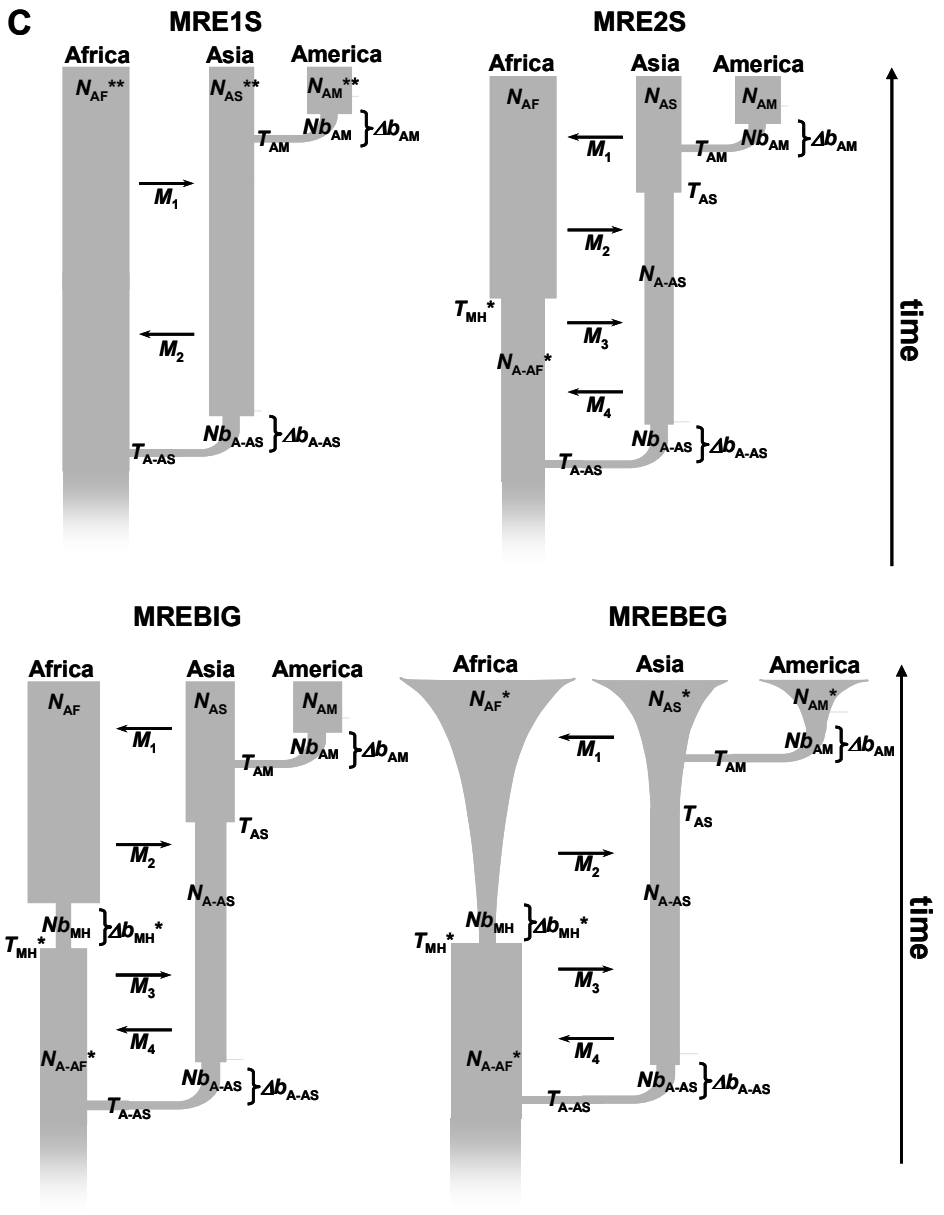


Figure S1 (continued). C. Multiregional evolution models with a single population size for archaic and modern populations (MRE1S), two populations sizes related to archaic and modern populations (MRE2S), a bottleneck in Africa and instantaneous growth for modern populations (MREBIG), and with a bottleneck in Africa and exponential growth for modern populations (MREBEG). See text and material and methods for further justification of these models. Values for these parameters are shown in table S2.

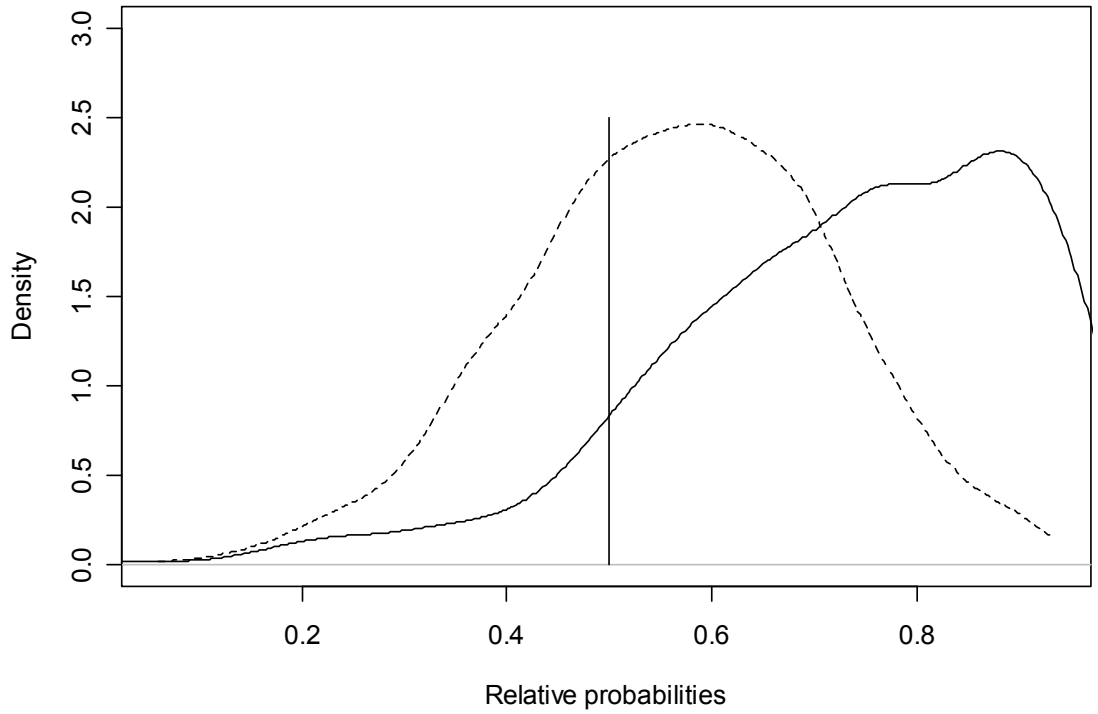


Figure S2. Empirical distributions of the estimated relative probabilities of the AFREG (solid line) and MREBIG (dashed line) models when they are the true models. In order to check that the probability associated to one model was not due to some bias in the model selection procedure, we simulated 1,000 datasets under the AFREG and MREBIG model, taking as parameter values the median estimates obtained under our ABC approach. The relative probability of the AFREG and MREBIG models were then estimated under our model selection procedure, as explained in the Material and Methods section. The area under the curve on the right of the vertical line represents the fraction of times the true model is recovered (relative probability >0.5) by our estimation procedure, which is 90.4% for the AFREG model and 66.8% for the MREBIG model. If there was no information in the data the distribution would be strongly clustered around our prior of 0.5.

Table S1. Evolutionary model and mutation rate estimated for each locus.

Locus	Evolutionary model	Mutation rate (per site per year)
T2557	TIM + I	1.34×10^{-9}
T2191	GTR + I	1.83×10^{-9}
T2568	HKY	6.84×10^{-10}
T1636	K81uf + I	2.60×10^{-9}
T2020	HKY	1.20×10^{-9}
T1584	TVM	1.75×10^{-9}
T2019	HKY	9.18×10^{-10}
T2568	HKY	9.22×10^{-10}
T953	K81uf + I	5.17×10^{-10}
T2021	TrN	1.61×10^{-9}
T2472	TVM + I	1.21×10^{-9}
T1469	K80	2.75×10^{-10}
T151	HKY	1.09×10^{-9}
T1364	F81	7.08×10^{-10}
T2609	GTR + I	1.54×10^{-9}
T2659	HKY + I	7.46×10^{-10}
T1251	F81	1.60×10^{-9}
T24894	HKY	1.39×10^{-9}
T2041	F81 + I	1.33×10^{-9}
T2294	K80	9.48×10^{-10}
T2984	TIM + I	1.74×10^{-9}
T10604	HKY	6.17×10^{-10}
T812	TrN	1.13×10^{-9}
T2920	TVM	1.12×10^{-9}
T2012	TVM	7.25×10^{-10}
T784	K81uf + I	1.76×10^{-9}
T787	K81uf	1.03×10^{-9}
T813	K81uf	7.22×10^{-10}
T2085	TVMef + I	1.05×10^{-10}
T2064	TrN	4.11×10^{-10}
T2352	TrN + I	1.01×10^{-10}
T2560	HKY	3.79×10^{-10}
T1412	HKY	9.16×10^{-10}
T1419	TrN + I	9.38×10^{-10}
T1482	TrN + I	1.25×10^{-10}

T2963	TrN	2.86×10^{-10}
T2265	HKY + I	9.49×10^{-10}
T2266	HKY	7.78×10^{-10}
T2558	HKY	8.44×10^{-10}
T2906	HKY	3.06×10^{-10}
T2987	JC	9.17×10^{-10}
T946	HKY	1.26×10^{-9}
T2988	HKY	1.18×10^{-9}
T866	K81uf	1.37×10^{-9}
T1506	F81	4.46×10^{-10}
T2563	HKY + I	1.31×10^{-9}
T2018	HKY	1.45×10^{-9}
T2924	HKY	1.62×10^{-9}
T1386	TrN	1.74×10^{-9}
T864	HKY	1.59×10^{-9}
Average	**-**	**1.10×10^{-9}**

Mutation rates were estimated from DNA sequences inferred with PHASE 2.0 [1] and an expectation maximization algorithm [2], two phasing methods shown as being accurate [3]. For 30 human and 16 chimpanzee loci, the phase inference was trivial since no individual was heterozygous for more than one site. Whenever the two phasing methods resulted in different estimates (which occurred for 15 human and 16 chimpanzee loci) the method suggesting the smallest number of different haplotypes was preferred. Human and chimpanzee haplotypes were then aligned, and the best sequence evolutionary model was selected using ModelTest [4]. The mutation rate for each locus was estimated using the average genetic distance between human and chimpanzee haplotypes under the most likely evolutionary model, and by assuming a generation time of 25 years and a divergence time of 6 million years [5] between the two species. The full names, references, and parameters for the evolutionary models can be found in the ModelTest manual (http://darwin.uvigo.es/software/modeltest.html).

Table S2. Prior distributions for the parameters of the tested evolutionary models. Relevant parameters for a given model are indicated with crosses (X)

Parameter	Model								Values		Distribution
	AFRIG	AFREG	ASIG	ASEG	MRE1S	MRE2S	MBIG	MBEG	Minimum	Maximum	
Effective population sizes											
Current size in Africa N_{AF}	**X**		**X**			**X**	**X**		5,000	1,000,000	Log-uniform
Current size in Asia N_{AS}	**X**		**X**			**X**	**X**		1,000	100,000	Log-uniform
Current size in America N_{AM}	**X**		**X**			**X**	**X**		1,000	100,000	Log-uniform
Current size in Africa N_{AF}*		**X**		**X**				**X**	5,000	5,000,000	Log-uniform
Current size in Asia N_{AS}*		**X**		**X**				**X**	1,000	1,000,000	Log-uniform
Current size in America N_{AM}*		**X**		**X**				**X**	1,000	1,000,000	Log-uniform
Current size in Africa N_{AF}**					**X**				100	1,000,000	Log-uniform
Current size in Asia N_{AS}**					**X**				100	100,00	Log-uniform
Current size in America N_{AM}**					**X**				100	100,000	Log-uniform
Archaic size in Africa $N_{A\text{-}AF}$	**X**	**X**	**X**	**X**					1,000	100,000	Log-uniform
Archaic size in Africa $N_{A\text{-}AF}$*						**X**	**X**	**X**	100	10,000	Log-uniform
Archaic size in Asia $N_{A\text{-}AS}$	**X**	**X**	**X**	**X**		**X**	**X**	**X**	100	10,000	Log-uniform
Size during the peopling of the Americas Nb_{AM}	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	2	5,000	Uniform
Size during modern out-of-Africa Nb_{AS}	**X**	**X**	**X**	**X**					2	5,000	Uniform

Size during modern human speciation Nb_{MH}	**X**	**X**	**X**	**X**			**X**	**X**	2	5,000	Uniform
Size during *H. erectus* out-of-Africa $Nb_{A\text{-}AS}$			**X**	**X**	**X**	**X**	**X**	**X**	2	5,000	Uniform
Forward in time migration rates											
Current migration rate from Africa to Asia M_1					**X**	**X**	**X**	**X**	10^{-7}	10^{-3}	Log-uniform
Current migration rate from Asia to Africa M_2					**X**	**X**	**X**	**X**	10^{-7}	10^{-3}	Log-uniform
Ancient migration rate from Africa to Asia M_3						**X**	**X**	**X**	10^{-7}	10^{-3}	Log-uniform
Ancient migration rate from Asia to Africa M_4						**X**	**X**	**X**	10^{-7}	10^{-3}	Log-uniform
Timing											
Time for the peopling of the Americas T_{AM}	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	300	1,600	Uniform
Time for modern out-of-Africa T_{AS}	**X**	**X**	**X**	**X**					1,600	4,000	Uniform
Time for population increase in Asia T_{AS}*						**X**	**X**	**X**	1,600	4,000	Uniform
Time for modern speciation T_{MH}	**X**	**X**	**X**	**X**					4,000	8,000	Uniform
Time for modern speciation T_{MH}*						**X**	**X**	**X**	1,600	8,000	Uniform
Time for H. erectus out-of-Africa $T_{A\text{-}AS}$			**X**	**X**	**X**	**X**	**X**	**X**	32,000	40,000	Uniform
Duration of the peopling of the Americas bottleneck Δb_{AM}	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	1	50	Uniform
Duration of the modern out-of-Africa bottleneck Δb_{AS}	**X**	**X**	**X**	**X**					1	50	Uniform
Duration of the modern speciation bottleneck Δb_{MH}	**X**	**X**	**X**	**X**					1	500	Uniform
Duration of the modern speciation bottleneck Δb_{MH}*							**X**	**X**	1	50	Uniform
Duration of the *H. erectus* out-of-Africa bottleneck Δb_{HE}			**X**	**X**	**X**	**X**	**X**	**X**	1	50	Uniform
Admixture level											
Proportion of archaic genes in current Asian population ADM			**X**	**X**					0.00	1.00	Uniform

Notes: Population sizes are in number of chromosomes, Times are in number of generations.

Table S3: Summary of the genetic diversity found at the 50 studied loci.

