Universidade de São Paulo

Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto

Departamento de Química

Programa de Pós-Graduação em Química

Uso de marcadores informativos de ancestralidade para desenvolvimento de painéis de inferência de ancestralidade e identificação humana em populações miscigenadas

Vítor Matheus Soares Moraes

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de Mestre em Ciências, Área: **Química**

RIBEIRÃO PRETO -SP

2022

UNIVERSIDADE DE SÃO PAULO

FACULDADE DE FILOSOFIA CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

Vítor Matheus Soares Moraes

Uso de marcadores informativos de ancestralidade para desenvolvimento de painéis de inferência de ancestralidade e identificação humana em populações miscigenadas

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de Mestre em Ciências, Área: **Química**

Orientador: Prof. Dr. Celso Teixeira Mendes Junior

RIBEIRÃO PRETO - SP

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Moraes, Vitor Matheus Soares

Análise de marcadores informativos de ancestralidade em amostra de população brasileira e desenvolvimento de painéis de inferência de ancestralidade e identificação humana. Ribeirão Preto, 2022.

224p. : il. ; 30cm

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de Mestre em Ciências, Área: **Química**

Orientador: Prof. Dr. Celso Teixeira Mendes Junior

1. Ancestralidade; 2. Identificação Humana; 3. Genética Forense; 4. Polimorfismos de Nucleotídeo Único; 5. Imputação.

Nome: Moraes, Vítor Matheus Soares

Título: Uso de marcadores informativos de ancestralidade para desenvolvimento de painéis de inferência de ancestralidade e identificação humana em populações miscigenadas

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de Mestre em Ciências, Área: **Química**

Aprovado em:

Banca Examinadora

Prof. Dr. _____ Instituição:_____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição:_____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição:_____

Julgamento: _____ Assinatura: _____

Para minha avó Carmen, a consolidação do seu carinho.

Agradecimentos

Agradeço a meus pais, pelo apoio e suporte por toda minha vida, nada do que concluí ou que irei concluir seria possível sem o sr. Lineu e a sra. Adriana. Reconhecerei sempre o esforço de vocês.

Toda minha família merece meu obrigado por ter me apoiado desde sempre, entender os motivos da ausência e por torcerem por cada uma das minhas conquistas, meus tios e tias, primos, avô e minha falecida vó Carmen, que participou em grande parte do meu processo acadêmico.

Agradeço ao pessoal da turma 53 do bacharelado em química da Filô, todos os "trancos e barrancos" eram superados com risadas nesse time.

Agradeço a toda equipe do Laboratório de Pesquisas Forenses e Genômica, o tempo da pandemia não favoreceu nossos encontros, muita gente saindo, e gente nova entrando. Nada me impediu de conhecê-los, Alison, Malu, Tamara e Thássia, meus veteranos mais próximos, devo muito a vocês, espero eu servir de exemplo aos recém chegados como vocês foram para mim.

Prof. Dr. Celso, agradeço a oportunidade de me deixar fazer parte dessa equipe, o senhor é um exemplo de professor e de pessoa. É também uma inspiração para qualquer aluno que pise no seu laboratório. Muito obrigado e continue assim.

Agradeço a infraestrutura cedida pela Universidade de São Paulo, aos professores, funcionários e a coordenação do programa de pós graduação em química.

A CAPES, a CNPq e a FAPESP, agradeço o apoio financeiro.

Agradeço por fim, os membros da banca examinadora, por suas contribuições e pela graça me ajudarem com meu desenvolvimento como cientista.

Resumo

MORAES, V. M. S. **Uso de marcadores informativos de ancestralidade para desenvolvimento de painéis de inferência de ancestralidade e identificação humana em populações miscigenadas** 2022. 225 páginas. Dissertação de Mestrado – Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

A partir da introdução da PCR no final do século XX, tornou-se possível a análise dos *Short Tandem Repeats*, sendo até os dias de hoje o padrão para as análises periciais no âmbito de investigações policiais. Os marcadores mais abundantes no material genético humano são os polimorfismos de base única, chamados de SNPs, onde devido ao tamanho de seus amplicons, a possibilidade de trabalho com amostras degradadas demonstra uma vantagem em relação aos STRs. Porém, é necessária uma quantidade de aproximadamente 3 vezes mais marcadores, quando comparados aos microssatélites, para a obtenção de valores estatísticos viáveis para a utilização investigativa. Marcadores informativos de ancestralidade (AIMs) são polimorfismos que apresentam frequências alélicas elevadas em uma população e não em outras, sendo então utilizados para identificar as contribuições de ancestralidade genéticas individuais e populacionais de uma amostragem. São compostos por SNPs (AISNPs) em sua vasta maioria, podendo ser aplicados de maneira a identificar riscos epidemiológicos ou em investigações policiais. Apresentam também a possibilidade de uso para identificação humana em populações miscigenadas devido ao aumento da variabilidade genética. O objetivo do projeto é desenvolver dois painéis distintos que contenham entre 30-50 AIMs, para inferência de ancestralidade e para a identificação individual, para utilização na população brasileira e miscigenadas da américa latina utilizando 399 AIMs pertencentes aos painéis PIMA, SNP*for*ID 34-plex, *Kidd Lab set of 55 AISNPs, Seldin's* 128 AIMs e *Santos's* 192 AIMs em amostra de 513 indivíduos da população de Ribeirão Preto – SP. As amostras foram sequenciadas a partir do sistema *HaloPlex Enrichment System*, e genotipados a partir do *Infinium Multi-Ethnic Global-8 Array* da Illumina. Posições que não estavam presentes em nenhum desses ensaios foram imputadas a partir de três ferramentas diferentes, o *Sanger Imputation Server,* o *Michigan Imputation Server,* e o Impute2, posteriormente mesclados a um único output. A inferência de ancestralidade foi realizada pra cada painel selecionado e para os dois painéis desenvolvidos e intitulados Painel Ancestralidade e Painel Identificação

Humana, com as populações de referência europeias, africanas e asiáticas presentes no âmbito do *1000 Genomes Diversity Project* (1KGEN) e do *Human Genome Diversity Project* (HGDP). As populações miscigenadas foram obtidas no âmbito do 1KGEN e as populações ameríndias no âmbito do HGDP. As inferências de ancestralidade foram realizadas a partir dos softwares Admix 95 e STRUCTURE. Os parâmetros forenses foram obtidos através do software STRAF e as estatísticas populacionais e genéticas a partir dos programas GenAlEx e Plink. As PCAs foram obtidas a partir da função prcomp nativa da linguagem R de programação. Apenas 3 posições não foram imputadas dos 399 selecionados, pertencentes ao painel de Seldin. Para as PCAs, observou-se que todos os painéis apresentaram poder para separação e de agrupamento de todas as populações referência analisadas, quando utilizadas as três primeiras componentes principais. As inferências de ancestralidade populacional para as populações miscigenadas no âmbito do 1KGEN demonstrou resultados muito semelhantes, independente do painel utilizado, e compatíveis com a referência obtidas a partir do GWAS em Martin et al. (2007). Para a população de Ribeirão Preto os resultados também se mostraram muito próximos da referência de Marcorin (2019), onde somente os painéis *Kidd Lab set of 55 AISNPs* e Ancestralidade superestimaram a ancestralidade AFR e subestimaram a ancestralidade EUR da amostragem, porém demonstrado como um desvio continuo a todas as amostras individuais pelo STRUCTURE. A amostra de Ribeirão Preto também se demonstrou melhor analisada quando a metodologia tetra-parental foi utilizada. O painel SNP*for*ID 34-plex, quando utilizado em conjunto ao PIMA, e o *Kidd Lab set of 55 AISNPs*, demonstraram parâmetros forenses com informatividade próximas aos dos 13 STRs do CODIS e do SNP*for*ID 52-plex, quando aplicados na amostra brasileira. O Painel Identificação Humana demonstrou resultados superiores a estes dois painéis, apresentando informatividade próxima a dos 20 STRs do CODIS. Ambos os painéis desenvolvidos neste projeto apresentaram viabilidade de utilização para inferência de ancestralidade e identificação humana na américa latina.

Palavras-chave: Ancestralidade; Identificação Humana; Genética Forense; Imputação; Single Nucleotide Polymorphisms

Abstract

MORAES, V. M. S. **Use of ancestry informative markers for the development of panels for ancestry inference and human identification in mixed populations.** 2022. 225 pages. Dissertation (Master's Degree) – Faculty of Philosophy, Sciences and Letters at Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo, Brazil.

Since the introduction of PCR at the end of the 20th century, the analysis of Short Tandem Repeats became possible, which is still the standard for expert analysis in the context of police investigations. The most abundant markers in human genetic material are single-base polymorphisms, SNPs, where, due to the size of their amplicons, the possibility of working with degraded samples demonstrates an advantage over STRs. However, an amount of approximately 3 times more markers is required, when compared to microsatellites, to obtain viable statistical values for investigative use. Ancestry informational markers (AIMs) are polymorphisms that present high allele frequencies in one population and not in others, and are used to identify the individual and population genetic ancestry contributions of a sample. The vast majority are composed of SNPs (AISNPs) and can be applied in order to identify epidemiological risks or in police investigations. They also present the possibility of use for human identification in mixed populations due to the increase in genetic variability. The objective of the project is to develop two distinct panels ranging from 30 to 50 AIMs, to infer ancestry and for individual identification, which could be used in the Brazilian and admixed population of Latin America, using 399 AIMs belonging to the PIMA, SNPforID 34-plex, Kidd Lab set of 55 AISNPs, Seldin's 128 AIMs and Santos's 192 AIMs panels in a sample of 513 individuals from the population of *Ribeirão Preto* – SP. Samples were sequenced using the HaloPlex Enrichment System, and genotyped using the Illumina Infinium Multi-Ethnic Global-8 Array. Positions that were not present in any of these tests were imputed using three different tools, Sanger Imputation Server, Michigan Imputation Server, and Impute2, later merged into a single output. Ancestry inference was performed for each selected panel and for the two panels developed and entitled *Painel Ancestralidade* and *Painel Identificação Humana,* with European, African and Asian reference populations present within the scope of the 1000 Genome Diversity Project (1KGEN) and the Human Genome Diversity Project (HGDP). The admixed populations were obtained under the 1KGEN and the Amerindian populations under the HGDP. Ancestry inferences were performed using Admix 95 and STRUCTURE software. The forensic parameters were obtained through the STRAF software and the population and genetic statistics from the GenAlEx and Plink software. The

PCAs were obtained from the *prcomp* function native to the R programming language. Only 3 positions were not assigned from the 399 selected, belonging to Seldin's panel. For PCAs, it was observed that all panels showed power for separation and grouping of all reference populations analyzed, when using the first three main components. Population ancestry inferences for mixed populations under 1KGEN showed very similar results, regardless of the panel used, and consistent with the reference obtained from the GWAS in Martin et al. (2007). For the population of *Ribeirão Preto*, the results were also very close to the reference by Marcorin (2019), where only the Kidd Lab set of 55 AISNPs and *Painel Ancestralidade* overestimated the AFR ancestry and underestimated the EUR ancestry of the sample, but demonstrated as a continue deviation to all individual samples by STRUCTURE. The sample from Ribeirão Preto was also better analyzed when the tetra-parental methodology was used. The SNPforID 34-plex panel, when used together with PIMA, and the Kidd Lab set of 55 AISNPs, demonstrated forensic parameters with informativeness close to those of the 13 CODIS STRs and the SNPforID 52-plex, when applied to the Brazilian sample. The *Painel Identificação Humana* showed superior results to these two panels, presenting informativity close to that of the 20 CODIS STRs. Both panels developed in this project showed feasibility of use for inference of ancestry and human identification in Latin America.