Locus	*L*	Sample size (individuals)			Number of segregating sites (*S*)			
		Africa	Asia	América	Africa	Asia	America	Total
T2557	431	10	8	12	3	2	3	4
T2191	405	10	8	12	1	0	0	1
T2568	439	10	8	12	1	2	2	3
T1636	506	10	8	12	8	0	0	8
T2020	488	10	8	12	1	1	1	1
T1584	563	10	8	11	1	0	0	1
T2019	442	10	8	12	1	1	1	1
T2568	439	10	8	11	1	1	0	1
T953	644	10	8	12	0	0	0	0
T2021	477	10	8	11	4	2	1	5
T2472	416	10	8	11	1	0	0	1
T1469	442	10	8	12	1	1	1	2
T151	412	10	8	12	2	1	1	2
T1364	417	10	8	12	0	0	0	0
T2609	452	10	8	12	2	0	0	2
T2659	455	10	8	12	2	1	1	3
T1251	529	10	8	12	6	0	0	6
T24894	476	10	8	12	3	0	0	3
T2041	429	10	8	9	0	0	0	0
T2294	700	10	8	12	5	5	4	7
T2984	478	10	8	12	3	0	0	3
T10604	511	10	8	12	2	2	0	3
T812	653	10	8	11	2	1	0	3
T2920	508	10	8	9	3	2	2	3
T2012	525	10	8	12	0	0	0	0
T784	541	10	8	12	3	4	1	6
T787	442	10	8	12	1	1	1	1
T813	528	10	8	10	1	1	1	1
T2085	577	10	8	11	3	0	0	3
T2064	522	10	8	11	0	0	2	2
T2352	558	10	8	12	4	1	1	4
T2560	499	10	8	12	2	1	0	3
T1412	543	10	8	11	4	2	2	4
T1419	452	10	8	12	2	0	0	2
T1482	479	10	8	11	2	2	1	3
T2963	432	10	8	11	3	0	0	3

T2265	453	10	8	12	3	0	0	3
T2266	477	10	8	12	2	0	0	2
T2558	509	10	8	12	4	2	2	4
T2906	585	10	8	12	1	0	0	1
T2987	316	10	8	10	2	1	1	2
T946	450	10	8	11	1	0	0	1
T2988	581	10	8	12	5	3	1	6
T866	349	10	8	12	1	1	0	2
T1506	476	10	8	12	0	1	1	1
T2563	526	10	8	10	1	0	0	1
T2018	522	10	8	11	2	1	0	3
T2924	460	10	8	10	3	3	3	3
T1386	418	10	8	11	4	2	1	5
T864	493	10	8	12	7	3	1	8
Overall	24425	-	-	-	114	51	36	137

Notes: *L*, length of the DNA sequence analyzed for all individuals. The four monomorphic loci are highlighted in grey shade.

Table S3: continued

π %			Tajima's *D*			F_{ST}			
Africa	Asia	America	África	Asia	America	Africa x Asia	Africa x America	Asia x America	Total
0.305	0.130	0.219	1.485	-0.189	0.434	0.153	0.068	-0.019	0.073
0.047	0.000	0.000	-0.592	NP	NP	0.037	0.066	NP	0.053
0.061	0.103	0.157	-0.086	-0.649	0.627	0.107	0.148	-0.016	0.082
0.583	0.000	0.000	1.035	NP	NP	0.235	0.288	0.000	0.266
0.039	0.082	0.095	-0.592	0.650	1.232	0.023	0.103	-0.038	0.034
0.018	0.000	0.000	-1.164	NP	NP	-0.012	0.005	NP	-0.003
0.043	0.053	0.105	-0.592	-0.448	1.232	-0.056	0.474	0.424	0.402
0.023	0.053	0.000	-1.164	-0.448	NP	-0.021	0.005	0.096	0.024
0.000	0.000	0.000	NP	NP	NP	NP	NP	NP	NP
0.156	0.159	0.106	-0.989	0.661	**1.471**	-0.001	0.092	0.010	0.039
0.065	0.000	0.000	-0.086	NP	NP	0.086	0.114	NP	0.102
0.043	0.074	0.078	-0.592	0.156	0.480	0.111	0.130	-0.054	0.063
0.089	0.111	0.095	-0.812	1.034	0.776	0.014	-0.013	-0.045	-0.016
0.000	0.000	0.000	NP	NP	NP	NP	NP	NP	NP
0.064	0.000	0.000	-1.141	NP	NP	0.021	0.047	NP	0.036
0.044	0.028	0.102	**-1.513**	-1.162	1.232	-0.003	0.226	0.145	0.170
0.211	0.000	0.000	-1.084	NP	NP	0.053	0.085	NP	0.071
0.063	0.000	0.000	**-1.723**	NP	NP	-0.012	0.009	NP	-0.001
0.000	0.000	0.000	NP	NP	NP	NP	NP	NP	NP
0.123	0.199	0.245	-1.197	-0.252	**1.641**	0.026	0.309	0.124	0.193
0.098	0.000	0.000	-1.191	NP	NP	0.047	0.077	NP	0.064
0.118	0.088	0.000	0.173	-0.649	NP	0.038	0.354	0.144	0.192
0.031	0.019	0.000	**-1.513**	-1.162	NP	-0.003	0.005	0.021	0.004
0.219	0.092	0.202	0.837	-0.578	1.866	0.408	-0.029	0.330	0.245
0.000	0.000	0.000	NP	NP	NP	NP	NP	NP	NP
0.110	0.154	0.094	-0.792	-0.966	1.505	0.568	0.617	-0.117	0.492
0.076	0.113	0.115	0.352	1.309	**1.505**	0.018	0.060	-0.051	0.011
0.099	0.101	0.100	**1.531**	**1.529**	**1.505**	-0.054	-0.047	-0.060	-0.053
0.263	0.000	0.000	**2.117**	NP	NP	0.338	0.382	NP	0.365
0.000	0.000	0.077	NP	NP	-0.603	0.000	0.106	0.087	0.100
0.223	0.058	0.062	0.294	0.156	0.480	0.074	0.089	-0.054	0.062
0.109	0.25	0.000	-0.090	-1.162	NP	0.176	0.249	0.026	0.202
0.207	0.103	0.113	-0.016	-0.189	0.273	0.120	0.110	-0.054	0.079
0.064	0.000	0.000	-1.141	NP	NP	0.021	0.047	NP	0.036
0.110	0.075	0.065	-0.156	-1.038	0.237	-0.003	-0.007	-0.034	-0.023

0.069	0.000	0.000	**-1.723**	NP	NP	-0.012	0.005	NP	-0.003
0.086	0.000	0.000	-1.441	NP	NP	0.013	0.038	NP	0.027
0.127	0.000	0.000	0.173	NP	NP	0.298	0.354	NP	0.330
0.236	0.197	0.154	0.186	**1.687**	1.014	0.042	0.011	-0.016	0.013
0.082	0.000	0.000	1.262	NP	NP	0.286	0.342	NP	0.319
0.157	0.166	0.060	-0.287	**1.474**	-0.592	0.123	0.544	0.220	0.346
0.042	0.000	0.000	-0.592	NP	NP	0.037	0.060	NP	0.050
0.280	0.176	0.067	0.471	0.388	0.776	0.024	0.134	0.070	0.080
0.029	0.067	0.000	-1.164	-0.448	NP	0.055	0.009	0.105	0.061
0.000	0.026	0.018	NP	-1.162	-1.159	0.014	-0.008	-0.050	-0.023
0.036	0.000	0.000	-0.592	NP	NP	0.037	0.053	NP	0.046
0.104	0.024	0.000	-0.090	-1.162	NP	0.176	0.238	0.021	0.196
0.259	0.286	0.248	1.086	1.323	0.936	-0.017	0.037	0.065	0.029
0.286	0.060	0.121	0.173	-1.498	**1.471**	0.117	0.028	0.204	0.096
0.246	0.213	0.070	-1.271	0.468	0.480	0.027	0.125	0.122	0.088
0.115	0.061	0.055	-0.452	-0.141	0.833	0.139	0.182	0.075	0.139

Note: π, nucleotide diversity; Tajima's *D* values where *P*-values are smaller than 0.05; are shown in bold; NP, non polymorphic loci

Choice of point estimates. Because several point estimates can be computed on posterior distributions (e.g. mean, median, mode, regression coefficient) obtained under the ABC approach, we performed a small accuracy study to define which estimator performed best (e.g. had the smallest associated relative Root Mean Square Error, or relative RMSE). The principle is to use test-datasets, from which we know the true values, to get estimated values and subsequently to assess the quality of the estimation by comparing the estimates with the true values. We used 1,000 simulations based on arbitrarily fixed values for all parameters, and we used our original 5 million simulated datasets to re-estimate the parameters using the ABC procedure mentioned above. From the set of 1,000 estimated parameters, we then evaluated the relative bias, the relative Root Mean Square Error (RMSE) as well as the index *Factor-2,* defined as the proportion of simulations whose estimated values were within an interval defined as 50%-200% of the true value [6]. The results of our test on the accuracy of different point estimators are reported in Table S4 below. They show that the median and the mode of the posterior distributions have overall the best properties. For simplicity, we have only reported the median estimator in Table 1, since this was the estimator upon which we based the TMRCA simulations and the evaluation of the power of our model-selection approach. However in Supporting information Table S5 we show both estimators for the best model in each set.

Table S4. Results of accuracy tests for the AFREG model on four point estimators

Parameter	True value	Median				Mode				Mean				Regression			
		Ave. Est.	*Bias*	*RMSE*	*F2*	*Ave. Est.*	*Bias*	*RMSE*	*F2*	*Ave. Est.*	*Bias*	*RMSE*	*F2*	*Ave. Est.*	*Bias*	*RMSE*	*F2*
T_{AM}	400	469	0.173	0.274	0.995	412	0.031	0.170	0.995	504	0.259	0.343	0.992	602	0.504	0.623	0.891
T_{AS}	2,300	2,182	-0.051	0.109	1	1,995	-0.133	0.171	1	2,260	-0.017	0.091	1	2,643	0.149	0.206	1
T_{MH}	6,200	5,108	-0.176	0.178	1	4,637	-0.252	0.253	1	5,311	-0.143	0.145	1	6,094	-0.017	0.033	1
N_{AF}*	150,000	151,902	0.013	0.587	0.757	99,457	-0.337	0.546	0.469	190,512	0.270	0.848	0.719	210,267	0.402	1.234	0.618
N_{AS}*	30,000	22,984	-0.234	0.536	0.629	13,226	-0.559	0.655	0.249	28,766	-0.041	0.665	0.671	27,447	-0.085	0.901	0.494
Nb_{MH}	1,000	1,388	0.388	0.447	0.999	898	-0.102	0.181	0.992	1,594	0.594	0.632	0.99	2,662	1.662	1.696	0.045
Nb_{AS}	600	724	0.207	0.553	0.878	553	-0.078	0.371	0.914	839	0.399	0.692	0.82	1,274	1.123	1.456	0.498
Nb_{AM}	800	1,039	0.299	0.507	0.934	696	-0.131	0.311	0.956	1,216	0.520	0.670	0.86	1,949	1.436	1.607	0.306
$N_{A\text{-}AF}$	30,000	13,936	-0.535	0.566	0.37	13,070	-0.564	0.590	0.318	14,858	-0.505	0.539	0.432	20,464	-0.318	0.411	0.732
Means over parameters				***0.417***	***0.840***			***0.361***	***0.766***			***0.514***	***0.832***			***0.907***	***0.620***

Ave. Est.: average estimated value; *Bias*: Relative Bias; *RMSE*: Relative Root Mean Square Error; *F2*: Factor-2. See Material and Methods (under *Choice of point estimates*) for the definition of these statistics. Population sizes are expressed in chromosomes, while times are given in generations. Due to its large RMSE (>2.0 for all estimators), the N_{AM}* parameter was omitted here in order to better discriminate means over parameters among the estimators.

Table S5. Median and mode estimates of all demographic parameters of the best models under each evolutionary scenario.

Parameters	R^2	Median	Mode	95% HPD
AFREG Model				
N_{AF}*	0.75	143,318	85,806	9,809 – 557,985
N_{AS}*	0.70	22,386	16,165	1,443 – 75,799
N_{AM}*	0.51	7,724	534	784 – 19,267
$N_{A\text{-}AF}$	0.57	14,649	14,293	7,753 – 22,815
Nb_{AM}	0.44	476	342	74 – 1,533
Nb_{AS}	0.46	317	253	57 – 917
Nb_{MH}	0.45	536	367	101 – 1,332
T_{AM}	0.22	9,350	8,875	7,600 – 13,375
T_{AS}	0.31	54,225	47,550	40,200 – 75,500
T_{MH}	0.09	144,825	133,375	103,975 – 191,025
ASEG Model				
N_{AF}*	0.71	182,238	118,752	18,160 – 690,420
N_{AS}*	0.73	16,108	11,059	1,347 – 50,156
N_{AM}*	0.47	11,663	5,505	773 – 32,651
$N_{A\text{-}AF}$	0.75	10,392	10,365	5,019 – 16,683
$N_{A\text{-}AS}$	0.05	341	154	52 – 1,720
Nb_{AM}	0.46	493	336	67 – 1,377
Nb_{AS}	0.36	516	388	98 – 1,318
Nb_{MH}	0.36	764	516	108 – 1,796
$Nb_{A\text{-}AS}$	0.03	124	51	2 – 621
T_{AM}	0.20	10,639	9,601	7,571 – 17,704
T_{AS}	0.12	54,415	49,498	40,265 – 77,926
T_{MH}	0.05	131,518	120,351	100,134 – 178,105
$T_{A\text{-}AS}$	0.01	833,880	816,420	800,233 – 921,080
ADM	0.67	0.005	0.003	0.000 – 0.015
MREBIG Model				
N_{AF}	0.64	33,963	24,649	5,075 – 130,942
N_{AS}	0.80	3,430	3,601	819 – 5,549
N_{AM}	0.64	1,277	985	525 – 2,506
$N_{A\text{-}AF}$*	0.19	504	248	52 – 2,866
$N_{A\text{-}AS}$	0.05	211	99	52 – 727
Nb_{AM}	0.25	589	259	8 – 1,649
Nb_{MH}	0.14	207	101	2 – 1,170

$Nb_{A\text{-}AS}$	0.04	103	51	2 – 438
$^{1}NM_{1}$	0.34	0.558	0.000	0.042 – 87.309
$^{2}NM_{2}$	0.55	0.563	0.630	0.003 – 1.265
$^{3}NM_{2}$*	0.41	0.039	0.017	0.001 – 0.177
$^{4}NM_{3}$	0.16	0.024	0.000	0.002 – 4.007
$^{5}NM_{4}$	0.07	0.017	0.000	0.002 – 2.394
T_{AM}	0.18	8,850	9,200	7,541 – 12,779
T_{AS}*	0.04	46,150	61,375	40,129 – 89,477
T_{MH}*	0.17	88,843	96,710	53,712 – 145,257
$T_{A\text{-}AS}$	0.01	820,630	865,168	800,453 – 960,078

Notes: Parameters labels are according to Fig. S1 and Table S3, except when indicated. Population size units are effective number of diploid individuals, times are in years, assuming 25y per generation. Parameters not presented in table S2 includes $^{1}NM_{1}$: $N_{AF} \times M_{1}$; $^{2}NM_{2}$: $N_{AS} \times M_{2}$; $^{3}NM_{2}$*: $N_{A\text{-}AS} \times M_{2}$; $^{4}NM_{3}$: $N_{A\text{-}AF} \times M_{3}$; $^{5}NM_{4}$: $N_{A\text{-}AS} \times M_{4}$. Additionally, for each parameter we report its multiple determination coefficient R^2 by summary statistics, as previous studies have shown that parameters with R^2>0.10 can be reasonably well estimated [7].

Table S6. Distribution of simulated TMRCA, in years, for autosomal loci under each scenario simulated on the basis of median estimates. Model labels as in fig. S1.