Keywords: Ancestry; Forensic Genetics; Human Identification; Imputation; Single Nucleotide Polymorphisms

Lista de Figuras

Figura 1. Proporção da população imigrante na cidade de Ribeirão Preto- SP nos anos de: (a) 1886, com população de aproximadamente 10.425 indivíduos e (b) 1902, com aproximadamente 53.418 habitantes (Lanza & Lamounier, 2015)..... 28

Figura 2. Análise de componentes principais baseada em 25 dos 26 marcadores que compõem o painel PIMA, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (65,92%) vs PC2 (21,49%); (b) PC1 vs PC3 (5,73%)..... 48

Figura 3. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel PIMA. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513)..... 51

Figura 4 Análise de componentes principais baseada nos 34 marcadores que compõem o painel SNP*for*ID 34-plex, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (46,83%) vs PC2 (38,42%); (b) PC1 vs PC3 (5,11%)..... 54

Figura 5 Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel SNP*for*ID 34-plex. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513)..... 57

Figura 6 Análise de componentes principais baseada nos 55 marcadores que compõem o painel *Kidd Lab set of 55 AISNPs*, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (53,12%) vs PC2 (31,09%); (b) PC1 vs PC3 (8,87%)..... 60

Figura 7 Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel *Kidd Lab set of 55 AISNPs*. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513)..... 63

Figura 8 Análise de componentes principais baseada nos 59 marcadores que compõem o painel PIMA + 34-plex, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (53,12%) vs PC2 (31,09%); (b) PC1 vs PC3 (8,87%)..... 66

Figura 9. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel PIMA

+ 34-plex. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 69

Figura 10. Análise de componentes principais baseada em 125 dos 128 marcadores que compõem o painel desenvolvido por Seldin, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (49,00%) vs PC2 (31,50%); (b) PC1 vs PC3 (8,08%). 72

Figura 11. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel desenvolvido por Seldin. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 75

Figura 12. Análise de componentes principais baseada nos 192 marcadores que compõem o painel desenvolvido por Santos, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (70,12%) vs PC2 (16,37%); (b) PC1 vs PC3 (6,13%). 78

Figura 13. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel desenvolvido por Santos. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 81

Figura 14. Análise de componentes principais baseada nos 396 marcadores presentes no trabalho, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (70,12%) vs PC2 (16,37%); (b) PC1 vs PC3 (6,13%). 84

Figura 15. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o conjunto de todos os painéis do trabalho (396 AIMs). As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 87

Figura 16 Análise de componentes principais baseada nos 33 marcadores presentes no Painel Ancestralidade, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (45,19%) vs PC2 (40,36%); (b) PC1 vs PC3 (4,81%). 94

Figura 17. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o Painel Ancestralidade (33 AIMs). As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 97

Figura 18. Análise de componentes principais baseada nos 50 marcadores presentes no Painel Identificação Humana, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas

(verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (54,18%) vs PC2 (31,22%); (b) PC1 vs PC3 (6,97%). 100

Figura 19. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o Painel Identificação Humana (50 AIMs). As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 103

Lista de Tabelas

Tabela 1. Ancestralidade biogeográfica inferidas para a população de Ribeirão Preto (Marcorin, 2021) e para as populações miscigenadas presentes no 1KGEN (Martin et al., 2017). 39

Tabela 2. Compatibilidade entre as posições genotipadas a partir do MEGA e do NGS, e das posições imputadas por diferentes metodologias. 46

Tabela 3. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel PIMA. 49

Tabela 4. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise de 25 dos 26 marcadores que compõem o painel PIMA. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 52

Tabela 5. Parâmetros forenses combinados para as populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto (*n* = 343) utilizando o painel PIMA (25 AIMs). 53

Tabela 6. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel SNP*for*ID 34-plex. 55

Tabela 7. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 34 marcadores que compõem o painel SNP*for*ID 34-plex. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 58

Tabela 8. Parâmetros forenses combinados para as populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto (*n* = 343) utilizando o painel SNP*for*ID 34-plex. 59

Tabela 9. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel *Kidd Lab set of 55 AISNPs*. 61

Tabela 10. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 55 marcadores que compõem o painel *Kidd Lab set of 55 AISNPs*. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do

HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 64

Tabela 11. Parâmetros forenses combinados para as populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto (*n* = 343) utilizando o painel *Kidd Lab set of 55 AISNPs*. 65

Tabela 12. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel PIMA + 34-plex. 67

Tabela 13. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 59 marcadores que compõem o painel PIMA + 34-plex. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 70

Tabela 14. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o painel PIMA + 34-plex (59 AIMs)..... 71

Tabela 15. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel desenvolvido por Seldin. 73

Tabela 16. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise de 125 dos 128 marcadores que compõem o painel desenvolvido por Seldin. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 76

Tabela 17. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o painel desenvolvido por Seldin (125 AIMs). 77

Tabela 18. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel desenvolvido por Santos..... 79

Tabela 19. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 192 marcadores que compõem o painel desenvolvido por Santos. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do

HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 82

Tabela 20. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o painel desenvolvido por Santos (192 AIMs)..... 83

Tabela 21. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os 396 marcadores presentes no trabalho..... 85

Tabela 22. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 396 marcadores que compõem o conjunto de todos os painéis do trabalho. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 88

Tabela 23. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o conjunto de todos os painéis presentes no trabalho (396 AIMs)..... 89

Tabela 24. Lista dos marcadores informativos de ancestralidade que melhor individualizam as populações de referência, de acordo com suas frequências alélicas. 90

Tabela 25. Lista dos marcadores informativos de ancestralidade que melhor discriminam os indivíduos de Ribeirão Preto, de acordo com os valores de PM e PD. 92

Tabela 26. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os 33 marcadores presentes no Painel Ancestralidade..... 95

Tabela 27. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 33 marcadores que compõem o Painel Ancestralidade. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 98

Tabela 28. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o Painel Ancestralidade (33 AIMs)..... 99

Tabela 29. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os 50 marcadores presentes no Painel Identificação Humana. 101

Tabela 30. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 50 marcadores que compõem o Painel Identificação Humana. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513). 104

Tabela 31. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando Painel Identificação Humana (50 AIMs). 105

Tabela 32. Variância observada para cada componente principal em todos os painéis analisados aplicados às populações de referência (1KGEN + HGDP). 110

Tabela 33. Diferença absoluta entre as inferências de ancestralidade populacional utilizando o software Admix 95 e os parâmetros de referência (Martin et al., 2007 e Marcorin, 2021). 115

Tabela 34. Diferença absoluta entre as inferências de ancestralidade populacional utilizando o software STRUCTURE e os parâmetros de referência (396 SNPs ou resultados de Marcorin, 2021). 118

Tabela 35. Desvio da clusterização ideal das populações de referência utilizando o software STRUCTURE. Os clusters são atribuídos como: 1- AFR, 2- AMR, 3- EAS e 4- EUR. 121

Tabela 36. Parâmetros forenses obtidos para os 343 indivíduos da população de Ribeirão Preto coletados de maneira randômica. 123

Sumário

1. Introdução 21

1.1 Genética Forense 21

1.2 Marcadores Genéticos Usados na Prática Forense 21

1.3 Marcadores Informativos de Ancestralidade 23

1.4 Bancos de Dados Genômicos Globais 25

1.5 População de Ribeirão Preto 27

2. Justificativa 30

3. Hipóteses 32

4. Objetivos 34

5. Materiais e Métodos 36

5.1 Sequenciamento e Genotipagem 36

5.2 Imputação 37

5.3 Inferência de Ancestralidade 38

5.3.1 ADMIX 95 40

5.3.2 STRUCTURE 40

5.3.3 Desenvolvimento de Painéis 42

5.4 Parâmetros Forenses 42

5.5 Estatísticas Genéticas e Populacionais 43

6. Resultados 46

6.1 Imputação 46

6.2 Estatísticas Populacionais 47

6.3 PIMA (25 AIMs) 48

6.4 SNP*for*ID 34-plex (34 AIMs) 54

6.5 Kidd's Lab Set of 55 AIMs 60

6.6 PIMA + 34-plex (59 AIMs) 66

6.7 Seldin's 128 AIMs (125 AIMs) 72

6.8 Santos's 192 AIMs (192 AIMs) 78

6.9 Conjunto de Todos os Painéis (396 AIMs) 84

6.9.1 Seleção dos Painéis 90

6.10 Painel Ancestralidade (33 AIMs) 94

6.11 Painel Identificação Humana (50 AIMs) 100

7. Discussão 107
7.1 Imputação 107
7.2 Desvios do Equilíbrio de Hardy-Weinberg 108
7.3 PCAs 109
7.4 Inferência de Ancestralidade 113
7.5 Parâmetros Forenses 122
8. Conclusões 126
9. Referências 129
Apêndices 140
Apêndice A 140
Apêndice B 164
Apêndice C 166
Anexo A – Termo de consentimento livre e esclarecido 186
Anexo B – Termo de consentimento livre e esclarecido para armazenamento de material biológico 189
Anexo C – Protocolo de análise laboratorial (HaloPlex) para sequenciamento de nova geração ... 192
Anexo D – Protocolo de processamento computacional de dados de sequenciamento de nova geração 200
Anexo E – Imputação 218

Introdução

1. Introdução

1.1 Genética Forense

O início da genética forense é dado ao fim do século XIX e no começo do século XX, com os estudos de Karl Landsteiner a respeito dos grupos sanguíneos, o que passou a tornar possível a discriminação e o agrupamento de indivíduos (Kyle & Shampo, 2001).