	Model		
	AFREG	ASEG	MREBIG
Quantiles for the TMRCA distribution			
0.00	97,225	10,3075	116,700
0.05	135,600	131,375	327,825
0.10	140,175	202,648	404,550
0.15	143,275	260,725	472,625
0.20	176,170	313,370	538,800
0.25	246,425	364,550	606,425
0.30	315,300	414,200	675,693
0.35	386,050	464,000	749,550
0.40	456,375	515,350	828,990
0.45	531,153	569,325	866,450
0.50	610,363	628,300	868,900
0.55	695,625	691,725	871,575
0.60	787,725	761,625	874,625
0.65	891,300	838,175	878,000
0.70	1,011,550	924,108	881,925
0.75	1,146,644	1,023,000	886,475
0.80	1,313,035	1,142,750	892,075
0.85	1,524,879	1,296,575	899,304
0.90	1,824,075	1,509,325	909,625
0.95	2,330,378	1,869,876	927,250
1.00	8,777,750	8,306,850	1,136,825
Average TMRCA over all loci	828,361	770,215	748,206
% of loci where MRCA is located in Africa	100.0	100.0	66.2
% of loci where MRCA is located in Asia	0.0	0.0	33.8
% of loci where MRCA is located in America	0.0	0.0	0.0

Table S7. Distribution of simulated TMRCA, in years, for uniparentally inherited loci under each scenario based on the median estimates. Model labels as in fig. S1.

	Model		
	AFREG	ASEG	MREBIG
Quantiles for the TMRCA distribution			
0.00	66,775	72,225	72,125
0.05	100,825	99,150	176,075
0.10	105,225	103,175	229,923
0.15	108,300	105,900	270,975
0.20	110,725	108,175	321,275
0.25	112,875	110,125	381,450
0.30	114,850	111,950	447,650
0.35	116,700	113,625	516,675
0.40	118,475	115,225	589,600
0.45	120,175	116,825	652,925
0.50	121,875	118,375	729,788
0.55	123,600	119,925	826,900
0.60	125,375	121,525	865,650
0.65	127,175	123,225	866,475
0.70	129,075	124,975	867,425
0.75	131,150	126,900	868,550
0.80	133,475	129,100	869,975
0.85	136,079	137,350	871,750
0.90	139,425	192,275	874,250
0.95	144,200	292,226	878,525
1.00	2,085,800	1,961,150	943,725
Average TMRCA over all loci	130,278	142,367	637,909
% of loci where MRCA is located in Africa	100.0	100.0	57.6
% of loci where MRCA is located in Asia	0.0	0.0	42.4
% of loci where MRCA is located in America	0.0	0.0	0.0

Table S8. Distribution of simulated continent specific TMRCA, in years, for autosomal loci, based on the median estimates of the best scenario (AFREG).

	Continent		
	Africa	Asia	Americas
Quantiles for the TMRCA distribution			
0.00	95,700	32,000	21,550
0.05	135,075	52,000	50,125
0.10	139,625	60,925	52,825
0.15	142,825	95,150	70,046
0.20	164,795	108,325	96,300
0.25	234,450	116,650	108,475
0.30	303,850	122,893	116,700
0.35	372,500	127,925	123,025
0.40	444,300	132,225	128,125
0.45	518,425	136,100	132,550
0.50	597,525	139,600	136,500
0.55	682,050	142,950	140,225
0.60	775,200	185,275	143,775
0.65	877,700	288,850	219,050
0.70	994,183	408,150	336,500
0.75	1,132,400	543,225	474,531
0.80	1,297,575	709,110	640,725
0.85	1,508,383	924,200	854,950
0.90	1,808,403	1,224,730	1,152,475
0.95	2,309,728	1,737,754	1,660,675
1.00	8,583,975	8,547,325	9,809,450
Average TMRCA over all loci	816,529	441,301	407,297
% of loci where MRCA is located in Africa	100.0	90.4	86.2
% of loci where MRCA is located in Asia	0.0	9.6	13.8
% of loci where MRCA is located in America	0.0	0.0	0.0

References

1. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978-989.

2. Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12: 921-927.

3. Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, et al. (2006) A comparison of phasing algorithms for trios and unrelated individuals. Am J Hum Genet 78: 437-450.

4. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.

5. Haile-Selassie Y, Suwa G, White TD (2004) Late Miocene Teeth from Middle Awash, Ethiopia, and Early Hominid Dental Evolution. Science 303: 1503-1505.

6. Excoffier L, Estoup A, Cornuet J-M (2005) Bayesian Analysis of an Admixture Model With Mutations and Arbitrarily Linked Markers. Genetics 169: 1727-1738.

7. Neuenschwander S (2006) Reconstruction of the post-glacial re-colonization of the Swiss Alps by the bullhead (*Cottus gobio*) based on spatially explicit computer simulations [PhD Thesis]. Berne: University of Berne. 229 p.

CAPÍTULO III

Mitochondrial Genomics and the Peopling of the Americas

Manuscrito submetido à Nature

Mitochondrial Genomics and the Peopling of the Americas

Nelson J. R. Fagundes[1,2*], Ricardo Kanitz[1*], Roberta Eckert[1], Ana C. S. Valls[1], Mauricio R. Bogo[1], Francisco M. Salzano[2], David Glenn Smith[3], Wilson A. Silva Jr.[4], Marco A. Zago[4], Andrea K. Ribeiro-dos-Santos[5], Sidney E. B. Santos[5], Maria Luiza Petzl-Erler[6] & Sandro L. Bonatto[1]

[1]Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, 90619-900 Porto Alegre, RS, Brazil. [2]Departamento de Genética, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil. [3]Molecular Anthropology Laboratory, Department of Anthropology, University of California - Davis, 95616 Davis, CA, USA. [4]Faculdade de Medicina, Universidade de São Paulo, 14051-140 Ribeirão Preto, SP, Brazil. [5]Departamento de Patologia, Universidade Federal do Pará, 66075-970 Belém, PA, Brazil. [6]Departamento de Genética, Universidade Federal do Paraná, 81531-590 Curitiba, PR, Brazil.

**These authors contributed equally to this work.*

In the complex history of human migrations, it is widely accepted that the Americas have been the last continent reached by *Homo sapiens*, most likely through Beringia, the landmass that connected Siberia and Alaska during the last ice age[1]. However, the precise time and mode of the colonization of the New World remain hotly disputed issues[1-6]. Here we show, using 244 mitochondrial genomes, that all Native American haplogroups, including the enigmatic haplogroup X[7], were part of a single founding population, refuting multiple migration models. A detailed demographic history of the mitochondrial genomes, estimated using a Bayesian coalescent method[8], suggests that this founding populations experienced a moderate bottleneck between ~23,000 and ~19,000 yr ago, followed by a short but

strong expansion that started ~18,000 and finished ~15,000 yr ago. Taken together, our results support a complex model for the Peopling of the Americas, where the initial differentiation in Asia ended with a population reduction in Beringia during the Last Glacial Maximum (LGM), followed, toward the end of the LGM, by the rapid settlement of the continent along a Pacific coastal route.

A popular model for the Peopling of the Americas suggests that the archaeological remains known as the Clovis complex (thought to be the oldest unequivocal evidence of humans in the Americas) represent the people that first colonized the continent after a Late Glacial migration through the ice-free corridor that separated the Laurentide and Cordilleran Ice Sheets[9]. However, the recently reevaluated age of the Clovis sites to only between *ca.* 12.7 to 13.2 thousand years ago (kya)[2] and the confirmed human presence at the Monte Verde site located in southern South America around 14.5 kya[3] challenge the Clovis-first model and call for alternative hypotheses. As the earlier date for Monte Verde implies that peopling of the Americas south of Beringia occurred before the ice-free corridor was formed, a first migration along the Pacific coast may be a viable route[4]. Unfortunately, archaeological verification of this scenario is very difficult since most of the late Pleistocene coast is currently underwater, as the see-level rose >120m since the end of the LGM[10].

The maternally inherited mitochondrial DNA (mtDNA) has also been widely used to understand the peopling of the Americas. Since the first studies, it was found that extant Native American populations exhibit almost exclusively five mtDNA haplogroups (A-D, and X)[1] classified in the autochthonous haplogroups A2, B2, C1, D1, and X2a[11]. Haplogroups A-D are found all over the New World and are frequent in Asia, supporting a northeastern Asian origin of these lineages[12]. This distribution, together with the haplogroups similar coalescence time was used to suggest a single migration model[13]. The history of haplogroup X is more elusive, as it is found at present

in the New World at a relatively low frequency and only in North America, it is rare in West Eurasians and almost absent in Siberia. In addition, some have claimed that Native American haplogroup X is less diverse and has a younger coalescence time than haplogroups A-D[7]. These differential features have been cited to argue that haplogroup X represents an independent migration to the Americas from Asia or even Europe[7]. Additionally, a different pattern of diversification and distribution of haplogroup B found in some studies led some authors to hypothesize that it represents a later and separate migration from the joint arrival of haplogroups A, C and D[14].

Another widely disputed issue concerns the timing of such major migrations. While recent studies have been argued to support a single migration with dates as early as 30 kya[1], the uncertainties about and range around these dates are very large. One cause for this variation is the limited information content of the mtDNA control region, which is also too divergent to allow reliable substitution rate estimation by comparison with the chimpanzee. Alternatively, the coding region of the mtDNA is being increasingly used to circumvent these limitations in studies of human migrations[15,16]. Here we analyze 244 mtDNA genomes (58 of them new) belonging to all five major Native American haplogroups (A2, B2, C1, D1, and X2a) to provide a better understanding of the timing and mode of the peopling of the New World.

The phylogeny of the Native American mtDNA genomes is shown in Fig. 1a. For each haplogroup, all Native American sequences trace back to a single founder haplotype that can be distinguished from Old World haplotypes by the presence of exclusive mutations or, in the case of haplogroup C, by specific sequence motifs[11]. The diversity pattern within each Native American haplogroups, including haplogroup X, is remarkably similar. Using the standard mutation rate of 1.26×10^{-8} per site per year for the mtDNA coding region[17], all haplogroups show coalescence times around 20 kya.

Similar values are found even when we relax the assumption of a molecular clock in a full Bayesian procedure or use an external calibration point (Table 1 and Fig. 1a).

Our data could also be used to better understand the demography of the process of colonization. All haplogroups exhibit a marked excess of low-frequency variants that is characteristic of a strong and recent population expansion (negative Tajima's D and Fu's Fs statistics, Supplementary Table 7), as well as single peaks in the mismatch distribution graphics (Supplementary Fig. 1a). These results strongly suggest a scenario in which all five haplogroups were part of a single founding population that ultimately led to the peopling of the whole continent, refuting former scenarios in which haplogroups X and/or B arrived separately from the others.

To get a more realistic picture of the complex demographic history associated with the colonization of the New World without assuming *a priori* any number of founding haplotypes for the haplogroups, we applied the Bayesian skyline plot approach to all Native American mtDNA genomes simultaneously[8]. The skyline plot (Fig. 1b) identifies a moderate population reduction between ~23-19 kya reaching a minimum of ~1,000 females followed by a strong and rapid size expansion beginning ~19-18 kya and ending ~16-15 kya. It is noteworthy that the time of the population reduction correlates very well with the LGM (23-18 kya) while the expansion dates are in excellent agreement with the end of the LGM, dated around 19-17 kya[18,19].

Overall, the Native American mtDNA genomic diversity suggests the following scenario for the peopling of the Americas: Native American haplogroups began to diverge from their ancestors in the route to northeast Asia. If we use the average number of mutations that are markers of the Native American haplogroups as a proxy for the length of isolation[11] (see Supplementary Information), we estimate that a period of ~11 thousand years elapsed between their separation from Asians and their diversification

and expansion through the Americas. This suggests that the divergence of the population that ultimately gave rise to Native Americans likely predates the LGM. Although we cannot determine where in northeast Asia this population stayed during this long period of isolation, Beringia represents the best candidate for that location. During the LGM, Beringia was mostly exposed, and even though archaeological evidence for human presence in Beringia around the LGM is controversial, there is evidence of human settlements in the Artic around 30 kya[20] and also that the Beringian environment could likely sustain human populations during the LGM[21]. We estimate that, beginning ~18-19 kya and ending ~15-16 kya (*i.e.* towards the end of the LGM), the founding population experienced a significant demographic growth process most likely associated with an extensive range expansion, which may mark the beginning of the effective colonization of the New World south of Beringia. Since the opening of the ice-free corridor is dated not earlier than ~14 kya, our results strongly support an expansion along the western coast of North America[5,22]. Recent data have shown that this route was largely ice-free by ~19 kya and that the environment improved rapidly, being capable of supporting bears ~15 kya[6]. Interestingly, the end of the intense expansion period coincides with the age of the southern South American Monte Verde site, ~14.5 kya[3]. The strong and rapid population growth suggested by our data is consistent with a model in which humans have traveled the >13,000 km along the coast from Alaska to the southern tip of Chile in a few thousand years[23].

This model could help explain why some of the earliest known sites are in coastal South America while more recent sites are more frequently situated inland. Associated with the end of the ice age, sea level rose rapidly between ~18 and ~10 kya, inundating most of North America's Pacific coast that was exposed during the earliest expansion southward. Some of the earliest sites might occur along the much larger South American western coastal plain because large portions of its prehistoric coastline are still exposed[24]. The human dispersal from the coast into the interior of the continent,

perhaps driven by growing population density, depletion of coastal resources and rising sea levels[4], was probably delayed by the need to cross the mountain ranges and change living strategies and technologies from those associated with coastal adaptations. Interestingly, a similar model was proposed to the first colonization of Asia ~65 kya[16].

Our results strongly support the hypothesis that haplogroup X was part of the gene pool of the Native American founding population together with the other four mtDNA haplogroups. However, we infer that haplogroup X experienced a more limited expansion in size and range than the former four haplogroups. If the founding haplotype of the Beringian representative of haplogroup X was present at a low frequency, a likely explanation for these observations would be that it was lost by successive founder effects and genetic drift as the expansion wave moves southward[25]. A similar explanation may be used to account for the existence of other similarly rare or even extinct haplogroups in the Americas, such as the recently described haplogroup M[26], without the need to postulate independent colonization events. The existence of additional, rare founding haplotypes is in agreement with the moderate bottleneck estimated by our data and by recent results from nuclear loci[27], but contradicts an extreme bottleneck hypothesis[28].

Methods

A detailed description of materials and methods is given in Supplementary Information.

Samples and sequencing. DNA samples were obtained from 58 individuals belonging to Native American populations and have been collected directly by some of the authors (F.M.S., S.E.B.S., M.A.Z., or D.G.S.). Supplementary Table 1 provides further details on the individuals studied. The PCR primers used, covering the entire mitochondrial genome, the amplification conditions as well as chromatograms assembling in individual genomes were performed as described elsewhere[29]. Although some mtDNAs

have a partial sequence already published[30], these were mostly re-sequenced to ensure maximum quality. Additionally, 186 Native American mtDNA genomes available in public databases have been used.

1. Schurr, T. G. The peopling of the New World: perspectives from molecular anthropology. *Annu. Rev. Anthropol.* **33**, 551-583 (2004).
2. Waters, M. R. & Stafford Jr, T. W. Redefining the age of Clovis: implications for the peopling of the Americas. *Science* **315**, 1122-1126 (2007).
3. Dillehay, T. D. *Monte Verde, A Late Pleistocene Settlement in Chile, Vol. 2; The Archaeological Context and Interpretation* (Smithsonian Institute Press, Washington DC, 1997).
4. Dixon, E. J. Human colonization of the Americas: timing, technology and process. *Quat. Sci. Rev.* **20**, 277-299 (2001).
5. Fladmark, K. R. Routes: alternate migration corridors for early man in North America. *Am. Antiquity* **44**, 55-69 (1979).
6. Kelly, R. L. Maybe we do know when people first came to North America; and what does it mean if we do? *Quatern. Int.* **109-110**, 133-145 (2003).
7. Brown, M. D. *et al.* MtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am. J. Hum. Genet.* **63**, 1852-1861 (1998).
8. Drummond, A., Rambaut, A., Shapiro, B. & Pybus, O. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185-1192 (2005).
9. Fagan, B. M. *Ancient North America: The Archaeology of a Continent* (Thames & Hudson, New York, 2000).

10. Clark, P. U. & Mix, A. C. Ice sheets and sea level of the last glacial maximum. *Quat. Sci. Rev.* **21**, 1-7 (2002).