Na segunda metade do século XX, iniciou-se a análise de polimorfismos de DNA. Isso se tornou possível pelo desenvolvimento do sequenciamento de Sanger (Sanger et al., 1977) e pela utilização da técnica de *Southern Blotting* por Alec Jeffreys, que acabou por desenvolver o que chamou de DNA *fingerprinting*. Enquanto o sequenciamento permite avaliar sítios polimórficos presentes em uma sequência de DNA, a técnica de DNA *fingerprinting* permite identificar fragmentos compostos por repetições de nucleotídeos altamente polimórficas chamadas de "Número Variável de Repetições em Tandem" (VNTRs) (Jeffreys et al., 1985).

Nas duas últimas décadas do mesmo século, a *Polymerase Chain Reaction* (PCR) trouxe um avanço significativo para a genética forense. Um procedimento que permitiu a amplificação de regiões específicas do DNA, mesmo que o material genético encontre-se em quantidades muito pequenas e até mesmo degradado, o que acabou por resolver dois dos problemas encontrados nas análises de VNTRs: a necessidade de uma grande quantidade de DNA e da integridade desse material genético, devido ao tamanho de seus *amplicons* (Goodwin, 2015). Assim, nos anos 90, as primeiras análises de *Short Tandem Repeats* são realizadas (Hagelberg et al., 1991; Jeffreys et al., 1992), sendo esses os tipos de marcadores mais utilizados na investigação forense atual (Phillips et al., 2019; Zhang et al., 2020).

1.2 Marcadores Genéticos Usados na Prática Forense

Marcadores genéticos são sequências específicas presentes no DNA que trazem informações de saúde, variabilidade genética e diferenças envolvendo bases nitrogenadas entre indivíduos e/ou regiões etnogeográficas amplamente utilizados na identificação humana, devido a sua grande capacidade de distinguir indivíduos, sendo os tipos de marcadores mais utilizados os *Short Tandem Repeats* (*STRs*), os *Single Nucleotide Polymorphysms* (*SNPs*) e as inserções/deleções (InDels) (Grover & Sharma, 2016).

Regiões do DNA com repetições que possuem entre 2pb e 7pb em comprimento são chamados de microssatélites, repetições de sequência simples (SSRs) ou, de maneira mais frequente, repetições curtas em tandem (STRs), que são facilmente amplificados por PCR (Butler & Hill, 2012). Devido à elevada diversidade alélica e à facilidade de determinação dos mesmos por eletroforese, os STRs consistem nos marcadores mais empregados para obtenção de perfis de DNA voltados para a identificação humana.

Uma variação de base única entre duas sequências homólogas é comumente referida como polimorfismo de nucleotídeo único (SNP). São os marcadores mais abundantes em todo o genoma humano e utilizados regularmente em estudos de doenças genéticas (Brookes, 1999). Tecnologias vem sendo desenvolvidas para a redução de trabalho e automação dos processos de análise de SNPs (Ku et al., 2010).

A informatividade forense observada em conjuntos de STRs é bem maior do que a apresentada pelos SNPs, quando apresentados em mesmo número, sendo necessário cerca de 40~60 SNPs para que se atinja o poder estatístico observado por 13~15 STRs (Butler et al., 2007). Apesar da menor informatividade, muitas vantagens existem na utilização de SNPs em investigações forenses, como exemplos: o tamanho do amplicon, que por ser pequeno (< 100pb) torna possível a obtenção de dados mesmo de amostras degradadas; baixa taxa de mutação, cerca de uma mudança de ordem a cada 10^8 gerações, o que minimiza a probabilidade de se observar inconsistências em um exame de paternidade envolvendo o pai biológico; o tratamento e análise de dados pode ser totalmente automatizada, pois uma separação baseada em tamanho não é necessária; e a habilidade de predizer ancestralidade e fenótipos morfológicos (Pontes et al., 2015).

Inserções/deleções (InDels) consistem no segundo tipo de variação genética mais encontrada no DNA, porém recebe menor atenção do que a dada para outros marcadores (Say, 2017). São polimorfismos de comprimento, caracterizados pela inserção ou deleção de um ou mais nucleotídeos em uma determinada região do genoma. Como os SNPs, apresentam baixa informatividade, sendo necessário um grande número de InDels para um resultado semelhante a um ensaio utilizando STRs (Tozzo et al., 2021). Entretanto, apresentam a maioria das vantagens anteriormente relatadas para os SNPs, além de poder ser analisado por PCR seguida de eletroforese.

1.3 Marcadores Informativos de Ancestralidade

Para que seja possível a predição da ancestralidade de um indivíduo, ou seja, determinar qual ou quais ascendências contribuíram para que tal indivíduo apresente seu genótipo final, é necessário que os marcadores genéticos selecionados possuam uma diferenciação grande entre populações autóctones. O estudo de um marcador em potencial para identificação de ancestralidade é feito a partir das diferenças de frequências alélicas entre populações, o que resulta em elevado índice de fixação (F_{ST}) entre as populações consideradas (Cheung et al., 2019).

Os SNPs e as InDels são os marcadores de escolha para ensaios envolvendo ancestralidade, mais especificamente os marcadores informativos de ancestralidade (AIMs). AIMs são polimorfismos genéticos que apresentam frequências alélicas divergentes entre grupos étnicos distintos entre si, populações europeias e as asiáticas, por exemplo, e que contribuem para identificar a ancestralidade individual e populacional de um determinado grupo de indivíduos (Pereira et al., 2017). Embora qualquer tipo de marcador possa apresentar tais propriedades, a maioria dos AIMs consiste em SNPs e InDels.

Para que um conjunto de SNPs informativos de ancestralidade (AISNPs) seja considerado uma boa ferramenta, ele deve nos trazer informações a respeito da relação entre uma amostra desconhecida e as regiões biogeográficas dos seus ancestrais, tendo uma boa probabilidade de que essa relação esteja correta (Barnholtz-Sloan et al., 2005; Kidd et al., 2014). Dessa maneira, genotipando um painel de pelo menos 30 AIMs bem escolhidos para aplicação em uma população, é possível, através das suas frequências alélicas, calcular as proporções de ancestralidade genômica que remetem aos diferentes continentes (Cheung et al., 2019).

Além de identificar a ancestralidade de uma pessoa, os AIMs são de extrema importância na identificação de pessoas desaparecidas e vítimas de desastres em massa, com a utilização da ancestralidade biogeográfica, e posterior cálculo probabilístico para identificação da localidade de origem do indivíduo (Ambers et al., 2020; Latham & Miller, 2019). Os AIMs também podem ser utilizados em estudos de associação caso-controle, onde os índices de ancestralidade estimados podem contribuir para identificar possíveis falsas associações (Pereira et al., 2017). Portanto, a ancestralidade individual pode ser utilizada como fator de ajuste em estudos de associação (Hardy et al., 2020). Outra importância dos

AISNPs é a utilização em estudos epidemiológicos, onde é possível identificar possíveis predisposições a doenças genéticas biogeograficamente dependentes (Durães et al., 2020; Pérez-Mayoral et al., 2019).

Muitos dos SNPs utilizados para predição de ancestralidade genética são encontrados em regiões codificadoras que foram subjugadas a pressões de seleção, o que inclui genes de pigmentação da pele e de reação a xenobióticos (Pereira et al., 2020). Sendo assim, é observada a presença de marcadores de pigmentação na maioria dos conjuntos de ancestralidade atualmente utilizados. O painel 34-plex (Amigo et al., 2008), bem como o painel desenvolvido por K.K. Kidd contendo 55 SNPs (Kidd et al., 2014), por exemplo, incluem os SNPs rs12913832 (*HERC2*), rs1426654 (*SLC24A5*) e rs16891982 (*SLC45A2*), que estão associados à pigmentação da pele, dos olhos e do cabelo.

O sistema SNP*for*ID 34-Plex (Phillips et al., 2007), foi um dos primeiros painéis de AIMs a ser introduzido. Tal painel é composto por 34 marcadores informativos de ancestralidade independentes, selecionados com a função de distinguir indivíduos de ancestralidades Africanas (AFR), Europeias (EUR) e Leste-Asiáticas (EAS) para utilização em investigações policiais (Pfaffelhuber et al., 2020). Consiste, provavelmente no conjunto de AIMs mais utilizado pela comunidade científica.

Outro conjunto utilizado em um grande número de trabalhos (Pakstis et al., 2017, 2019; Phillips et al., 2019), o sistema *KiddLab set of 55 AISNPs*, baseado em SNPs bialélicos, foi desenvolvido para determinação de ancestralidade individual e populacional, sendo testado em uma grande diversidade de amostras populacionais disponíveis no HapMap e HGDP-CEPH (Kidd et al., 2014). Esse conjunto vem sendo adicionado a kits comercializados pela *Thermo Fisher Scientific* (*Precision ID Ancestry Panel*) e *Illumina* (*ForenSeq*^{TM}) (Pakstis et al., 2015).

O *Precision ID Ancestry Panel* também inclui 123 marcadores dentre um conjunto de 128 AIMs desenvolvido com o objetivo de avaliar a ancestralidade continental e proporções de mistura em populações das Américas (Kosoy et al., 2009). Foi desenvolvido com base em indivíduos europeus, leste-asiáticos, ameríndios, africanos, sul-asiáticos, mexicanos e porto-riquenhos, como uma simples e eficiente ferramenta para estudos de associação em populações etnicamente diversas. O teste para otimização e validação do conjunto foi feito utilizando 825 indivíduos (Kosoy et al., 2009).

Em 2016, um conjunto de 192 AIMs foi desenvolvido baseando-se exclusivamente na população brasileira, com o objetivo de determinar a ancestralidade com uma melhoria do custo-benefício e da relação tempo-eficiência (Santos et al., 2016). A partir de um total de 365.116 SNPs, foram selecionados 192 marcadores com elevada capacidade de separação das populações Africanas, Europeias e Nativo-Americanas e que resultou em elevada acurácia para fins de determinação de ancestralidade.

Por fim, muito recentemente um conjunto de AIMs foi criado com o foco nas populações do continente americano, o *Population Informative Multiplex for the Americas* (PIMA), onde o foco principal consistiu em selecionar marcadores que apresentassem grande diferencial de frequências alélicas entre as populações nativo-americanas e populações africanas, europeias e leste-asiáticas. O PIMA foi desenvolvido como uma ferramenta complementar ao conjunto 34-Plex mencionado anteriormente. É composto por 26 SNPs autossômicos, um SNP do cromossomo X e o marcador de sexo amelogenina (Gontijo et al., 2020).

Normalmente os AIMs são pouco eficientes no que se refere à identificação humana em populações caracterizadas pela ausência de miscigenação, uma vez que os genótipos são pouco variáveis em tais populações devido à baixa diversidade genética dentro do mesmo continente. Entretanto, além da aplicação em análise de ancestralidade, foi demonstrado que AIMs podem também ser utilizados para identificação individual em populações miscigenadas (Oliveira et al., 2009), visto que o diferencial de frequência alélica intercontinental muitas vezes é superior a 30-50%, o que garante a existência de elevada diversidade em populações caracterizadas por múltiplas ancestralidades. Porém, é necessário um número maior de marcadores para que a informatividade forense equivalente à de um conjunto de STRs seja atingida (Oliveira et al., 2009). Dessa maneira, a análise de marcadores informativos de ancestralidade tem se tornado altamente relevante em populações miscigenadas, com o propósito adicional de identificar a origem dos indivíduos em casos policiais (Tvedebrink & Eriksen, 2019).

1.4 Bancos de Dados Genômicos Globais

O conjunto de tecnologias de Sequenciamento de Nova Geração (NGS, do inglês, *Next-Generation Sequencing*) ou Sequenciamento Massivo em Paralelo (MPS, do inglês,

Massively Parallel Sequencing), permitem o sequenciamento em paralelo de milhões de sequências simultaneamente, de modo rápido e com custos bem inferiores, tornando possível sequenciar um genoma humano em até mesmo um dia. Diversas iniciativas de grande porte, envolvendo sequenciamento em larga escala, somente se tornaram possíveis devido ao surgimento de tecnologias NGS. Dentre as principais iniciativas voltadas para o estudo da diversidade genética global, destacam-se o *The 1000 Genomes Project* (1KGEN) (Auton et al., 2015), o *The Simons Genome Diversity Project* (SGDP) (Mallick et al., 2016) e o *The Human Genome Diversity Project* (HGDP) (Bergström et al., 2020).

O *The 1000 Genomes Project* (Auton et al., 2015) possui um total de 2504 genomas sequenciados em alta cobertura (30X) de indivíduos de 26 diferentes populações do mundo. O projeto foi focado na África, Europa, Ásia e Latino-Americanos. Tinha como objetivo identificar variações genéticas comuns ao realizar sequenciamentos de todo o genoma de indivíduos provenientes de diversas populações, detectando ao menos 95% dos sítios em que o alelo menos frequente apresentasse frequência superior a 1% (The 1000 Genomes Project Consortium, 2010).

O *The Simons Genome Diversity Project* (Mallick et al., 2016) apresenta 300 genomas analisados por sequenciamento de alta cobertura (43X), provenientes de 142 populações diferentes. Em comparação ao *The 1000 Genomes Project*, o SGDP traz dados obtidos de uma diversidade maior de populações distintas de outras partes do mundo, incluindo as populações indígenas de todos os continentes habitados.

Sabendo que sequências genômicas provenientes de diversos grupos populacionais humanos podem revelar padrões globais de diversidade genética em nossa espécie e elucidar a história demográfica de diferentes populações, fornecendo subsídios para desenvolvimento e interpretação de dados gerados em estudos médicos, o *The Human Genome Diversity Project* (Bergström et al., 2020) analisou um total de 929 genomas sequenciados em alta cobertura, provenientes de 54 populações humanas diversas e distintas, com foco em populações do sul e do centro da África, Américas e Oceania.

Estes bancos de dados armazenam e disponibilizam acesso, portanto, a dados genotípicos de populações caracterizadas por diferentes padrões de ancestralidade, quer sejam populações autóctones ou populações oriundas de diferentes padrões de miscigenação, o que os tornam fundamentais em estudos voltados para a ancestralidade genômica.

1.5 População de Ribeirão Preto

A população residente em terras brasileiras antes da colonização europeia era estimada ser em torno de dois milhões de nativos. Antes da década de 1820, os descendentes de portugueses eram a maioria, embora 3,5 milhões de africanos (oeste, centro-oeste e sudeste da África) tenham sido trazidos como escravos. (Ferreira et al., 2006). Entre as décadas de 1890 e 1920, o governo do estado de São Paulo desenvolveu uma política de subsidiar a imigração europeia, que ocasionou a criação de comunidades de imigrantes no Sul e Sudeste do Brasil, sendo em sua maioria italianos, portugueses, espanhóis e alemães. Adicionalmente, em 1908, um grande número de japoneses chegou no país. De acordo com o portal do Governo de São Paulo (https://www.saopaulo.sp.gov.br/ultimas-noticias/estado-tem-cerca-de-1-milhao-de-japoneses-e-descendentes, acesso em 29/06/2022, às 19h33), atualmente vivem no Brasil cerca de 2 milhões de japoneses, onde 1,3 milhão estão presentes no estado de São Paulo. Até o ano de 1920, um total de 30 diferentes nacionalidades estavam presentes em solos brasileiros (Craig & Faria, 2021). A **Figura 1** traz o número de imigrantes e suas origens para os anos de 1886 e 1902 em Ribeirão Preto, onde em 1902, com uma população de 53.418 habitantes, 62% dos moradores eram imigrantes e 96,7% desses imigrantes eram europeus, em sua vasta maioria composto por italianos.

Em Ribeirão Preto – SP, a partir da década de 1870, houve o início da grande expansão do café. A facilidade trazida pela ferrovia da Companhia Mogiana, o início do período de industrialização, e a oferta de trabalho em grandes fazendas de café, em conjunto com a imigração em massa desse período, contribuíram para o afluxo de italianos, portugueses, espanhóis, austríacos e africanos, sendo estes últimos trazidos como escravos. Sabe-se que durante o período entre 1901-1940, 95.687 imigrantes chegaram na região de Ribeirão Preto, Cravinhos, Sertãozinho e Pontal (Lanza & Lamounier, 2015).

No censo de 2010 a população era de 604.682 indivíduos, sendo eles autodeclarados 69,8% brancos, 22,8% pardos, 6,4% pretos, 0,9% amarelos e 0,1% indígenas (http://produtos.seade.gov.br/produtos/retratosdesp/view/index.php?temaId=1&indId=5&locId=3543402&busca=R, acesso em 18/07/2022, às 13h54).

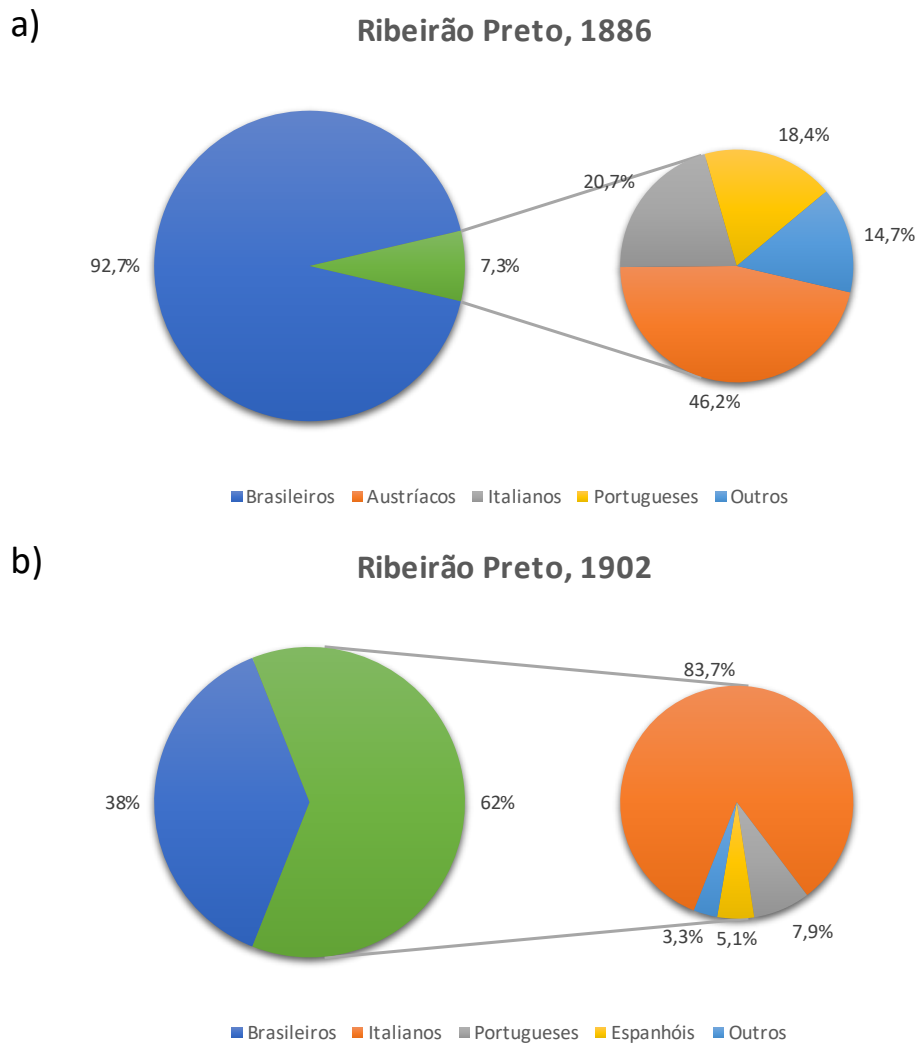


Figura 1. Proporção da população imigrante na cidade de Ribeirão Preto- SP nos anos de: (a) 1886, com população de aproximadamente 10.425 indivíduos e (b) 1902, com aproximadamente 53.418 habitantes (Lanza & Lamounier, 2015).

Justificativa

2. Justificativa

O uso de SNPs em análises policiais está se tornando essencial, já que os STRs apresentam dificuldades de análise quando o assunto é DNA degradado (Gettings et al., 2015; Phillips et al., 2008). A otimização para o menor número de marcadores informativos de ancestralidade que apresente uma eficiência adequada é importante para o custo benefício de análises laboratoriais e para a facilitação desses ensaios (Santos et al., 2016).