11. Bandelt, H. J. *et al.* Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann. Hum. Genet.* **67**, 512-524 (2003).

12. Merriwether, D. A., Hall, W. W., Vahlne, A., & Ferrell, R. E. mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. *Am. J. Hum. Genet.* **59**, 204-212 (1996).

13. Bonatto, S. L. & Salzano, F. M. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Amer. J. Hum. Genet.* **61,** 413-1423 (1997).

14. Starikovskaya, Y. B. *et al.* MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am. J. Hum. Genet.* **63**, 1473-1491 (1998).

15. Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713 (2000).

16. Macaulay, V. *et al.* Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. *Science* **308**, 1034-1036 (2005).

17. Mishmar, D. *et al.* Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* **100**, 171-176 (2003).

18. Yokoyama, Y., Lambeck, K., Deckker, P. D., Johnston, P. & Fifield L. K. Timing of the last glacial maximum from observed sea-level minima. *Nature* **406**, 713-716 (2000).

19. Schaefer, J. M. *et al.* Near-synchronous interhemispheric termination of the last glacial maximum in mid-latitudes. *Science* **312**, 1510-1513 (2006).

20. Pitulko, V. V. *et al.* The Yana RHS site: humans in the Arctic before the last glacial maximum. *Science* **303**, 52-56 (2004).

21. Brubaker, L. B., Anderson, P. M., Edwards, M. E. & Lozhkin, A. V. Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *J. Biogeogr.* **32**, 833-848 (2005).

22. Dixon, E. J. *Quest for the Origins of The First Americans* (University of New Mexico Press, Albuquerque, 1993).

23. Surovell, T. A. Simulating coastal migration in New World colonization. *Curr. Anthropol.* **44**, 580-591 (2003).

24. Dillehay, T. D. The late Pleistocene cultures of South America. *Evol. Anthopol.* **7**, 206-216 (1999).

25. Klopfstein, S., Currat, M. & Excoffier, L. The Fate of Mutations Surfing on the Wave of a Range Expansion. *Mol. Biol. Evol.* **23**, 482-490 (2006).

26. Malhi, R. S. *et al.* Mitochondrial haplogroup M discovered in prehistoric North Americans. *J. Archaeol. Sci.* **34**, 642-648 (2007).

27. Battilana, J. *et al.* Molecular variability of the 16p13.3 region in Amerindians and its anthropological significance. *Ann. Hum. Genet.* **71**, 64-76 (2007).

28. Hey, J. On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol.* **3**, 0965-0975 (2005).

29. Rieder, M. J., Tayler, S. L., Tobe, V. O. & Nickerson, D. A. Automating the identification of DNA variations using quality-based fluorescence re-sequensing: analysis of the human mitochondrial genome. *Nucleic Acids Res.* **26**, 967-973 (1998).

30. Silva, W. A. *et al.* Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. *Am. J. Hum. Genet.* **71**, 187-192 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Grant support was from the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo a Pesquisa do Rio Grande do Sul (S.L.B.) and by a CAPES scholarship (N.J.R.F.). We are also grateful to Institutos do Milênio and Programas de Apoio a Núcleos de Excelência for extra support (F.M.S.) and to the National Institutes of Health (D.G.S.). Thanks to Cladinara R. Sarturi, Ronaldo R. Ferreira, Luana Cardoso-Silva, Renata Schmitt, Andre Schnorr, Gabrielle D. Salton, and Marina O. Favarini for technical help, and to Kim Hill, A. Magdalena Hurtado, Ramiro Barrantes, and Luis Rodriguez-Delfin for sample donations, as well as to all individuals who, by contributing their own samples, made this study possible. We thank Claudio Bravi for help with checking mutations and Eduardo Eizirik for suggestions on the manuscript.

Author Information Mitochondrial genomes obtained in this project have been deposited in GenBank under accession numbers xxxxx. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L.B. (slbonatto@pucrs.br).

Table 1. Coalescence times for the five Native American haplogroups based on median-joining calculation (ρ) and on Bayesian estimation.

Haplogroup	ρ – TMRCA (95%CI)*	Bayesian – TMRCA (95% CI)
A2	17,009 (15,354 – 18,663)	23,200 (17,850 – 30,790)
B2	21,301 (19,198 – 23,404)	22,980 (17,960 – 29,340)
C1	20,642 (16,929 - 24,355)	23,160 (18,030 – 30,280)
D1	19,653 (17,192 – 22,114)	21,330 (16,830 – 27,080)
X2a	17,983 (11,764 – 24,202)	20,160 (15,570 – 27,730)
average	19,318	22,166

*Estimated as ρ ±2xSD.

Figure 1 Phylogenetic tree and Bayesian skyline plot from Native American mtDNAs. **a**, Maximum likelihood tree from 80 different Native American mtDNA coding region haplotypes. The time axis (in kya) was estimated using a parametric molecular clock model calibrated assuming human x chimpanzee divergence at 6.5 million years ago; **b**, mtDNA Bayesian skyline plot showing the Native American population size trend using a log-normal relaxed clock with the standard substitution rate of 1.26×10^{-8} sites/years. The *y* axis is the effective number of females. The thick solid line is the median estimate and the thin lines (blue) show the 95% highest posterior density limits estimated using 60 million chains. Approximate dates for the LGM, Monte Verde, and Clovis sites are shown in the middle panel (see Supplementary Information for additional details).

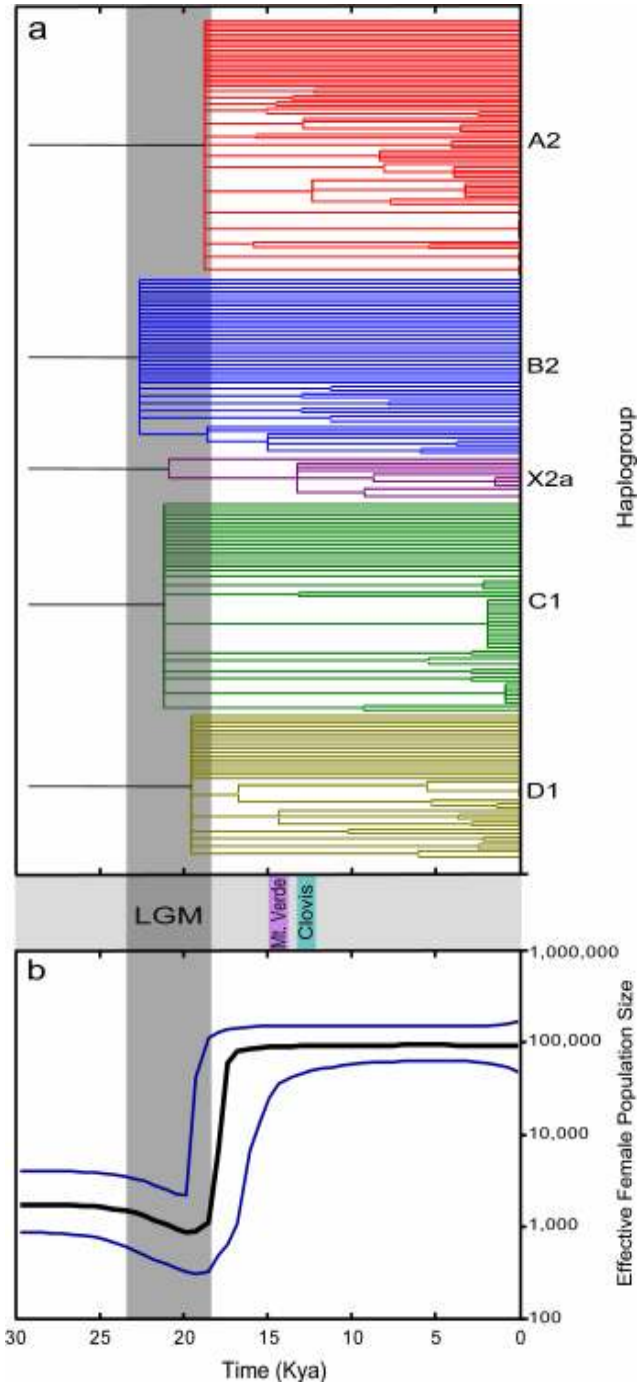

a
A2
B2
X2a
C1
D1
Haplogroup
LGM
Mt. Verde
Clovis
b
1,000,000
100,000
10,000
1,000
100
Effective Female Population Size
30
25
20
15
10
5
0
Time (Kya)

Fig 1.

Supplementary Information

Mitochondrial Genomics and the Peopling of the Americas

Nelson J. R. Fagundes, Ricardo Kanitz, Roberta Eckert, Ana C. S. Valls, Mauricio R. Bogo, Francisco M. Salzano, David Glenn Smith, Wilson A. Silva Jr., Marco A. Zago, Andrea K. Ribeiro-dos-Santos, Sidney E. B. Santos, Maria Luiza Petzl-Erler & Sandro L. Bonatto

Supplementary Methods

PCR, sequencing and Contig assembling. Given the low quantity of same of our DNA samples, we performed a genomic pre-amplification protocol using the GenomiPhi® kit (GE Healthcare) on these. Sequencing reactions covering the entire mitochondrial genome for both strands[29] were read in a MegaBACE 1000 (Ge Healthcare) using the ET Terminators® kit (GE Healthcare), following manufacturer's instructions. Chromatograms were assembled in the Phred-Phrap-Consed package[31,32]. After an initial visual inspection for low quality regions in the assembly, we aligned the contigs generated for every individual to each other and to the corrected Cambridge Reference Sequence (rCRS)[33,34] and checked all variable positions in the original chromatograms. Possible phantom mutations were again verified in the chromatograms and, whenever needed, re-sequenced from a new PCR product[35]. Polymorphic sites per haplogroup for the genomes obtained here are shown in Supplementary Tables 2-6.

Additional data. To the 58 genomes obtained here, we added 28 complete mtDNA genomes published scattered on the literature (see Supplementary Table 1). This comprises a small dataset of 86 complete mtDNA genomes characterized from mainly Native American individuals. Besides, two large scale databases[36,37] encompassing Native American mtDNAs but obtained from non-native individuals have been added to

generate a dataset of 244 mtDNA genomes comprising all available sequences from the five Native American haplogroups. The first database[36] included individuals sampled in urban centers whose mtDNA have been classified as Native American or Asian a posteriori. The second database[37] was aimed at assessing the power of mtDNA coding region sequence to discriminate among "Hispanic" individuals sharing the same control region sequence. To check the robustness of our conclusions against potential bias caused by the sampling origin and strategies employed in the two large databases[36,37], we performed the main analyses with both the total (n=244) and the n=86 native dataset. See Supplementary Notes below for further information.

Data analysis. All analyses were done using the slowly evolving mtDNA coding region (positions 577-16022) only. Control region sequence was used to confirm haplogroup assignment. Basic diversity statistics and neutrality tests were calculated with Arlequin 3.11 (ref. 38). Mismatch distributions for each haplogroup estimated using Arlequin are shown in Supplementary Fig. 1a.

Maximum likelihood phylogenetic trees were constructed using PAUP* 4.0 (ref. 39) under the HKY+G evolutionary model assuming an alpha parameter of 0.12 (ref. 16). The assumption of a molecular clock was tested using the PAML package[40], under the HKY+G evolutionary model assuming an alpha parameter of 0.12. For the native dataset (n=86) the null hypothesis of a molecular clock cannot be rejected (P=0.13). However, for the total dataset (n=244), the hypothesis of the molecular clock is rejected (P<0.01). Median-joining networks[41] were constructed with the program Network 4.1.0.2 (http://www.fluxus-engineering.com). Coalescence times were then calculated based on ρ using a rate of 1.26×10^{-8} substitutions per site per year[17] for the mtDNA coding region (positions 577-16022).

The tree with the time to most recent common ancestor (TMRCA) for the Native American mtDNA haplogroups given in Fig. 1 was estimated using the software r8s 1.7

(http://ginger.ucdavis.edu/r8s/). A maximum likelihood tree estimated in PAUP* with the HKY+G evolutionary model described above were optimized with the Langley-Fitch model and the Powell algorithm using the optimal smoothing value (S=1) obtained by a cross-validation procedure. We calibrated our estimates by assuming that the *Pan* and *Homo* lineages had separated from each other completely by 6 million years and added 500 ky for lineage sorting[16,17]. This procedure avoids the assumption of a substitution rate known *a priori*. This tree was constructed using sequences available in the GenBank from *Pan* (D38113, D38116, X93335) and individuals belonging to other haplogroups (AF346902, AF346966, AF346968, AF346974-5, AF346991-2, AF346994-5, AF346998-9, AF347000, AF347007-8, AF347014-5, AF381986, AP008566, AP008568, AY195747, AY195755, AY195760, AY195762, AY195766, AY195772-4, AY195777, AY195780, AY195783-4, AY195789, AY195796, AY255149, AY289066, AY289075, AY289078, AY289082, AY289086, AY289089), including Asians from haplogroups A-D, that were used to break long branches to improve phylogenetic reconstruction.

To investigate whether our inferences were robust when relaxing the assumption of a strict molecular clock we used the Bayesian approach for the estimation of the coalescence times (TMRCAs)[8] implemented in BEAST v1.4 (http://evolve.zoo.ox.ac.uk/beast/) which applies Markov Chain Monte Carlo (MCMC) integration for parameter estimation over the space of all equally likely trees. Population size dynamics through time (i.e. a Bayesian Skyline plots)[8] were also estimated using this approach in BEAST. Estimations were carried out assuming HKY+G model using the same rate used for ρ time estimations but allowing lognormal relaxation. The analysis was run for 60 million iterations, with the first 10% discarded as burn-in. Genealogies and model parameters were sampled every 1,000 iterations thereafter. The posterior probability density for the TMRCA of each haplogroup is shown in Supplementary Fig. 1b.

To check for mutations separating Native American and Old World haplogroups, we compared our sequences with sequences belonging to Asian (haplogroups A-D) and European (haplogroup X) individuals available in the literature[15,17,42-50]. Overall, there are 11 coding region mutations separating New World haplogroups from their Old World counterparts, that is, that are haplogroup markers (in exact agreement with 11), giving an average of 2.2 mutations per haplogroup. These are, for haplogroup A, 8027,12007; for haplogroup B, 3547, 4977, 6473, 9950, 11177; haplogroup D, 2092; haplogroup X, 8913, 12397, 14502. Using the above cited rate that is equivalent to one substitution per 5,138 years we obtain around 11,000 for the period of divergence and isolation of the founding population before the expansion.

Supplementary Notes

We have deliberately restricted our analysis to the populations known as "Amerindians", leaving aside people from Eskimo-Aleuts and Na-Dené linguistic groups. We[51] and others (reviewed in 1) have already demonstrated that the latter were part of the single founding population that gave origin of all Native Americans. However, there is also ample evidence[1] that the Eskimo-Aleuts and Na-Dené diverged from Amerindians >10 kya and since then underwent independent population contractions and re-expansions around the circumartic region. Methods such as Bayesian skyline plot, neutrality tests, etc. are only applicable to a group of populations that share the same demographic history.

In Supplementary Table 7 we present a comparison of the estimates of same basic statistics for the total dataset and for the reduced, native dataset. TMRCAs estimates based on the ρ statistic and on BEAST for the reduced dataset are reported in Supplementary Table 8 and the Bayesian skyline plot is shown in Supplementary Fig. 2. Mismatch distributions are shown in Supplementary Fig. 3a, and the density of TMRCAs estimated using BEAST are shown in Supplementary Fig. 3b. The differences

between the results of the reduced and total datasets are very small in all analysis. Therefore, our results with the complete dataset (244 sequences) are very robust and authentically represent present day mtDNA diversity in "Amerinds".

31. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175-185 (1998).

32. Gordon, D., Abajian, C., Green, P., Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195-202 (1998).

33. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-465 (1981).

34. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).

35. Bandelt, H.-J., Quintana-Murci, L., Salas, A. & Macaulay, V. The fingerprint of phantom mutations in mitochondrial DNA data. *Am. J. Hum. Genet.* **71**,1150-1160 (2002).

36. Herrnstadt, C. *et al.* Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am. J. Hum. Genet.* **70**, 1152-1171 (2002).

37. Parsons, T. J. GenBank accession nos. DQ282387-DQ282487.

38. Excoffier, L., Laval, G. & Schneider, S., Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinf. Online* **1**, 47-50 (2005).

39. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)*, Sunderland, MA (1998).

40. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**, 555-556 (1997).

41. Bandelt, H.-J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37-48 (1999).

42. Levin, B.C., Cheng, H. & Reeder, D.J. A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and utation detection. *Genomics* **55**, 135–146 (1999).

43. Finnilä, S., Lehtonen, M.S. & Majamaa, K. Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.* **68**, 1475–1484 (2001).

44. Maca-Meyer, N., González, A.M., Larruga, J.M., Flores, C. & Cabrera, V.C. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* **2**, 13 (2001).

45. Ingman, M. & Gyllensten ,U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.* **13**,1600–1606 (2003).

46. Kong, Q. P. *et al.* Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet.* **73**, 671–676 (2003).

47. Reidla, M. *et al.* Origin and diffusion of mtDNA Haplogroup X. *Am. J. Hum. Genet.* **73**, 1178-1190 (2003).

48. Tanaka, M. *et al.* Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res.* **14**, 1832–1850 (2004).

49. Starikovskaya, E. B. *et al.* Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann. Hum. Genet.* **69**, 67-89 (2005).

50. Kivisikd, T. *et al.* The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**, 373-387 (2006).

51. Bonatto, S.L. & Salzano, F.M. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc. Natl. Acad. Sci. USA* **94**, 1866-1871 (1997).

Supplementary Tables

Supplementary Table 1. MtDNA sequences obtained in this work or gathered from literature used for the analyses. To these sequences we added the mtDNA genome reported in Herrnstadt et al.[36] and Parsons[37].

Haplogroup	Original ID	GenBank ID	Tribe/Population	Reference
A2	ACHE30		Ache	this work
A2	WWAI01		Waiwai	this work
A2	WWAI25		Waiwai	this work
A2	ZORO02		Zoró	this work
A2	SURUI01		Suruí	this work
A2	WPI167		Waiãpi	this work
A2	Y655		Yanomama	this work
A2	PTJ03		Poturujara	this work
A2	Y623		Yanomama	this work
A2	KKT13		Kriketun	this work
A2	KTN130		Katuena	this work
A2	GRC149		Guarani/Rio das Cobras	this work
B2	ACHE78		Ache	this work
B2	GAVIAO23		Gavião	this work
B2	POMO01		Pomo/North California	this work
B2	WWAI24		Waiwai	this work
B2	XAVAN04		Xavante	this work
B2	XAVAN12		Xavante	this work
B2	1876		Guarani	this work
B2	1880		Guarani	this work
B2	1881		Guarani	this work
B2	GRC169		Guarani/Rio das Cobras	this work
B2	KBK23		Kubemkokre	this work
B2	KBK39		Kubemkokre	this work
B2	KKT01		Kriketun	this work
B2	KRC33		Guarani/Rio das Cobras	this work
B2	KTN209		Katuena	this work
B2	Y637		Yanomama	this work
C1	WWAI16		Waiwai	this work
C1	ZORO19		Zoró	this work
C1	ZORO31		Zoró	this work

C1	1875		Guarani	this work
C1	1878		Guarani	this work
C1	ARL58		Arara/Arara do Laranjal	this work
C1	PTJ68		Poturujara	this work
C1	Y591		Yanomama	this work
C1	Y650		Yanomama	this work
C1	Y669		Yanomama	this work
D1	GAVIAO12		Gavião	this work
D1	GAVIAO26		Gavião	this work
D1	SURUI22		Suruí	this work
D1	WWAI05		Waiwai	this work
D1	ZORO23		Zoró	this work
D1	GRC131		Guarani/Rio das Cobras	this work
D1	KTN18		Katuena	this work
D1	PTJ01		Poturujara	this work
D1	TYR04		Tiryó	this work
D1	TYR16		Tiryó	this work
X2a	CHIP20		W. Chippewa/NE	this work
X2a	CHIP44		W. Chippewa/NE	this work
X2a	CHIP76		W. Chippewa/NE	this work
X2a	CHIP85		W. Chippewa/NE	this work
X2a	CHIPSAM2		Chippewa/NE	this work
X2a	CHIPSW097		Chippewa/NE	this work
X2a	JEMEZ22		Jemez/SE	this work
X2a	JEMEZ435		Jemez/SE	this work
X2a	JEMEZ990		Jemez/SE	this work
X2a	SIOUAN59		Siouan/SE	this work
A2	Na5A	AY195786	Native American*	17
A2	N/A	AF346971	Chukchi	15
A2	haplotype A	AF382010	Canary	44
A2	AM17	DQ112832	Auca	50
B2	Na1B	AY195749	Native American*	17
B2	N/A	AF347001	Pima	15
B2	AM12	DQ112889	Mayan	50
B2	AM15	DQ112790	Colombian Indian*	50
B2	AM16	DQ112791	Colombian Indian*	50
C1	Na4C	AY195759	Native American*	17

C1	haplotype C	AF382009	Canary	44
C1	N/A	AF347012	Warao	15
C1	N/A	AF347013	Warao	15
C1	AM03	DQ112789	Colombian Indian*	50
C1	AM04	DQ112888	Mayan	50
C1	AM06	DQ112846	Navajo	50
D1	Na2D	AY195748	Native American*	17
D1	N/A	AF346984	Guarani	15
D1	AM01	DQ112772	Brazilian Indian*	50
D1	AM02	DQ112776	Brazilian Indian*	50
D1	AM07	DQ112871	Quechua	50
D1	AM08	DQ112872	Pima	50
D1	AM09	DQ112773	Brazilian Indian*	50
D1	AM10	DQ112774	Brazilian Indian*	50
D1	AM11	DQ112775	Brazilian Indian*	50
D1	AM14	DQ112843	Guarani	50
X2a	NA22	N/A	Ojibwa	11
X2a	Na3X	AY195787	Navajo	17

* No further information available.

Supplementary Table 2: Variables sites for each sequence compared to the corrected Cambridge Reference Sequence (rCRS) for haplogroup A2.

					1111	1111111111	1111111111	1111
	111122	2333444444	4466677888	8889990001	1122233444	4444445555	5555
	6677124727	8046022457	8825707067	7880393562	7904719114	5677893347	8999
	4605183340	5839914556	1217223246	9679969898	1400092277	6656671273	2456
	3390778666	0353268409	1465382784	4006268948	9476518786	6856187604	6618
rCRS	AAGAAAAAAA	TTCGGTTTTA	AATAGCAGGG	CATTAGAGAC	GTGGCTGATG	ACACGAGATG	ACAT
ACHE30	.G.G..GG.G	.C....C..G	.G...T.A..	TGC.....	A.A.T.....	...T...G.A	....
WAIWAI01	TG.G..GG.G	....A.C..G	.G..AT.A..	TG.C.....	A.A.TC.C..	...T...G..	..G.
WAIWAI25	.GAGG.GG.G	......C..G	.G...T.A..	TG.....	A.A.T.....	...T...GC.	....
ZORO02	.G.G.GGG.G	......C..G	.GC..TGA..	TG.....A..	A.A.T....A	...T...G..	....
SURUI01	.G.G..GG.G	..T...C..G	.G...T.A..	TG.....	A.A.T.....	.T.T...G..	....
WPI167	.G.G..GG.G	......C.CG	GG.G.T.A..	TG....G...	A.A.T.....	...T.G.G..	....
Y655	.G.G..GG.G	...A..CC.G	.G...T.A..	TG.....	A.A.T.C...	...T..AG..	....
PTJ03	.G.G..GG.G	......C..G	.G...T.A..	TG.....	ACA.T.....	...T...G..	.T..
Y623	.G.G..GG.G	...A..CC.G	.G...T.A..	TG......G.	A.A.T.C...	...T..AG..	....
KKT13	.G.G..GG.G	C....CC..G	.G...T.AAA	TG..GA....	A.A.T.....	G..T...G..	G..C
KTN130	.G.G..GG.G	..T...C..G	.G...T.A..	TG.......T	A.AAT...C.	..GTA..G..	....
GRC149	.G.G..GGGG	......C..G	.GC..T.A..	TG.....	A.A.T.....	...T...G..	....

Supplementary Table 3: Variables sites for each sequence compared to the corrected Cambridge Reference Sequence (rCRS) for haplogroup B2.

						111111	1111111111	1111111111	111
	11223	3334444455	5666666777	7777788888	8899000001	1111222233	3344444455	555
	6789947074	5692378926	9122447022	2446822457	7819168991	1178136757	9900144713	579
	0526831506	4113862795	7178775225	7092157351	3685009572	5712991790	2349117602	381
	6071289663	7582590715	8926135871	8386017525	6020140465	0791208108	8294000666	544
rCRS	AAATAAGGAA	AAGTAAGTTT	ACATACGCGT	TAGCCGTATT	TAGTTTACCC	GCGAGCGGGG	GCCTTGTCGA	CTA
ACHE78	.GG..G..G.	G....GAC..	.....T.T..	....	.G.C......	.TAG....A.	...C.A.T.G	T..
GAVIAO23	.GG..G..G.	G....GAC..	....GT.T..	....	.G.CC....T	.TA..T.AA.	.......T.G	T..
POMO01	.GG..G..G.	G...GGAC..	.....TAT..	C...T.C...	.G.C......	.TA.....A.	....C..T.G	T..
WAIWAI24	.GG.GG..G.	G....GAC..	.T...T.T..	.G........	.GACC.....	.TA.....A.	..T....T.G	T..
XAVANTE04	.GG..G.AG.	G....GACC.	...C.TAT..	....	.G.C......	.TA.....A.	CT.....T.G	T..
XAVANTE12	.GG..G..G.	G....GAC..	G....T.T..	..A......A	.G.C......	.TA.....A.	......CT.G	T..
1876	.GG..G..G.	G....GAC..	.....T.T..	....	.G.C......	.TA...A.A.	.......T.G	T..
1880	GGG..G..G.	G....GAC..	.....T.T..	....	.G.C......	.TA.A...A.	.......T.G	T..
1881	.GG..G..GG	G....GAC..	.....T.T..	....	.G.C......	.TA.....A.	.......T.G	T..
GRC169	.GG..G..G.	G....GAC..	.....T.T..	...T......	.G.C......	.TAG....A.	.......T.G	T..
KBK23	.GG..GA.G.	G.AC.GAC..	.....TAT..	........C.	.G.C.C....	.TA.....AA	.......T.G	TC.
KBK39	.GGC.G..G.	G....GAC..	..G..T.T..	....	CG.C..TTT.	.TA.....A.	.......T.G	T..
KKT01	.GGC.G..G.	G....GAC..	..G..T.T..	....	.G.C......	.TA.....A.	.......T.G	T.C
KRC33	.GG..G..G.	GG.C.GAC..	.....T.TTC	....	.G.C......	.TA.....A.	.......T.G	T..
KTN209	.GG..G..G.	G....GAC.C	.....T.T..	.......G..	.G.C......	ATA.....A.	.......T.G	T..
Y637	.GG..G..G.	G....GAC..	.....TAT..	.....A....	.G.C......	.TA.....AA	.......TAG	T..

Supplementary Table 4: Variables sites for each sequence compared to the corrected Cambridge Reference Sequence (rCRS) for haplogroup C1.

			111111	1111111111	111111
	123344444	5677788888	8899000001	1222333344	445555
	7475704779	2401603557	8855334587	9178235637	770334
	5305619162	3529978080	4644190871	1907624516	884028
	0862441594	8686783341	8005080639	4359365686	383167
rCRS	AAATCCGAAG	CGCCGGTTGA	TATAGACGTG	GACTATCTTC	TTGGAA
WAIWAI16	GGGA...GG.	..TAA..CAG	.GCG.GT.CA	A.TCG...CT	C.AAGT
ZORO19	GGGA...GG.	..TA..C.AG	.GCG.GTACA	AGT.G...CT	C.AAGT
ZORO31	GGGA...GG.	..TA..C.AG	.GCG.GTACA	AGT.G...CT	C.AAGT
1875	GGGA.TAGG.	..TA....AG	.GCG.GT.CA	A.T.G..CCT	C.AAGT
1878	GGGA...GG.	.ATA....AG	.GCG.GT.CA	A.T.G.T.CT	C.AAGT
ARL58	GGGAT..GGA	..TA.A..AG	.GCG.GT.CA	A.T.G...CT	C.AAGT
PTJ68	GGGA...GG.	T.TA....AG	.GCG.GT.CA	A.T.G...CT	CCAAGT
Y591	GGGA...GG.	..TA....AG	CGCG.GT.CA	A.T.GC..CT	C.AAGT
Y650	GGGA...GG.	..TA....AG	CGCG.GT.CA	A.T.GC..CT	C.AAGT
Y669	GGGA...GG.	..TA....AG	CGCGAGT.CA	A.T.GC..CT	C.AAGT

Supplementary Table 5: Variables sites for each sequence compared to the corrected Cambridge Reference Sequence (rCRS) for haplogroup D1.

			111111111	1111111111	111111111
	122334445	5557778889	9000000111	1122223444	445555555
	7407032781	3480124785	7134888124	7947880006	770133357
	5390111687	2722371064	5590177548	1100195356	684001211
	0826068938	4118484100	3980634082	9465029458	633613649
rCRS	AACAGGTACC	CGGCATCAAT	GCACATCGAT	GGGCATCTCC	CTGGGTATT
GAVIAO12	GGTGA..GTA	T..T.CTGGC	..GT.C....	A..T...CTT	TCA.A.G..
GAVIAO26	GGTGA..GTA	..AT..TGGC	..GT.C....	A..T.....T	TCA.ACG.C
SURUI22	GGTGA..GTA	.AAT..TGGC	..GT.C....	A..T.....T	T.A.ACG.C
WAIWAI05	GGTGA..GTA	...T..TGGC	..GT.CT...	...T.....T	TCA.A.G..
ZORO23	GGTGAA.GTA	...T..TGGC	A.GT.C..GC	A..T.C...T	TCA.A.G..
GRC131	GGTGA..GTA	..AT..TGGC	..GTTC....	AA.T..T..T	TCA.A.GC.
KTN18	GGTGA..GTA	...T..TGGC	.GGT.CT...	...T.....T	TCA.A.G..
PTJ01	GGTGA.CGTA	...T..TGGC	..GT.C.A..	A..T.....T	TCAAA.G..
TYR04	GGTGA..GTA	...TC.TGGC	..GT.CT...	A.ATG....T	TCA.A.G..
TYR16	GGTGA..GTA	...T..TGGC	..GT.CT...	A.ATG....T	TCA.A.G..

Supplementary Table 6: Variables sites for each sequence compared to the corrected Cambridge Reference Sequence (rCRS) for haplogroup X2a.