O desenvolvimento de um painel de marcadores informativos de ancestralidade, com um poder de distinção entre ancestralidades asiáticas e nativo americanas demonstra benefícios nas correções em estudos de associação (Andrade et al., 2018) e na inclusão social dos indivíduos que detenham tais descendências para a população do estado de São Paulo.

Hipóteses

3. Hipóteses

Painéis de AISNPs possuem informatividade adequada para uso simultâneo em atividades de inferência de ancestralidade e de identificação humana, seja para investigações criminais ou em casos de pessoas desaparecidas.

A população de Ribeirão Preto e do estado de São Paulo apresenta uma miscigenação tetra- ou poli-híbrida.

Objetivos

4. Objetivos

O projeto tem por objetivo geral montar e validar dois painéis de AIMs para utilização em simples ensaios laboratoriais de análises de inferência de ancestralidade e para identificação individual em população brasileira e outras populações miscigenadas da América Latina, através de análise de uma amostra constituída por 513 indivíduos de Ribeirão Preto- SP, a partir de cinco conjuntos de AIMs previamente desenvolvidos. Como objetivos específicos, propomos:

- Determinar para cada marcador, as frequências alélicas, genotípicas e aderência ao equilíbrio de Hardy-Weinberg;
- Testar a ancestralidade genômica populacional e individual da amostra populacional brasileira e de outras populações miscigenadas presentes nos repositórios do 1KGEN e HGDP, com os diferentes conjuntos de marcadores;
- Avaliar a acurácia das estimativas de ancestralidade através de comparação com as estimativas de ancestralidade individual geradas por 336.112 marcadores presentes no *Multi-Ethnic Global-8 array* realizado por Letícia Marcorin (Marcorin, 2021), *Illumina inc.*, realizadas anteriormente pela equipe do laboratório em contribuição com os professores Esteban Parra, Universidade de Toronto e Heather Norton, Universidade de Cincinnati.
- Estimar parâmetros forenses do conjunto aqui desenvolvido e dos cinco conjuntos originais, a fim de avaliar aplicabilidade de AIMs na atividade de identificação humana para investigações criminais.

Materiais e Métodos

5. Materiais e Métodos

A coleta de amostras e obtenção dos dados para realização deste presente projeto de pesquisa foi aprovada em seus aspectos éticos pelo Comitê de Ética em Pesquisa da FFCLRP-USP, de acordo com o Processo CEP-FFCLRP CAAE n.º 25696413.7.0000.5407. Foram avaliados 513 indivíduos amostrados junto ao Hemocentro e ao Hospital das Clínicas de Ribeirão Preto – SP, bem como no Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, sendo que 343 pessoas foram amostradas de maneira randômica enquanto que o restante foi coletado de maneira a priorizar fenótipos de pigmentação mais raros, de acordo com necessidade de outros projetos do laboratório. Para cada participante coletou-se, ainda, dados referentes à autodeclaração de cor/raça (abordagem do IBGE) e ancestralidade por meio de preenchimento de questionário (**Anexos A e B**).

O DNA das amostras de sangue foi extraído através de protocolo modificado de s*alting-out* (Miller et al., 1988). Posteriormente foi feita a quantificação do DNA por meio de ensaios espectrofotométricos (*NanoDrop™ 1000 Spectrophotometer – Thermo Scientific*) e fluorimétricos (*Qubit – Life Technologies*), para a determinação da concentração. Para avaliar eventual degradação, as amostras foram submetidas a eletroforese em gel de agarose. Os protocolos de extração e quantificação estão presentes no **Anexo C.**

5.1 Sequenciamento e Genotipagem

Boa parte dos marcadores a serem analisados no presente estudo, incluindo todos os SNPs dos sistemas *KiddLab 55 AISNPs*, SNP*for*ID 34-Plex, foram analisados por sequenciamento de nova geração. As bibliotecas de fragmentos para sequenciamento foram geradas por meio de ensaio customizado implementado no kit *HaloPlex Target Enrichment System* (*Agilent Technologies, Inc.*). A validação da biblioteca de fragmentos de cada indivíduo foi feita através do *2100 Bioanalyzer* (*Agilent Technologies, Inc.*). Quantidades equimolares de amostras de até 96 indivíduos foram combinadas em *pool* único, a fim de proceder para a etapa de sequenciamento na plataforma *MiSeq Personal Sequencer* (*Illumina Inc.*) utilizando o *MiSeq Reagent Kit V3* (*600 cycles*), segundo recomendações do fabricante.

Para a etapa do *Genotype Calling* foram utilizados os seguintes programas: Cutadapt (https://cutadapt.readthedocs.io/en/stable/installation.html), para remoção de adaptadores e

filtragem das *reads* por qualidade; Burrows-Wheeler Aligner (BWA-MEM) (H. Li & Durbin, 2009), para alinhamento das *reads* ao genoma de referência versão Hg19; GATK (https://software.broadinstitute.org/gatk/download/) *HaplotypeCaller*, para realizar a etapa de genotipagem propriamente dita; e VCFx (http://www.castelli-lab.net/apps/apps_vcfx.php), para remoção de genótipos de baixa qualidade. O protocolo do sequenciamento encontra-se no item 6 do **Anexo C,** já os procedimentos computacionais encontram-se no **Anexo D.**

Essa amostragem de DNA também foi submetida a ensaio de genotipagem 1.748.250 SNPs utilizando o *Infinium Multi-Ethnic Global-8 array* (MEGA) *Kit* (*Illumina*) e a plataforma *HiScanSQ System* (*Illumina Inc.*), realizado no Centro de Genômica Funcional da Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ-USP), em parceria com o Prof. Dr. Esteban J. Parra, da Universidade de Toronto, e Profa. Dra. Heather L. Norton, da Universidade de Cincinnati. Os dados de genotipagem podem ser acessados por meio do *software GenomeStudio* (*Illumina*), o qual também interroga os genótipos com baixa confiabilidade geradas pelo ensaio. O protocolo está disponível em https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/infinium_assays/infinium_lcg_assay/infinium-lcg-assay-reference-guide-15023139-04.pdf.

Resumidamente, o processo consiste em seis etapas:

a) Amplificação do DNA genômico;
b) Fragmentação enzimática do DNA;
c) Precipitação e ressuspensão do DNA;
d) Distribuição e hibridização do DNA com os *primers* contidos na *BeadChip*;
e) Extensão do *primer* em uma única base marcada e identificação do sinal emitido;
f) Leitura da *BeadChip*, a fim de coletar dados de intensidade da coloração.

Os processos desse tópico foram realizados antes do desenvolvimento deste trabalho.

5.2 Imputação

Para os 107 marcadores que não estavam presentes nos nossos ensaios laboratoriais, utilizou-se ferramentas de imputação (*Sanger Imputation Server* (SIS), IMPUTE2 e *Michigan Imputation Server* (MIS)), a fim de determinar os genótipos com base em padrões globais de

desequilíbrio de ligação (Das et al., 2016; B. Howie et al., 2009; 2011). Os protocolos de imputação estão presentes no **Anexo E.**

Os três outputs foram comparados para a obtenção dos resultados mais acurados possíveis para utilização nas análises seguintes, sendo assim, utilizamos a "Regra dos três genótipos". Essa regra foi montada para utilização apenas nos marcadores dos painéis estudados nesse trabalho e que não estavam presente no sequenciamento de NGS ou na genotipagem a partir do MEGA.

A fórmula foi desenvolvida em Excel e segue as seguintes condições:

- 1ª – Os marcadores não precisam estar em fase. Como trabalhamos com cálculos de frequência alélica, *locus* a *locus*, sem considerarmos haplótipos, resultados (0|1), (1|0) foram considerados equivalentes.
- 2ª – Se pelo menos duas das três imputações determinassem alelos equivalentes para um *loci*, este seria tomado como correto.
- 3ª – Se nenhuma das metodologias de imputação tivessem atribuído alelos equivalentes a uma posição, esta seria atribuída como um duplo *missing*.

5.3 Inferência de Ancestralidade

A inferência de ancestralidade da amostra populacional foi realizada a partir dos 5 painéis de ancestralidade genômica já desenvolvidos: *Kidd Lab set of 55 AISNPs* (Kidd et al., 2014), SNP*for*ID 34-*plex*, (Phillips et al., 2007), Seldin's 128 AISNPs (Kosoy et al., 2009), PIMA (Gontijo et al., 2020) e 192 AISNPs (Santos et al., 2016). Visto que alguns desses marcadores estão presentes em mais de um painel, o total observado de marcadores únicos é de 399 marcadores. O **Apêndice A** apresenta cada um desses marcadores e seus respectivos painéis.

Os softwares utilizados são ADMIX 95, desenvolvido por Bernardo Bertoni e disponível em www.genetica.fmed.edu.uy (Chakraborty, 1986), para ancestralidade populacional e ADMIXTURE 2.3.4 (Pritchard et al., 2000), para inferência de ancestralidade populacional e individual.

Os dados genotípicos de populações parentais foram obtidos no âmbito do *The 1000 Genomes Project* (1KGEN) (Auton et al., 2015) e do *The Human Genome Diversity Project* (HGDP-CEPH) (Bergström et al., 2020). Não foram utilizados, no entanto, a totalidade dos

indivíduos desses bancos de dados, pois o intuito do projeto é comparar a população local com as parentais homogêneas disponíveis. Assim, a partir do *1000 Genomes Project*, foram utilizados dados de 404 europeus (CEU, TSI, GBR e IBS), 504 africanos (YRI, LWK, GWD, MSL, ESN) e 504 indivíduos do leste asiático (CHB, JPT, CHS, CDX e KHV), totalizando 1412 indivíduos, cujos genótipos foram obtidos em ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/. A partir do HGDP, foram utilizados 484 indivíduos de populações parentais europeias, africanas, americanas e leste-asiáticas, dentre os 929 genomas disponíveis em ftp://ngs.sanger.ac.uk/production/hgdp. Três diferentes modelos de análise foram investigados, o modelo tetra-híbrido, onde as populações europeias (EUR), africanas (AFR), nativo-americanas (NAM) e asiáticas (EAS) foram utilizadas em conjunto e dois modelos tri-híbridos, variando entre as populações de referência NAM e EAS. As relações de indivíduos e populações estão dispostas no **Apêndice B**.