		1	111111111
	112233466	6677788880	112234445
	7473755712	3602648893	273794573
	5319025612	7829924618	219067062
	0893672931	1089722039	997560266
rCRS	AAGCACTAAT	CTCAGAAAAT	CGACATTCA
CHIP20	GGA.GGCG.C	TCTG.G.GG.	.AGTGCCTG
CHIP44	GGA.G.CG.C	T.T..G.GG.	TAGTGCCTG
CHIP76	GGA.G.CG.C	T.T...GGG.	.AGTGCCTG
CHIP85	GGA.G.CG.C	TCTG.G.GG.	.AGTGCCTG
CHIPSAM2	GGA.G.CG.C	TCTG.G.GG.	.AGTGCCTG
CHIPSWO97	GGA.G.CG.C	T.T....GG.	.AGTGCCTG
JEMEZ22	GGATG.CGGC	T.T....GG.	.AGTGCCTG
JEMEZ435	GGATG.CGGC	T.T....GG.	.AGTGCCTG
JEMEZ990	GGATG.CGGC	T.T....GG.	.AGTGCCTG
SIOUAN59	GGA...CGGC	T.T.A..GGC	.AGTGCCTG

Supplementary Table 7. Summary statistics for the total and for the reduced datasets.

Haplogroup	n	S	π (SD) %	Tajima's D	Fu's Fs
total dataset					
A2	87	185	0.0425 (0.0225)	-2.787**	-25.074**
B2	48	152	0.0526 (0.0276)	-2.755**	-24.872**
C	57	93	0.0456 (0.0241)	-2.272**	-24.715**
D1	40	96	0.0478 (0.0253)	-2.462**	-25.005**
X2a	12	20	0.0304 (0.0180)	-1.277*	-2.410*
reduced dataset					
A2	16	58	0.0512 (0.0282)	-2.333**	-9.897**
B2	21	72	0.0504 (0.0273)	-2.468**	-15.997**
C	17	44	0.0417 (0.0233)	-2.097**	-7.200**
D1	20	44	0.0484 (0.0263)	-1.594**	-7.280**
X2a	12	20	0.0304 (0.0180)	-1.277*	-2.410*

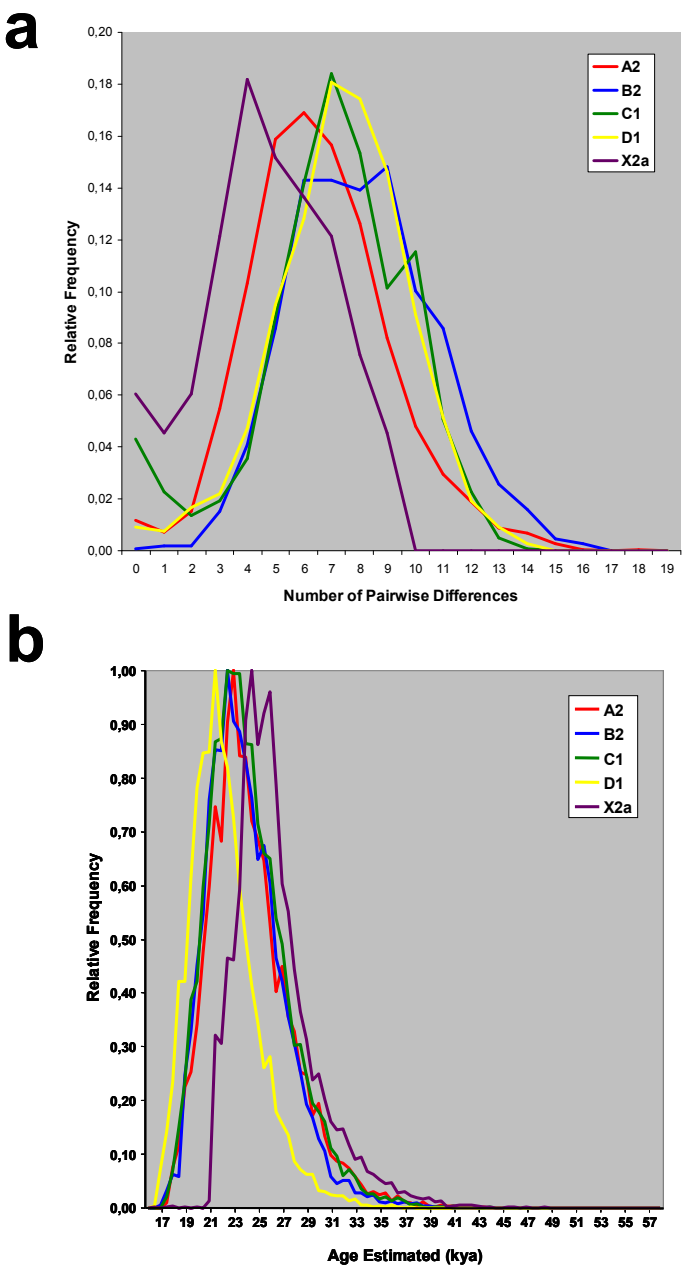
*P<0.10, **P<0.05.

Supplementary Table 8. TMRCAs in years for the reduced dataset of 86 sequences, based on a median-joining (ρ) or Bayesian approach (Beast).

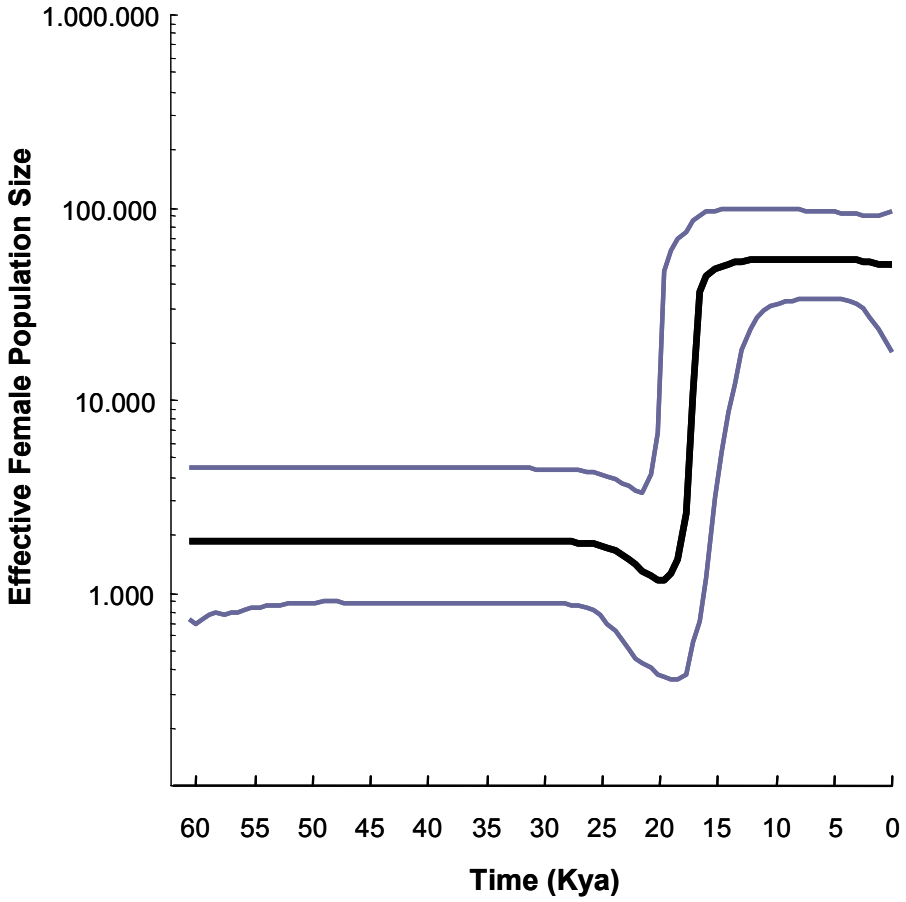
Haplogroup	ρ (95%CI)*	Bayesian (95%CI)
A2	20,552 (14,953-26,151)	21,290 (16,550-28,130)
B2	20,307 (15,246-25,369)	22,140 (17,570-28,730)
C1	17,227 (11,461-22,994)	20,680 (16,830-26,260)
D1	21,580 (13,263-29,896)	21,430 (16,850-28,730)
X2a	17,983 (6,056-29,910)	20,730 (16,100-29,000)
Mean	19,530	21,254

*95% CI estimated as average ± (2×SD).

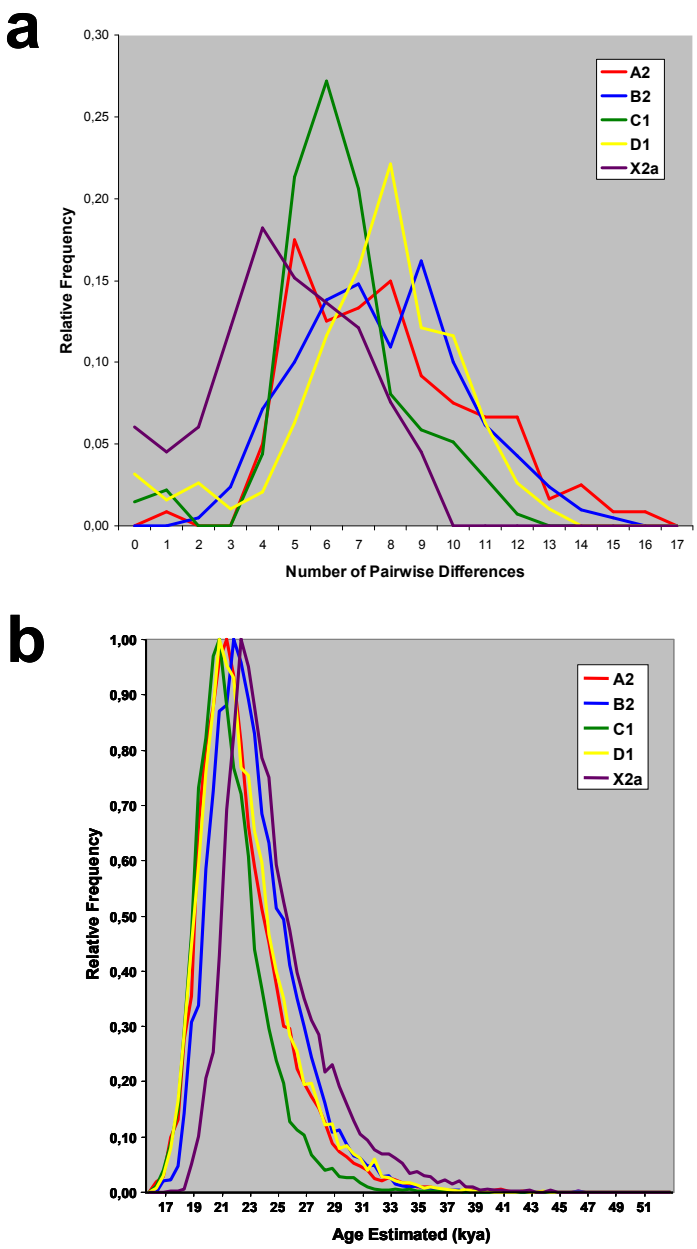
Supplementary Figures and Legends



Supplementary Fig. 1. **a**, Mismatch distributions for each Native American haplogroup; and **b**, Bayesian estimation of TMRCA density for each Native American haplogroup.



Supplementary Fig. 2. Bayesian skyline plot considering the reduced dataset of 86 sequences. Median estimate is shown as the black line. The 95% credible interval is limited by the thin (blue) lines.



Supplementary Fig. 3. Results with the reduced dataset of 86 sequences. **a**, Mismatch distributions for each Native American haplogroup; and **b**, Bayesian estimation of TMRCA density for each haplogroup.

CAPÍTULO IV

Discussão Geral

IV. DISCUSSÃO GERAL

A presente tese apresenta dois trabalhos que mostram como marcadores moleculares podem contribuir para um melhor entendimento de questões atuais pertinentes à evolução das populações humanas. Embora cada trabalho tenha suas particularidades e já tenha sido discutido em seu contexto próprio, é também relevante uma análise de como ambos os trabalhos podem ser entendidos num contexto mais amplo, principalmente a partir do ponto de convergência de ambos: o povoamento das Américas.

IV.1. A total "modernidade" dos nativos americanos

É consenso que o nordeste asiático e as Américas foram ocupadas apenas pelo *H. sapiens* moderno, não havendo nenhum vestígio de ocupação por hominídios arcaicos (Hoffecker e Elias, 2003). Entretanto, a possibilidade de hibridização entre humanos modernos e arcaicos na Ásia sugerida por alguns autores (Templeton, 2002; 2005; Eswaran *et al.*, 2005; Garrigan, 2005b; Evans *et al.*, 2006) permite que caso a assimilação de linhagens arcaicas tenha ocorrido na Ásia, tais linhagens tenham viajado até as Américas através da população fundadora asiática. Os resultados obtidos e apresentados no capítulo II da presente tese deixam claro que um cenário onde haveria a presença de linhagens arcaicas na Ásia e nas Américas possui uma probabilidade relativa extremamente baixa em relação a um cenário de origem africana e substituição.

IV.2. Estimativas da idade do povoamento do continente americano

Em relação aos parâmetros calculados para o povoamento das Américas podemos comparar as estimativas obtidas nos capítulos II e III. Usando marcadores nucleares, a estimativa de idade para o povoamento das Américas é entre 7.600 e 13.375 anos atrás, ao passo que usando o mtDNA essa mesma estimativa (supondo que a idade de expansão é a que melhor representa a colonização do continente) fica entre 18.000 e 19.000 anos. Como a diferença é significativa, podemos nos perguntar que fatores podem ter influenciado tal discrepância entre as idades e se, afinal de contas, ambas são congruentes. Uma diferença evidente é o ponto de calibragem entre humanos e chimpanzés. Enquanto o estudo dos 50 locos utilizou uma calibragem de 6 milhões de anos (Goodman *et al.*, 1998; Haile-Selassie

et al., 2004), a taxa evolutiva calculada para o mtDNA por Mishmar *et al.* (2003) usou um ponto de calibragem de 6,5 milhões de anos. Embora essa pequena (~8%) diferença seja insuficiente para explicar totalmente a diferença entre as estimativas, outros fatores podem também ter contribuído para tanto, como veremos a seguir.

Um ponto importante a ser considerado no caso dos locos nucleares é o próprio desenho amostral. Marcadores haplóides mostram maior incidência de haplótipos raros na América do Norte (Schurr, 2004; Salzano, 2007). Como nosso estudo utilizou uma amostragem restrita a indivíduos centro e sul-americanos, é possível que essa estratégia tenha sub-estimado o grau de variabilidade genética e, por conseqüência, o tempo de povoamento do continente. No caso do mtDNA, além de contarmos com uma amostragem maior, existe um número grande de mutações dentro de cada haplogrupo, de modo que a estimativa média de mutações dentro de cada haplogrupo, que influencia a estimativa do tempo de expansão é mais robusta à inclusão de novos haplótipos.

A própria estimativa obtida com o mtDNA, aliás, não é livre de pressupostos que podem introduzir algum viés nos cálculos. Alguns autores sugerem que há evidências de seleção purificadora no mtDNA humano contra mutações não-sinônimas (Moilanen e Majamaa, 2003; Kivisild *et al.*, 2006), e esses últimos autores sugerem também uma taxa evolutiva que em princípio, corrigiria os efeitos de seleção. Essa taxa evolutiva é consideravelmente mais rápida do que a taxa sugerida por Mishmar *et al.* (2003) e usada na análise dos dados de mtDNA apresentada no capítulo III. Por exemplo, o uso dessa taxa mais rápida geraria uma estimativa de 14.133 anos (para detalhes metodológicos, por favor consultar o capítulo III), cerca de 5.000 anos mais recente do que a estimativa usando a taxa de Mishmar e colaboradores. De modo similar, a estimativa do início de expansão mostrado no gráfico de *skyline* cai para cerca de 15 mil anos. Bandelt *et al.* (2006) publicaram uma análise profunda sobre a estimativa e taxas evolutivas para o mtDNA e concluíram que embora deva haver algum impacto de seleção purificadora no mtDNA humano, o cálculo apresentado por Kivisild *et al.* (2006) apresenta alguns problemas. Porém, é bastante provável que uma re-calibragem da taxa evolutiva do mtDNA humano que leve em conta adequadamente o efeito da seleção purificadora poderá conduzir a alguma aproximação entre os valores estimados para o mtDNA e os locos autossômicos, embora a diferença deva ser menor do que aquela apresentada usando a taxa de Kivisild *et al.* (2006).

Concluindo, parece possível que uma série de fatores de incerteza teriam se somado para produzir estimativas de valor distintas nos dois conjuntos de marcadores, que na verdade, sugerem o mesmo cenário de povoamento. Entretanto, cabe ressaltar que essa não é a única alternativa. Um cenário onde a variabilidade genética autossômica seria influenciada por duas migrações, sendo uma mais recente do que aquela sugerida pelo mtDNA, poderia ser compatível com as datas mais recentes para os locos autossômicos. Essa possibilidade será discutida em mais detalhes abaixo. É interessante ressaltar que Hey (2005) num conjunto de dados que incluía marcadores no mtDNA, cromossomo Y, X e autossomos, também encontrou datas muito recentes para o povoamento das Américas, em torno de 7.000 anos na estimativa de ponto, embora o intervalo de confiança tenha sido bastante grande (até cerca de 40.000 anos atrás).

IV.3. Estimativas de tamanho populacional

Em relação ao tamanho efetivo antes e após a colonização das Américas, nenhum dos dois conjuntos de dados reproduziu o cenário de efeito extremo de gargalo-de-garrafa proposto por Hey (2005). Interessantemente, as estimativas obtidas com o mtDNA foram maiores do que aquelas obtidas com os marcadores autossômicos. Para a população fundadora, os marcadores autossômicos estimaram um tamanho efetivo aproximado entre 75 e 1.350 indivíduos (~500 na mediana), enquanto para o mtDNA, os valores sugeriram um tamanho efetivo entre 300 e 2.000 mulheres (~800 na mediana). Assim como foi previamente discutido para a estimativa da idade do povoamento, há necessidade de uma amostragem na América do Norte para verificar se a variabilidade para estes marcadores não está ligeiramente subestimada, com reflexos na estimativa do tamanho da população. Por outro lado, o método de gráfico *skyline* Bayesiano é relativamente recente (Drummond *et al.*, 2005), e nunca foi exaustivamente testado para cenários complexos. Um estudo onde uma versão anterior desse método foi aplicado a um cenário demográfico simples sugeriu que a estimativa do tamanho populacional pode ser ligeiramente super-estimada (Strimmer e Pybus, 2001).

A diferença no tamanho populacional, entretanto, é maior para os números após a expansão populacional. Enquanto as estimativas usando marcadores autossômicos sugerem um crescimento de aproximadamente 20 vezes, sugerindo valores aproximados entre 800 e

20.000 indivíduos (~8.000 na mediana), os dados de mtDNA sugerem um crescimento de ~100 vezes, num tamanho populacional atual aproximadamente entre 50.000 e 200.000 indivíduos (~100.000 na mediana). Além dos fatores já discutidos anteriormente, cabe aqui ressaltar que o mtDNA, devido ao seu menor tamanho efetivo e à sua taxa de mutação mais elevada, recupera-se de um evento gargalo-de-garrafa mais rapidamente do que os marcadores de seqüência autossômicos (Fay e Wu, 1999). Esta característica pode também ser responsável pelo fato de que para o mtDNA, o padrão de crescimento populacional é praticamente o de uma expansão instantânea. Porém, para os marcadores autossômicos, um modelo de expansão exponencial recebeu muito maior suporte do que um de expansão súbita (para todas as populações conjuntamente). Uma possível explicação para isso é que para locos autossômicos, poucas mutações novas teriam surgido desde a saída da África, de modo que a escolha do modelo de crescimento (instantâneo ou exponencial) acabaria sendo dominada pela história das populações africanas, cujas evidências sugerem um crescimento populacional antigo e lento (p. ex. Reich e Goldstein, 1998; Marth *et al.*, 2004). Em princípio, poderíamos então supor que um modelo com crescimento exponencial para populações africanas e instantâneo para populações não-africanas deveria receber maior suporte. Entretanto, o teste exaustivo de modelos demográficos para populações modernas está fora do escopo do presente trabalho. Além disso, o teste de muitas hipóteses simultaneamente só é possível caso o conjunto de dados tenha poder suficiente para diferenciar entre todas elas.

Os resultados apresentados nos capítulos II e III, que apóiam a noção de um efeito gargalo-de-garrafa moderado durante o povoamento das Américas, encontra suporte em outros estudos recentes, como o de Battilana *et al.* (2006), apresentado no anexo II da presente tese, Heller *et al.* (2004), Mateus Pereira *et al.* (2005), Battilana *et al.* (2007) e mesmo Fagundes *et al.* (2005). Em relação a este último trabalho, apresentado no anexo I desta tese, embora os autores tenham encontrado evidência de algum efeito fundador durante o povoamento do continente americano, o fato do gene estudado ter alguma evidência de seleção balanceadora poderia ter acentuado este efeito. Além disso, num cenário de redução populacional inicial moderada, como o aqui proposto, é esperado que alguns locos específicos apresentem sinais mais evidentes desse processo.

IV.4. A questão morfológica: inferências possíveis

Os dois artigos incluídos na presente tese podem enriquecer o debate sobre as diferentes morfologias encontradas no continente? Em relação ao estudo de locos autossômicos, uma possibilidade para a idade recente inferida para o povoamento do continente seria a existência de duas ondas migratórias com extensa hibridização. Por outro lado, o mtDNA apóia claramente um modelo onde toda a variabilidade atual desse marcador genético teria sua origem em uma única onda migratória.

A hipótese de Neves *et al.* (1999; 2005; 2007; Neves e Hubbe, 2005), na qual uma população de morfologia semelhante à apresentada atualmente por nativos americanos modernos teria substituído a outra de morfologia "paleoamericana" não pode ser testada com dados de DNA de populações recentes, uma vez que a população extinta não teria contribuído para o conjunto gênico atual. Porém, uma dificuldade para esta hipótese está nas datações obtidas para os marcadores genéticos, uma vez que algumas formas "paleoamericanas" parecem ter persistido em registro arqueológico tão recente quanto 7.000 anos atrás. É bastante difícil conciliar um cenário de substituição total de populações humanas com a entrada das populações de morfologia "recente" há cerca 15.000 anos, e a persistência, no continente, de populações "paleoamericanas" por milênios sem que houvesse miscigenação entre elas. A evidência apresentada pelo mtDNA de uma expansão rápida no continente torna ainda menos provável um cenário onde a substituição da população paleoamericana teria ocorrido ao longo de milênios. Um outro problema vem da própria análise morfológica que sugere que algumas populações ameríndias modernas possuem características morfológicas semelhantes às dos "paleoamericanos" (González-José *et al.*, 2003).

Por outro lado, o modelo proposto por González-José *et al.* (submetido) sugere que após uma primeira migração de populações de morfologia "paleoamericana" seguiu-se um período de troca genética entre as populações que estavam já no continente e as de morfologia "mongolóide" vindas da Ásia. Currat e Excoffier (2004; 2005) estudaram a dinâmica das linhagens gênicas em situações de interação entre "camadas" populacionais como no caso da interação entre humanos e neandertais, além de processos envolvendo migrações paleolíticas e neolíticas, relacionada ao povoamento da Europa. Esses autores concluíram que dependendo da configuração de alguns parâmetros-chave quando uma

nova população invade um território já ocupado, as suas linhagens têm poucas chances de persistir após sucessivas gerações a não ser na ausência de interação genética entre as duas camadas populacionais, ou se as linhagens estão sob seleção positiva. Além disso, espera-se que a maior parte da contribuição da população mais recente fique concentrada próxima à origem geográfica de sua expansão.

No caso específico do continente americano, podemos interpretar as populações "paleoamericana" e "mongolóide" como duas camadas distintas. Dessa forma, se considerarmos o mtDNA como um marcador neutro, ou pelo menos não responsável por um maior sucesso adaptativo da população mongolóide, esperaríamos que sua história fosse marcada pela camada mais antiga "paleoamericana", não apresentando, necessariamente, evidências de mais de uma onda migratória. Em relação aos marcadores autossômicos, esperaríamos uma contribuição da camada "mongolóide" principalmente em genes associados à morfologia (no caso da morfologia mongolóide apresentar alguma vantagem seletiva). Um ponto obscuro seria o grau de seleção necessário para fazer com que essas linhagens pudessem espalhar-se por todo o continente em um espaço de tempo relativamente curto (Neves e Pucciarelli, 1991). Se os resultados dos locos autossômicos de fato tiverem sido influenciados por uma segunda migração, o modelo de González-José *et al.* (submetido) pode ser uma representação adequada do processo de colonização. Curiosamente, se as estimativas obtidas a partir de autossomos forem totalmente concordantes com os resultados de mtDNA, nem um modelo de migração única, nem o modelo de fluxo gênico entre os dois componentes morfológicos poderiam ser refutados, visto que nenhum dos marcadores analisados está sabidamente associado a regiões funcionais associadas à morfologia crânio-facial.

A metodologia de computação Bayesiana aproximada (ABC) apresentada no capítulo II pode fornecer um excelente meio para o teste das hipóteses de migração única e de fluxo gênico. Entretanto, por tratar-se de um cenário onde os eventos demográficos são recentes, e onde as populações parentais de ambas camadas genéticas podem não ser muito diferenciadas geneticamente, a discriminação entre esses cenários através de estatísticas que resumem os dados pode não ser trivial. Para tanto, é necessário que se estabeleça um conjunto de dados envolvendo um grande número de populações e de locos, preferencialmente distribuídos por todo o genoma.

CAPÍTULO V

Referências Bibliográficas

V. Referências Bibliográficas

Aiello L (1993) The fossil evidence for modern human origins in Africa; a revised view. Am Anthropol 95:73-96.

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA e Kruglyak L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2:e286. http://biology.plosjournals.org/

Bandelt H–J, Kong Q–P, Richards M e Macaulay V (2006) Estimation of mutation rates and coalescence times: some caveats. In H–J Bandelt, Macaulay V e Richards M (eds) Human Mitochondrial DNA and the Evolution of *Homo sapiens*. Springer, Berlin, pp. 47-90.

Barbujani G, Magarini A, Minch E e Cavalli-Sforza LL (1997) An apportionment of human DNA diversity. Proc Natl Acad Sci USA 94:4516-4519.

Battilana J, Fagundes NJR, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvsin N, Krylov M, Benevolenskaya L, Arnett FC, Batzer MA, Deininger PL, Salzano FM e Bonatto SL (2006) *Alu* insertion polymorphisms in Native American and related Asian populations. Ann Hum Biol 33:142-160.

Battilana J, Cardoso-Silva L, Barrantes R, Hill K, Hurtado AM, Salzano FM e Bonatto SL (2007) Molecular variability of the 16p13.3 region in Amerindians and its anthropological significance. Ann Hum Genet 71:64-76.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannu PA, Scheer WD, Herrera RJ e Deininger PL (1994) African origin of human-specific polymorphic *Alu* insertions. Proc Natl Acad Sci USA 91:12288-12292.

Bernal V, Perez SI e Gonzalez PN (2006) Variation and causal factors of craniofacial robusticity in Patagonian hunter-gatherers from the late Holocene. Am J Hum Biol 18:748-765.

Bianchi NO, Catanesi CI, Baillet G, Martinez-Marignec VL, Bravi CM, Vidal-Rioja LB, Herrera RJ e Lopez-Camelo JS (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 63:1862-1871.

Bolnick DA, Shook BA, Campbell L e Goddard I (2004) Problematic use of Greenberg's linguistic classification of the Americas in studies of Native American genetic variation. Am J Hum Genet 75:519-522.

Bonatto SL e Salzano FM (1997a) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc Natl Acad Sci USA 94:1866-1871.

Bonatto SL e Salzano FM (1997b) Diversity and age of the four major mtDNA haplogroups and their implications for the peopling of the New World. Am J Hum Genet 61:1413-1423.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D e Ruiz-Linares A (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73:524-539.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Kidd JR e Cavalli-Sforza LL. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455-457.

Bräuer G (1992) Africa's place in the evolution of *Homo sapiens*. In: Bräuer G e Smith F (eds) Controversies in *Homo sapiens* Evolution. Balkema, Rotterdam, pp 83-98.

Brown WM, George M e Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76:1967-1971.

Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F e Wallace DC (1998) mtDNA haplogroup X: an ancient link between Europe/western Asia and North America? Am J Hum Genet 63:1852-1861.

Brumfield RT, Beerli P, Nickerson DA e Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history Trends Ecol Evol 18:249-256.

Callegari-Jaques SM, Salzano FM, Constans J e Maurieres P (1993) Gm haplotype distribution in Amerinds: relationship with geography and language. Am J Phys Anthrop 90:427-444.

Cann RL, Stoneking M e Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31-36.

Cavalli-Sforza LL, Piazza A e Menotti P (1994) History and Geography of Human Genes. Princeton University Press, Princeton, 1088p.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH e Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. Genome Res 15:1496-1502.

Coon CS (1962) The Origin of Races. Alfred A Knopf, New York, 724 p.

Currat M e Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. PLoS Biol 2:e421. http://biology.plosjournals.org/

Currat M e Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. Proc R Soc B 272:679-688.

Dornelles CL, Bonatto SL, Freitas LB e Salzano FM (2005) Is haplogroup X present in extant South American Indians? Am J Phys Anthropol 127:439-448.

Drummond AJ, Rambaut A, Shapiro B e Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22:1185-1192.

Duarte C, Maurício J, Pettitt PB, Souto P, Trinkaus E, van der Plicht H e Zilhão J (1999) The early Upper Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia. Proc Natl Acad Sci USA 96:7604-7609.

Eswaran V, Harpending HC e Rogers AR (2005) Genomics refutes an exclusively African origin of humans. J Hum Evol 49:1-18.

Evans PD, Mekel-Bobrov N, Vallender EJ, Hudson RR e Lahn BT (2006) Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic *Homo* lineage. Proc Natl Acad Sci USA 103:18178-18183.

Excoffier L (2002) Human demographic history: refining the recent African origin model. Curr Opin Genet Dev 12:675-682.

Fagundes NJR, Salzano FM, Batzer MA, Deininger PL e Bonatto SL (2005) Worldwide genetic variation at the 3'-UTR of the *LDLR* gene: possible influence of natural selection. Ann Hum Genet 69:389-400.

Fay JC e Wu C-I (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol 16:1003-1005.

Frissé L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J e Di Rienzo A (2001) Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am J Hum Genet 69:831-843.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ e Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229.

Garrigan D, Mobasher Z, Kingan SB, Wilder JA e Hammer MF (2005a) Deep haplotype divergence and long-range linkage disequilibrium at xp21.1 provide evidence that humans descend from a structured ancestral population. Genetics 170:1849-1856.

Garrigan D, Mobasher Z, Severson T, Wilder JA e Hammer MF (2005b) Evidence for archaic Asian ancestry on the human X chromosome. Mol Biol Evol 22:189-192.

Gonzáles-José R, Gonzáles-Martín A, Hernández M, Pucciarelli HM, Sardi M, Rosales A e van der Molen S (2003) Craniometric evidence for Palaeoamerican survival in Baja California. Nature 425:62-65.

González-José R, Bortolini MC, Santos FR e Bonatto SL (submetido) The peopling of the Americas: an interdisciplinary reappraisal. Curr Anthropol.

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G e Groves CP (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9:585-598.

Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M e Pääbo S (2006) Analysis of one million base pairs of Neanderthal DNA. Nature 444:330-336.

Greenberg JH, Turner II CG e Zegura SL (1986) The settlement of the Americas: a comparison of the linguistic dental and genetic evidence. Curr Anthropol 27: 477-497.

Haile-Selassie Y, Suwa G e White TD (2004) Late Miocene teeth from Middle Awash, Ethiopia, and early hominid dental evolution. Science 303:1503-1505.