Para motivos de conferência, foram utilizados os resultados de inferência de ancestralidade obtidos a partir de 336.112 marcadores presentes no MEGA, pela colaboradora do Laboratório de Pesquisas Genômicas e Forenses (LPFG), Dra. Letícia Marcorin (Marcorin, 2021). A ancestralidade das populações miscigenadas presentes no 1KGEN, assumindo modelos trí-híbridos, foi inferida no trabalho de Martin et al. (2017) a partir do genoma completo dessas amostras. As análises de ambos foram realizadas através do software ADMIXTURE (Alexander & Lange, 2011) e está disposta na **Tabela 1**.

Tabela 1. Ancestralidade biogeográfica inferidas para a população de Ribeirão Preto (Marcorin, 2021) e para as populações miscigenadas presentes no 1KGEN (Martin et al., 2017).

Pop	AFR	AMR	EAS	EUR
RPT	0,183	0,067	0,026	0,724
ACB	0,880	0,003	-	0,117
ASW	0,756	0,031	-	0,213
CLM	0,078	0,257	-	0,666
MXL	0,043	0,470	-	0,487
PEL	0,025	0,773	-	0,202
PUR	0,139	0,129	-	0,732

5.3.1 ADMIX 95

O software utiliza da similaridade genética a partir do coeficiente de identidade genética para sua determinação de ancestralidade populacional. Para definir a probabilidade de que dois alelos escolhidos aleatoriamente são idênticos por natureza assume-se que: se a população híbrida "h" recebe m_1 fração dos seus alelos da população 1 e (1-m_1) da população 2, a identidade genética entre a população 1 e a "h", J_{1h}, é composta por duas partes, a probabilidade de o alelo ter vindo da população 1 (J_{11}) e a probabilidade de o alelo ter vindo da população 2 (J_{12}):

$$J_{1h} = m_1 J_{11} + (1 - m_1) J_{12}$$

Então Chakraborty, (2016), usou esta metodologia para estimar J_{ij} de maneira que:

$$\hat{J}_{ij} = \left[\sum_{l=1}^{L} \sum_{k=1}^{r_l} x_{ikl}\, x_{jkl} \right] \Big/ L$$

Onde x_{ikl} é a frequência do alelo "k" no locus "l" na população "i", "r_l" é o número de alelos no locus "l" e "L" o total de loci estudados.

Logo depois, estendeu-se a função para a obtenção da metodologia dos menores quadrados:

$$E\left[J_{in} - J_{ip}\right] = \sum_{j=1}^{p-1} m_j (J_{ij} - J_{ip})$$

Para i = 1, 2, ..., p.

Todas as relações entre populações onde m_j são negativas, são excluídas. Seu *input* é baseado nas frequências alélicas na população híbrida e nas parentais.

5.3.2 STRUCTURE

O *STRUCTURE* funciona utilizando de metodologias bayesianas e de simulações utilizando *machine learning.* Os dados genômicos de um indivíduo são comparados então a populações fornecidas ao programa como populações de referência. Essas populações possuem diferentes frequências, levando então ao software diferentes informações *a priori* da inferência sendo atribuídas a:

$$\Pr(Z, P, Q \mid X)$$

Onde “X” é o genótipo dos indivíduos amostrados, “Z” a população de onde pertence o indivíduo, que será determinado, “P” as frequências alélicas das populações estudadas no ensaio e “Q” a proporção de miscigenação para cada indivíduo (Pritchard et al., 2000).

Para que a etapa computacional ocorra de maneira a se obter os melhores resultados *a posteriori,* os valores de “*burn in”* e da “*Markov Chain Monte Carlo”,* devem ser selecionados de acordo com a quantidade de marcadores e de indivíduos amostrados. O *burn in* é a etapa onde ocorre a obtenção de dados e simulações iniciais para que se minimize o efeito das configurações iniciais e os possíveis erros, já o MCMC funciona de maneira a obtermos os resultados de inferência de ancestralidade, para ambos os comandos, o valor atribuído padrão varia de 10.000 a 100.000, mas cada amostragem deve ser analisada de maneira individualizada (Porras-Hurtado et al., 2013).

Além de incluir genótipos, o *input* inclui três informações importantes para cada indivíduo: *“POPDATA”,* que indica o grupo ao que o indivíduo pertence, “*POPFLAG”* refere-se a um valor binário (0/1) onde “0” indica que o programa não deve utilizar os grupos definidos em “*POPDATA*” para fins de clusterização e “1” indica que a informação de “*POPDATA*” deve ser utilizada quando a opção USEPOPINFO estiver habilitada, e “*LOCDATA”* indica a localidade de onde o indivíduo foi coletado.

Após carregar o *input* no software é necessário criar os parâmetros das corridas. Para nossos estudos, escolhemos os valores de *burn in* e MCMC de 200.000 para painéis de até 50 marcadores, 150.000 para os painéis de 50 a 59 marcadores e a partir de 125 marcadores, os valores escolhidos foram de 100.000, todos com 10 iterações para cada valor de *K*, o qual variou de 2 a 10. As opções escolhidas foram: *Use Admixture Model, Use sampling location as prior (LOCPRIOR), Alelle Frequencies Correlated, Update alelle frequencies using only individuals with POPFLAG=1 data* e *Compute data probability (for estimating K).*

Após finalizada as análises, os resultados foram tratados com o Clumpak, software online, disponível em http://clumpak.tau.ac.il/ (Kopelman et al., 2015). O aplicativo de browser junta três grandes softwares de análise de resultados do STRUCTURE, o Harvester (Earl & vonHoldt, 2012), utilizado para determinar os melhores valores de *K* para sua análise, o Clumpp (Jakobsson & Rosenberg, 2007), que analisa a similaridade entre as réplicas de

corridas para cada valor de *K* e o Distruct (Rosenberg, 2003), que produz um gráfico de barras para uma réplica única.

O input para o Clumpak consiste nos três resultados de corridas que apresentam os maiores valores de lnP(D) (Logaritmo natural estimado para a probabilidade dos dados) para cada valor de *K*, totalizando 27 corridas para os nove valores de *K* testados (*K* = 2~10).

5.3.3 Desenvolvimento de Painéis

O desenvolvimento dos painéis foi realizado adotando duas metodologias: i. Seleção dos marcadores que melhor individualizem as populações de referência a partir da diferença de frequências alélicas entre si, sendo essa diferença maior do que 30%; ii. Seleção dos marcadores que apresentarem maior relevância forense em questão de probabilidade de *match* e poder de exclusão. A quantidade de marcadores a serem selecionados foi estabelecida de maneira a se manter a possibilidade de análise via uma única reação multiplex para análise de mini-sequenciamento com o menor número de marcadores, mantendo a informatividade para obtenção da ancestralidade e para a identificação humana (Phillips et al., 2007; Sanchez et al., 2006).

5.4 Parâmetros Forenses

Os parâmetros forenses foram obtidos a partir do programa em *browser* chamado STRAF 2.0.8, disponível em http://straf.fr, baseado em R (Gouy & Zieger, 2017).

Os valores obtidos estão definidos a seguir.

Diversidade genética (GD), ou heterozigose esperada (H_{exp}):

$$H_{exp} = GD = \frac{n}{n-1}\left(1 - \sum_{i=1}^{n}(p_i)^2\right)$$

Onde *n* é o número de cromossomos amostradas e p_i a frequência do alelo "*i*" na população.

Probabilidade de *match* (PM), definido como a probabilidade de que dois indivíduos não relacionados apresentem o mesmo genótipo:

$$PM = \sum_{i} (G_i)^2$$

Onde G_i é a frequência observada do genótipo "*i*" em um dado *locus* da população.

O poder de discriminação (PD), definido como a probabilidade de diferenciar dois indivíduos não relacionados:

$$PD = 1 - PM$$

Conteúdos de informações de polimorfismo (PIC), nos dá a probabilidade de conseguirmos deduzir se um alelo presente em um marcador genético de uma dada criança foi transmitido pelo pai ou pela mãe:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

Onde p_i e p_j são as frequências alélicas.

O poder de exclusão (PE), define a probabilidade de excluir-se um indivíduo em um estudo de determinação de paternidade:

$$PE = h^2(1 - 2hH^2)$$

Sendo h a proporção de indivíduos heterozigotos e H a proporção de homozigotos esperados na amostra.

E o índice de paternidade típico (TPI), sendo o valor médio do índice de paternidade para um homem qualquer não excluído para um dado locus, dado por:

$$TPI = 1/(2H)$$

5.5 Estatísticas Genéticas e Populacionais

As frequências alélicas foram obtidas a partir do pacote desenvolvido para o *Microsoft Excel* chamado GenAlEx 6.503, disponível gratuitamente em https://biology-

assets.anu.edu.au/GenAlEx/Download.html (Peakall & Smouse, 2012). Para o cálculo do equilíbrio de Hardy-Weinberg, foi utilizado o software Plink 1.9, disponível em https://www.cog-genomics.org/plink/ (Chang et al., 2015). As análises de componente principal foram realizadas a partir da função *prcomp*, nativa da linguagem de programação R, dentro do ambiente RStudio (https://www.rstudio.com/products/rstudio/download/), utilizando R 4.1.1 (https://cran.r-project.org/bin/windows/base/). Os gráficos das PCAs foram montados a partir da função "scatter2D" presente no pacote para R chamado "Plot3D".

Resultados

6. Resultados

6.1 Imputação

A **Tabela 2** mostra a compatibilidade entre as posições que foram sequenciadas através de sequenciamento de nova geração (NGS) e das posições genotipadas a partir do *Multi-Ethnic Global-8 Array* (MEGA), para os marcadores utilizados neste trabalho, traz também os valores de compatibilidade entre as posições imputadas por cada servidor entre si e comparadas ao NGS e ao MEGA. Dos 399 marcadores selecionados para o projeto, 290 estavam presentes no MEGA e no NGS, sendo 26 AIMs sequenciados por NGS, 199 AIMs foram genotipados pelo MEGA e 65 foram sequenciados e genotipados. 106 AIMs foram imputados com sucesso enquanto que três marcadores, pertencentes ao painel desenvolvido por Seldin (rs1340093, rs10954737 e rs10236187), não foram imputados por qualquer uma das três metodologias. O marcador rs17287498 possui uma segunda nomenclatura, rs201381710, atribuído ao painel de Santos, devido a essa confusão, este marcador não foi utilizado no painel PIMA. A regra dos três genótipos foi aplicada nos 106 marcadores imputados, que também foram comparados para compatibilidade entre as três metodologias de imputação.

Tabela 2. Compatibilidade entre as posições genotipadas a partir do MEGA e do NGS, e das posições imputadas por diferentes metodologias.

Metodologias Comparadas			Compatibilidade %	Número de Marcadores	Número de *loci* Comparados
MEGA	vs	NGS	99,025	65	29.705
SIS	vs	MIS	98,630	106	54.378
SIS	vs	Impute2	98,311	54	27.702
SIS	vs	NGS	97,314	24	8.880
MIS	vs	Impute2	98,383	54	27.254
MIS	vs	NGS	97,455	24	8.880
Impute2	vs	NGS	90,790	24	4.392
Regra dos Três	vs	NGS	99,734	24	8.832

Os 106 marcadores imputados pelo SIS e pelo MIS não apresentaram posições interrogadas ou atribuídas a *missing*. Porém o sistema de imputação Impute2 não atribuiu valores para 6,76% das posições.

O sistema Impute2 apresentou também menor compatibilidade com o NGS quando comparada aos outros painéis. O MIS foi o sistema que apresentou maior compatibilidade com o sequenciamento de nova geração. Após a junção das três metodologias de imputação, seguindo a "Regra dos Três Genótipos" a compatibilidade com o sequenciamento atingiu o valor de 99,734%.

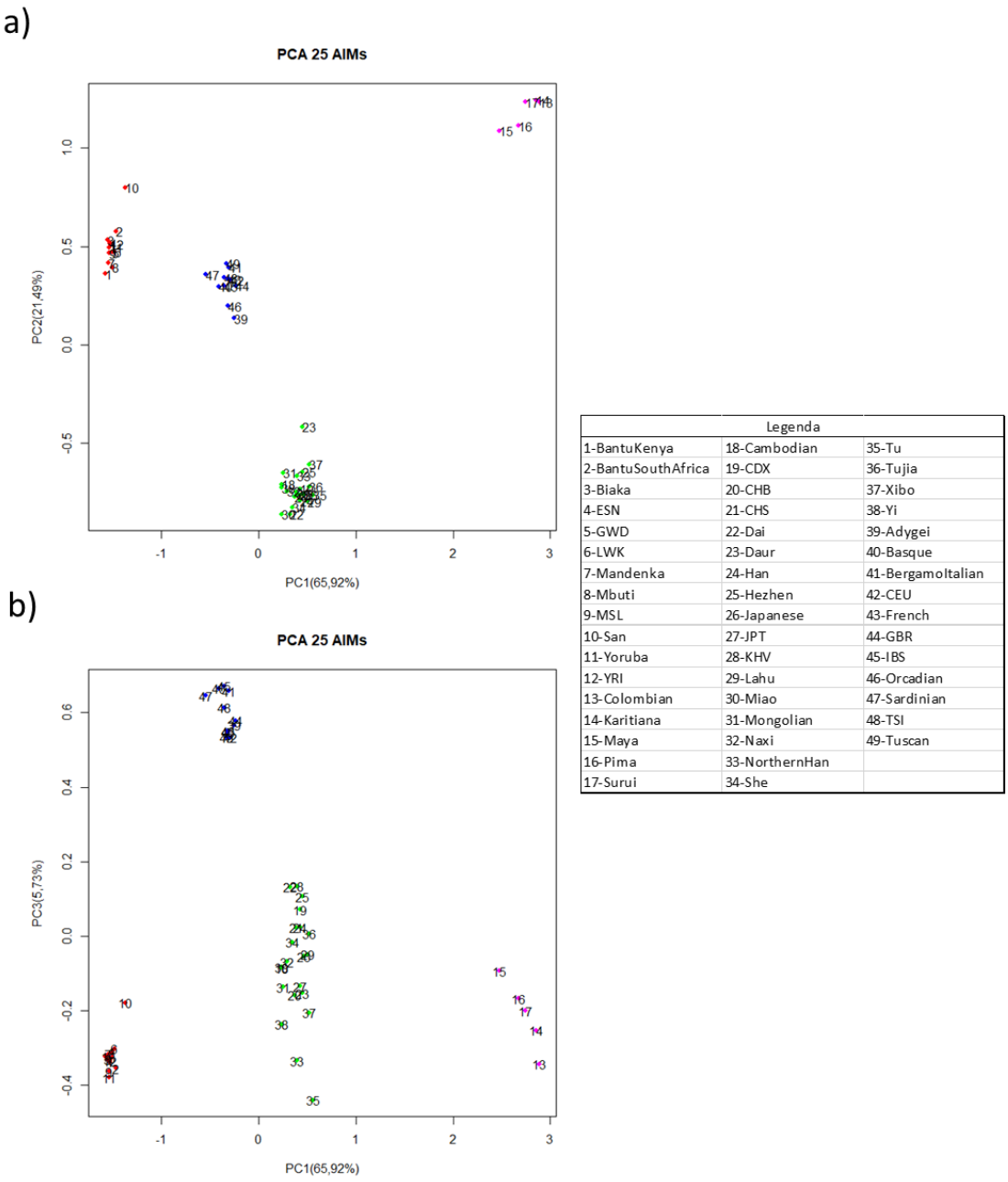
6.2 Estatísticas Populacionais

As frequências alélicas para cada um dos 396 marcadores, em cada população estudada está disposta no **Apêndice C.**

A **Tabela A1,** do **Apêndice A** traz os valores de heterozigose esperada e observada, bem como o valor de *P-value* para o Equilíbrio de Hardy-Weinberg para cada um dos 396 marcadores estudados neste projeto, considerando apenas os 343 indivíduos de Ribeirão Preto que foram coletados de maneira randômica. Apenas o SNP rs2572307 apresentou-se fora do equilíbrio após a aplicação da correção de Bonferroni para múltiplos testes ($\alpha_{BONFERRONI}$ = 0.05/396 = 0.000126). A **Tabela A2**, dispõe os resultados para todos os 513 indivíduos amostrados, onde 18 SNPs apresentaram-se fora do equilíbrio após a mesma correção para múltiplos testes.

6.3 PIMA (25 AIMs)

A análise de componentes principais (**Figura 2)** resultou em três principais componentes que somados explicam 92,64% da variância, distribuída da seguinte forma: 65,42% para a PC1, 21,49% para a PC2 e de 5,73% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das duas primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-Bergamoltalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 2. Análise de componentes principais baseada em 25 dos 26 marcadores que compõem o painel PIMA, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (65,92%) vs PC2 (21,49%); (b) PC1 vs PC3 (5,73%).

Os valores apresentados na **Tabela 3** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 3. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel PIMA.

População	AFR	AMR	EAS	EUR	R^2
ACB	0,9243 ± 0,0069	0,0284 ± 0,0021	-0,0511 ± 0,0042	0,0985 ± 0,0061	0,999807
ASW	0,7517 ± 0,0219	0,0690 ± 0,0069	-0,0741 ± 0,0125	0,2534 ± 0,0191	0,994268
CLM	0,0617 ± 0,0019	0,2278 ± 0,0007	0,0671 ± 0,0015	0,6435 ± 0,0018	0,999836
MXL	-0,0061 ± 0,0172	0,4560 ± 0,0067	0,0347 ± 0,0152	0,5155 ± 0,0167	0,999499
PEL	0,0437 ± 0,0029	0,7127 ± 0,0012	0,0268 ± 0,0029	0,2169 ± 0,0030	0,999997
PUR	0,1531 ± 0,0035	0,1413 ± 0,0017	-0,0263 ± 0,0030	0,7319 ± 0,0036	0,996892
RPT	0,1938 ± 0,0028	0,0658 ± 0,0016	0,0111 ± 0,0028	0,7293 ± 0,0022	0,996616
ACB	0,9235 ± 0,0164	0,0251 ± 0,0048	-	0,0514 ± 0,0119	0,999342
ASW	0,7530 ± 0,0302	0,0714 ± 0,0094	-	0,1756 ± 0,0214	0,993338
CLM	0,0621 ± 0,0086	0,2405 ± 0,0023	-	0,6975 ± 0,0078	0,998181
MXL	0,0054 ± 0,0234	0,4659 ± 0,0069	-	0,5287 ± 0,0174	0,999099
PEL	0,0385 ± 0,0082	0,7133 ± 0,0030	-	0,2482 ± 0,0054	0,999971
PUR	0,1554 ± 0,0039	0,1416 ± 0,0019	-	0,7030 ± 0,0030	0,995862
RPT	0,1926 ± 0,0001	0,0689 ± 0,0001	-	0,7385 ± 0,0000	0,999998
ACB	0,8465 ± 0,0087	-	-0,0033 ± 0,0100	0,1568 ± 0,0102	0,997984
ASW	0,5997 ± 0,0274	-	0,0641 ± 0,0268	0,3362 ± 0,0278	0,96054
CLM	-0,3496 ± 0,0865	-	0,3010 ± 0,0933	1,0485 ± 0,0639	0,821204
MXL	-0,8392 ± 0,1229	-	0,4387 ± 0,1853	1,4005 ± 0,1331	0,863213
PEL	-1,2108 ± 0,1547	-	0,5544 ± 0,2894	1,6565 ± 0,2138	0,877412
PUR	-0,0242 ± 0,0696	-	0,1840 ± 0,0619	0,8402 ± 0,0527	0,458755
RPT	0,1448 ± 0,0288	-	0,1063 ± 0,0302	0,7490 ± 0,0331	0,553335

A **Figura 3**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 4**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