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature 378:376-378.

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS e Clegg JB (1997) Archaic African *and* Asian lineages in the genetic ancestry of Modern Humans. Am J Hum Genet 60:772-789.

Hawks J, Hunley K, Lee S e Wolpoff MH (2000) Population bottlenecks and Pleistocene human evolution. Mol Biol Evol 17:2-22.

Hayakawa T, Aki I, Varki A, Satta Y e Takahata N (2006) Fixation of the human-specific CMP-N-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. Genetics 172:1139-1146.

Heller AH, Salzano FM, Barrantes R, Krylov M, Benevolenskaya L, Arnett FC, Munkhbat B, Munkhtuvshin N, Tsuji K, Hutz MH, Carnese FR, Goicoechea AS, Freitas LB e Bonatto SL (2004) Intra and intercontinental molecular variability of an Alu insertion in the 3'UTR of the LDLR gene. Hum Biol 76:591-604.

Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. PLoS Biol 3:e193. http://biology.plosjournals.org/

Hoffecker JF e Elias SA. Environment and archeology in Beringia. Evol Anthropol 12:34-49.

Ingman M, Kaessmann H, Pääbo S and Gyllensten U (2000) Mitochondrial genome variation and the origin of modern Humans. Nature 408:708-713.

Jorde LB, Banshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiac PA, Carpenter KD, Soodyall H, Jenkins T e Rogers AR (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523-538.

Jorde LB, Rogers AR, Bamshad M, Scott Watkins W, Krakowiak P, Sung S, Kere J and Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100-3103.

Kaessmann H, Heissdig F, von Haeseler A e Pääbo S (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat Genet 22:78-81.

Karafet T , Zegura L, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, Knijff P, Wiebe V, Griffiths RC, Templeton AR e Hammer MF (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. Am J Hum Genet 64:817-831.

Kern A (1998) Antecedentes Indígenas. Editora da Universidade Federal do Rio Grande do Sul, Porto Alegre, 139p.

Kidd JR, Black FL, Weiss KM, Balazs I e Kidd KK (1991) Studies of three Amerindian populations using nuclear DNA polymorphisms. Hum Biol 63:775-794.

Knowles LL e Maddison WP (2002) Statistical phylogeography. Mol Ecol 11:2623-2635.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M e Paabo S (1997) Neandertal DNA sequences and the origins of modern humans. Cell 90:19-30.

Kings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grossschmidt K, Possnert G, Paunovic M e Paabo S. (2000) A view of Neandertal genetic diversity. Nat Genet 26: 144-146.

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL e Oefner PJ (2006) The role of selection in the evolution of human mitochondrial genomes. Genetics 172:373-387.

Lahr MM e Foley R (1998) Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Ybk Phys Anthropol 41:137-176.

Leakey LSB (1966) *Homo habilis*, *Homo erectus* and the australopithecines. Nature 209:1279-81.

Leakey LSB e Goodall VM. Unveiling Man's Origins. Schenkman Publishing, Cambridge, 220 p.

Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, Troup GM e Wallace DC (1997) Y chromosome polymorphisms in Native Americans and Siberian populations: identification of Native American Y chromosome haplotypes. Hum Genet 100:536-543.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA e Wallace DC (2002) The dual origin and Siberian affinities of Native American Y chromosomes. Am J Hum Genet 70:192-206.

Lewin B (1998) Principles of Human Evolution: A Core Textbook. Blackwell Science, Malden, 526 pp.

Marjoram P e Tavare S (2006) Modern computational approaches for analysing molecular genetic variation data. Nat Rev Genet 7:759-770.

Marth G, Schuler G, Yeh R, Davenport R, Agawala R, Church D, Wheelan S, Baker J, Ward M, Kholodov M, Phan L, Czaberka E, Murval J, Cultier D, Wooding S, Rogers A, Chakravarti A, Harpending HC, Kwok P-Y e Sherry ST (2003) Sequence variations in public human genome data reflect a bottlenecked population history. Proc Natl Acad Sci USA 100:376-381.

Marth GT, Czabarka E, Murvai J e Sherry ST (2004) The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics 166:351-372.

Mateus Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Bonatto SL, Salzano FM e Herrera RJ (2005) Phylogenetic information in polymorphic L1 and *Alu* insertions from East Asians and Native American populations. Am J Phys Anthropol 128:171-184.

McDougall I, Brown FH e Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature 433:733-736.

Mellars P (2005) The impossible coincidence: a single-species model for the origins of modern human behavior in Europe. Evol Anthropol 14:12-27.

Meltzer DJ (1993) Pleistocene peopling of the Americas. Evol Anthropol 1:157-169.

Meltzer DJ (1997) Monte Verde and the Pleistocene peopling of the Americas. Science 276:754-755.

Merriwether DA, Rothhammer F e Ferrel RE (1995) Distributions of the four founding lineage haplotypes in Native Americans suggest a single wave of migration for the New World. Am J Phys Anthropol 98:411-430.

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A e Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100 :171-176.

Moilanen JS e Majamaa K (2003) Phylogenetic network and physicochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA. Mol Biol Evol 20 :1195-1210.

Nei M e Roychoudhury A (1982) Genetic relationship and evolution of human races. Evol Biol 14:1-59.

Neves WA e Hubbe M (2005) Cranial morphology of early Americans from Lagoa Santa, Brazil: implications for the settlement of the New World. Proc Natl Acad Sci USA 102 :18309-18314.

Neves WA e Pucciarelli HM (1991) The origin of the first Americans : an analysis based on the cranial morphlogy of early South American human remains. J Hum Evol 21 :261-273.

Neves WA, Munford D, Zanini MC e Pucciarelli HC (1999) Cranial morphological variation and the colonization of the New World: towards a four migration model? Ciência e Cultura 51:151-165.

Neves WA, Hubbe M, Okumura MM, González-José R, Figuti L, Eggers S e De Blasis PA (2005) A new early Holocene human skeleton from Brazil: implications for the settlement of the New World. J Hum Evol 48:403-414.

Neves WA, Hubbe M e Pilo LB (2007) Early Holocene human skeletal remains from Sumidouro Cave, Lagoa Santa, Brazil: history of discoveries, geological and chronological context, and comparative cranial morphology. J Hum Evol 52 :16-30.

Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Paabo S, Pritchard JK e Rubin EM (2006) Sequencing and analysis of Neanderthal genomic DNA. Science 314 :1113-1118.

Novick GE, Novick CC, Yunis J, Yunis E, Mayolo PA, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA e Herrera RJ (1998) Polymorphic Alu insertions and the Asian origin of Native American populations. Hum Biol 70:23-39.

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K e Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404 :490-493.

Pena SDJ, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B e Oyunsuren T (1995) A major founder Y-chromosome haplotype in Amerindians. Nat Genet 11:15-16.

Plagnol V e Wall JD (2006) Possible ancestral structure in human populations. PloS Genet 2 :e105. http://genetics.plosjournals.org/

Prous A (1995) Archeological analysis of the oldest settlements on the Americas. Braz J Genet 18:689-699.

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW e Cavalli-Sforza LL (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci USA 102 :15942-15947.

Ray N, Currat M, Berthier P e Excoffier L (2005) Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. Genome Res 15 :1161-1167.

Reich DE e Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. Proc Natl Acad Sci USA 95:8119-8123.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R e Lander ES. (2001) Linkage disequilibrium in the human genome. Nature 411 :199-204.

Rogers CA, Rogers LA e Martin CP (1992) How the door opened: the peopling of the New World. Hum Biol 64:281-302.

Roosevelt AC, Lima da Costa M, Lopes Machado C, Michab M, Merciek N, Valladas H, Feathers J, Barnett W, Inazio da Silveira M, Henderson A, Silva J, Chernoff B, Reese DS, Holman JA, Toth N e Schick K (1996) Paleoindian cave dwellers in the Amazon: the peopling of the Americas. Science 272:373-384.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA e Feldman MW (2002) Genetic structure of human populations. Science 298 :2381-2385.

Rougier H, Milota S, Rodrigo R, Gherase M, Sarcina L, Moldovan O, Zilhão J, Constantin S, Franciscus RG, Zollikofer CPE, Ponce de León M e Trinkaus E (2007) Pestera cu Oase 2 and the cranial morphology of early modern Europeans. Proc Natl Acad Sci USA 104:1165-1170.

Salzano FM (2007) The prehistoric colonization of the Americas. In: Crawford MH (ed) Anthropological Genetics: Theory, Methods, and Applications. Cambridge University Press, Cambridge, pp 433-455.

Santos FR, Rodriguez-Delfin L, Pena SDJ, Moore J e Weiss KM (1996) North and South Amerindians may have the same major founder Y chromosome haplotype. Am J Hum Genet 58:1369-1370.

Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH e Mitchell RJ (1999) The Central Siberian origin for Native American Y chromosomes. Am J Hum Genet 64:619-628.

Sardi ML, Novellino PS e Pucciarelli HM (2006) Craniofacial morphology in the Argentine Center-West: consequences of the transition to food production. Am J Phys Anthropol 130:333-343.

Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ e Altshuler D (2006) Calibrating a coalescent simulation of human genome sequence variation. Genome Res 15:1576-1583.

Schurr TG (2004) The peopling of the New World: Perspectives from molecular anthropology. Annu. Rev. Anthropol. 33:551-83.

Schurr TG, Ballinger SW, Gan Y-Y, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM e Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613-623.

Seielstad M, Yulldasheva N, Singh N, Underhill P, Oefner P, Shen P e Spencer Wells R (2003) A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. Am J Hum Genet 73:700-705.

Silva WA, Bonatto SL, Holanda AJ, Ribeiro-dos-Santos AKC, Paixão BM, Goldman GH, Abe-Sandes K, Rodriguez-Delfin L, Barbosa M, Paçó-Larson ML, Petzl-Erler ML, Valente V, Santos SEB e Zago MA (2003) Correction: Mitochondrial DNA variation in Amerindians. Am J Hum Genet 72:1346-1349.

Smith F (1992) The role of continuity in modern human origins. In: Bräuer G e Smith F (eds) Controversies in *Homo sapiens* Evolution. Balkema, Rotterdam, pp 145-156.

Stanford D e Bradley B (2002) Ocean trails and prairie paths? Thoughts about Clovis origins. In Jablonski NG (ed) The First Americans: The Pleistocene Colonization of the New World. Memoirs of the California Academy of Sciences, San Francisco, pp 255-271.

Starikovskaya Y, Sukernik RI, Schurr TG, Kogelnick AM e Wallace DC (1998) MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. Am J Hum Genet 63:1473-1491.

Steele DG e Powell JF (1993) Paleobiology of the first Americans. Evol Anthropol 1:138-146.

Strimmer K e Pybus OG (2001) Exploring the demographic history of DNA sequences using the generalized skyline plot. Mol Biol Evol 18:2298-2305.

Stringer C (2002) Modern human origins: progress and prospects. Phil Trans R Soc B 357: 563-579.

Takahata N, Lee S-H e Satta Y (2001) Testing multiregionality of modern human origins. Mol Biol Evol 18:172-183.

Tamura K e Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512-526.

Tarazona-Santos E e Santos FR (2002) The peopling of the Americas: a second major migration? Am J Hum Genet 70:1377-1380.

Tattersal I e Schwartz JH (1999) Hominids and hybrids: the place of Neanderthals in human evolution. Proc Natl Acad Sci USA 96:7117-7119.

Templeton AR (2002) Out of Africa again and again. Nature 416:45-50.

Templeton AR (2005) Haplotype trees and modern human origins. Am J Phys Anthropol 121:33-59.

The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299-1320.

Thorne A e Wolpoff MH (1992) The multiregional evolution of modern humans. Sci Am 266:76-83.

Torroni A, Schurr TG, Yang C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM e Wallace DC (1992) Native mitochondrial DNA analysis indicates that the Amerind and the Na-Dene populations were founded by two independent migrations. Genetics 130:153-162.

Torroni A, Achilli A, Macaulay V, Richards M e Bandelt H-J (2006) Harvesting the fruit of the human mtDNA tree. Trends Genet 22:339-345.

Trinkaus E (2007) European early modern humans and the fate of the Neanderthals. Proc Natl Acad Sci USA 104:7367-7372.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL e Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat genet 26:358-361.

Vigilant L, Stoneking M, Harpendinh HC, Hawkes K e Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503-1507.

Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR e Di Rienzo A (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc Natl Acad Sci USA 102:18508-18513.

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, Lynch JR, Falust AG, Weatherall DJ e Clegg JB. (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319:491-493.

Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol 37:613-623.

Waters MR e Stafford Jr TW (2007) Redefining the age of Clovis: implications for the peopling of the Americas. Science 315:1122-1126.

Weidenreich F (1946) Apes, Giants, and Man. University of Chicago Press, Chicago, 122 p.

Wolpoff MH, Hawks J, Frayer DW e Hunley K (2001) Modern human ancestry at the peripheries: a test of the replacement theory. Science 291:293-297.

Yu N, Zhao Z, Fu Y-X, Sambuughin N, Ramsay M, Jenkins T, Leskinen E, Patthy L, Jorde LB, Kuromori T e Li W-H (2001) Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol Biol Evol 18:214-222.

Yu N, Chen F-C, Ota S, Jorde LB, Pamilo P, Patthy P, Ramsay M, Jenkins T, Shyue S-K e Li W-H (2002) Larger genetic differences within Africans than between Africans and Eurasians. Genetics 161:269-274.

Zegura SL, Karafet TM, Zhivotovsky LA e Hammer MF (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21:164-175.

Zhao Z, Jin L, Fu Y-X, Ramsay M, Jenkins T, Leskinen E, Pamilo P, Trexler M, Patthy L, Jorde LB, Ramos-Onsins S, Yu N e Li W-H (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci USA 97:11354-11358.

Zhivotovsky LA, Rosenberg NA e Feldman MW (2003) Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. Am J Hum Genet 72:1171-1186.

Zilhão J (2006) Neandertals and moderns mixed, and it matters. Evol Anthropol 15:183-195.

CAPÍTULO VI

Anexos

VI. Anexos

Os trabalhos a seguir, embora não façam parte do corpo principal da tese, foram elaborados com a minha participação durante o período do doutorado. Como abordam o tema geral da tese, foram incluídos como apêndices à mesma.

O primeiro artigo, intitulado "Worldwide genetic variation at the 3'-UTR region of the LDLR gene: possible influence of natural selection" indica a ação de seleção balanceadora sobre este gene, além de mostrar evidências de alguma perda de variabilidade em relação ao mesmo nas populações nativas americanas.

O segundo "*Alu* insertion polymorphisms in Native Americans and related Asian populations" sugere a ausência de um efeito gargalo-de-garrafa forte durante o povoamento das Américas, bem como um papel importante para a deriva genética e o endocruzamento na formação da variabilidade genética apresentada por populações nativas americanas atualmente.

Finalmente, o terceiro trabalho, "Mitochondrial DNA and *Alu* insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay" sugere que os índios Ayoreo sofreram um efeito fundador que afetou sua variabilidade no mtDNA, mas não em marcadores de grupos sangüíneos + proteína ou de DNA autossômicos.

Anexo I

Worldwide genetic variation at the 3'-UTR of the *LDLR* gene: possible influence of natural selection.

Fagundes NJR, Salzano FM, Batzer MA, Deininger PL e Bonatto SL (2005)

Ann Hum Genet 69:389-400.

Anexo II

Alu insertion polymorphisms in Native American and related Asian populations.

Battilana J, Fagundes NJR, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvsin N, Krylov M, Benevolenskaya L, Arnett FC, Batzer MA, Deininger PL, Salzano FM e Bonatto SL (2006)

Ann Hum Biol 33:142-160.

Anexo III

Mitochondrial DNA and *Alu* insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay.

Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL e Salzano FM (2004)

Am J Hum Biol 16:479-488.