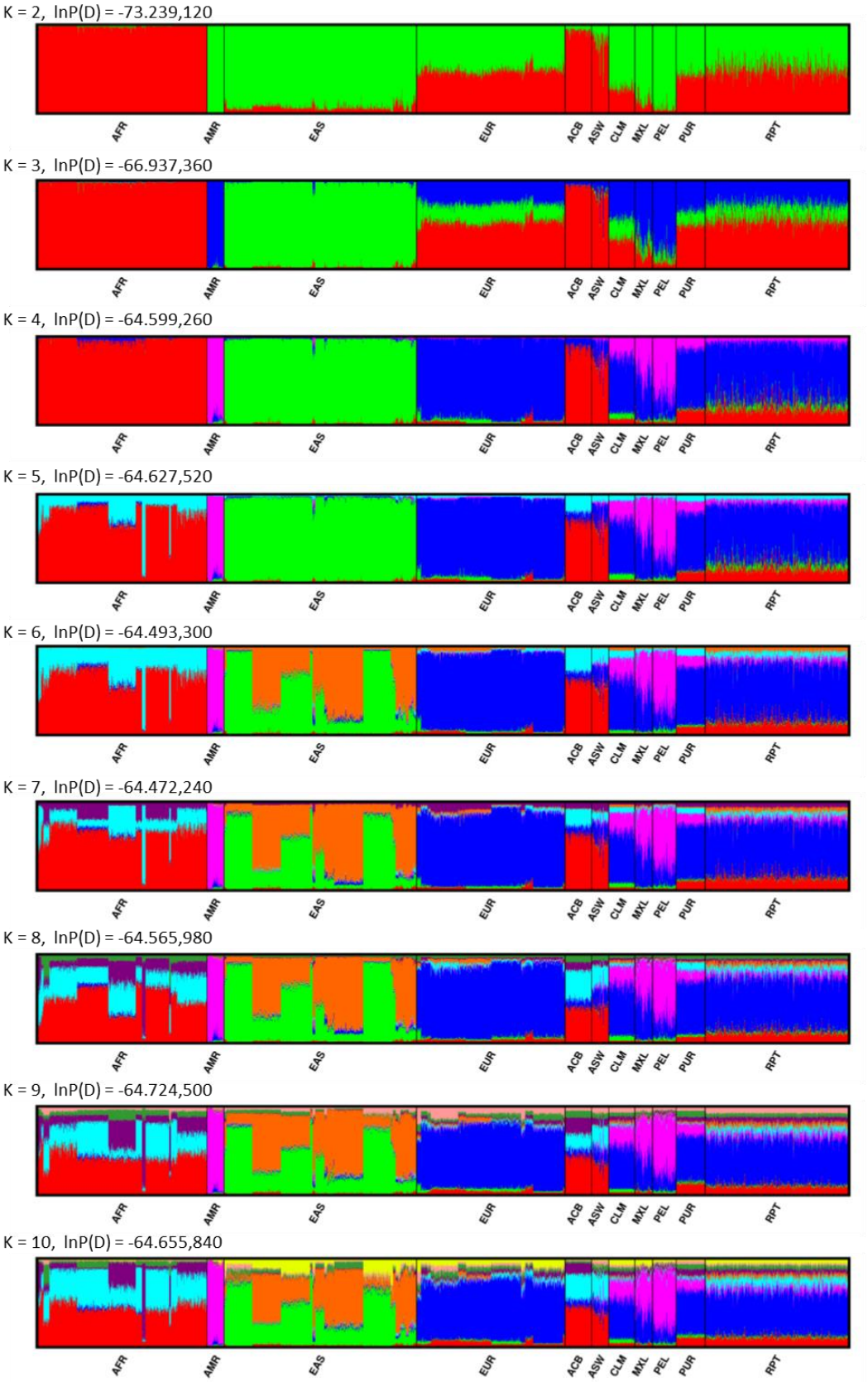


Figura 3. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel PIMA. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 4. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (K = 4), resultante da análise de 25 dos 26 marcadores que compõem o painel PIMA. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (n = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,9804**	0,0006	0,0024	0,0165
AMR	0,0039	**0,9359**	0,0177	0,0426
EAS	0,0106	0,0047	**0,9745**	0,0102
EUR	0,0243	0,0076	0,0305	**0,9377**
ACB	0,8863	0,0047	0,0043	0,1047
ASW	0,7291	0,0474	0,0099	0,2136
CLM	0,0578	0,2171	0,0676	0,6576
MXL	0,0303	0,4416	0,0216	0,5065
PEL	0,0268	0,6655	0,0365	0,2711
PUR	0,1619	0,1324	0,0184	0,6873
RPT	0,2015	0,0552	0,0691	0,6742

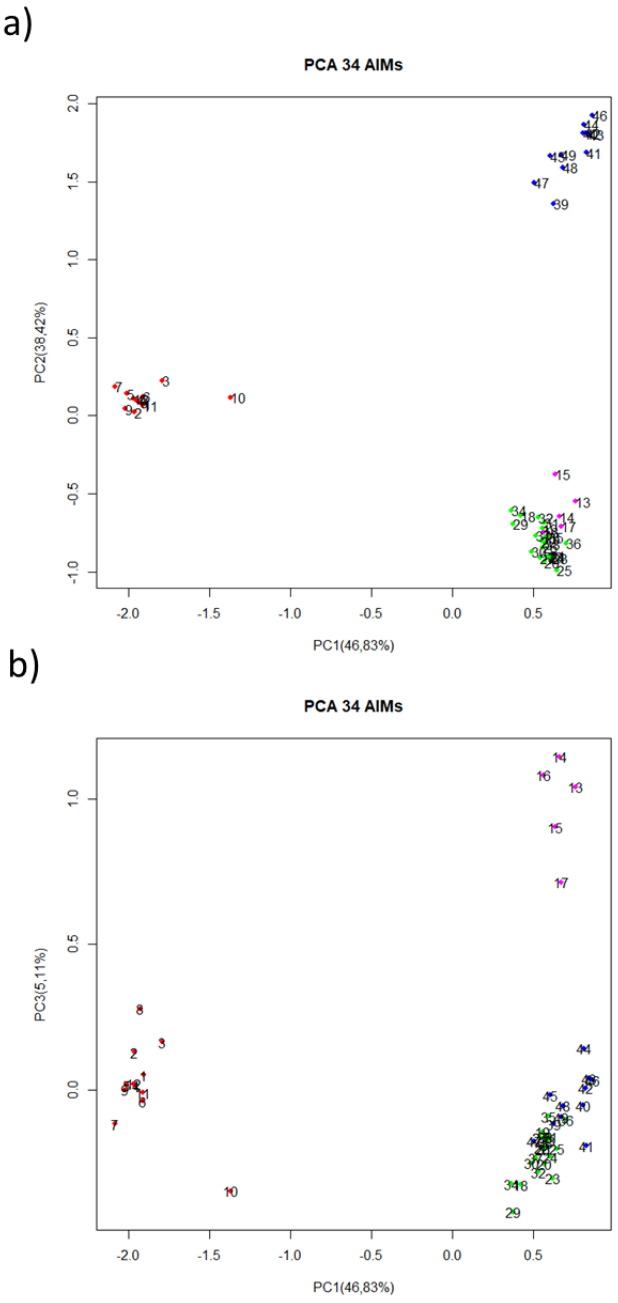
A **Tabela 5** mostra os valores de probabilidade de *match* combinada (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade típico combinado (IPC) obtidos para as populações miscigenadas analisadas junto ao projeto 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando o painel PIMA.

Tabela 5. Parâmetros forenses combinados para as populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto (*n* = 343) utilizando o painel PIMA (25 AIMs).

População	PMC	PDC	IPC	PEC
ACB	1,0848E-04	0,9998915206	6,2972E-06	0,640212848
ASW	1,3481E-06	0,9999986519	1,0416E-04	0,816136600
CLM	3,6846E-10	0,9999999996	7,3861E-02	0,981893262
MXL	2,9932E-10	0,9999999997	2,0654E-02	0,967587144
PEL	1,2918E-08	0,9999999871	1,2913E-03	0,908275618
PUR	2,6898E-09	0,9999999973	3,0171E-02	0,976760729
RPT	6,2519E-09	0,9999999937	3,9453E-03	0,946751696

6.4 SNP*for*ID 34-plex (34 AIMs)

A análise de componentes principais (**Figura 4)** resultou em três principais componentes que somados explicam 90,36% da variância, distribuída da seguinte forma: 46,83% para a PC1, 38,42% para a PC2 e de 5,11% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das três primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-Bergamoltalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 4 Análise de componentes principais baseada nos 34 marcadores que compõem o painel SNP*for*ID 34-plex, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (46,83%) vs PC2 (38,42%); (b) PC1 vs PC3 (5,11%).

Os valores apresentados na **Tabela 6** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 6. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel SNP*for*ID 34-plex.

População	AFR	AMR	EAS	EUR	R^2
ACB	0,8905 ± 0,0029	0,0152 ± 0,0057	-0,0034 ± 0,0062	0,0978 ± 0,0000	0,999900
ASW	0,7503 ± 0,0017	0,0494 ± 0,0031	0,0003 ± 0,0033	0,2000 ± 0,0005	0,999958
CLM	0,0955 ± 0,0011	0,2824 ± 0,0032	-0,0446 ± 0,0026	0,6667 ± 0,0020	0,999076
MXL	0,0209 ± 0,0064	0,4859 ± 0,0210	0,0675 ± 0,0194	0,4257 ± 0,0092	0,992293
PEL	0,0186 ± 0,0004	0,7792 ± 0,0011	0,0026 ± 0,0012	0,1996 ± 0,0004	0,999992
PUR	0,1421 ± 0,0032	0,0966 ± 0,0117	0,0749 ± 0,0091	0,6864 ± 0,0087	0,992206
RPT	0,1890 ± 0,0020	0,0715 ± 0,0090	0,0344 ± 0,0064	0,7052 ± 0,0067	0,997773
ACB	0,8892 ± 0,0024	0,0164 ± 0,0031	-	0,0943 ± 0,0024	0,999865
ASW	0,7494 ± 0,0015	0,0523 ± 0,0019	-	0,1982 ± 0,0013	0,999941
CLM	0,0933 ± 0,0014	0,2410 ± 0,0020	-	0,6657 ± 0,0020	0,998803
MXL	0,0254 ± 0,0072	0,5242 ± 0,0102	-	0,4504 ± 0,0081	0,993133
PEL	0,0187 ± 0,0004	0,7819 ± 0,0005	-	0,1995 ± 0,0004	0,999991
PUR	0,1601 ± 0,0046	0,1422 ± 0,0083	-	0,6977 ± 0,0080	0,985523
RPT	0,1965 ± 0,0029	0,0902 ± 0,0073	-	0,7133 ± 0,0065	0,995381
ACB	0,8880 ± 0,0032	-	0,0142 ± 0,0039	0,0977 ± 0,0025	0,999828
ASW	0,7477 ± 0,0027	-	0,0449 ± 0,0031	0,2074 ± 0,0019	0,999862
CLM	0,1000 ± 0,0022	-	0,2099 ± 0,0028	0,6900 ± 0,0028	0,997345
MXL	0,0254 ± 0,0115	-	0,4610 ± 0,0160	0,5136 ± 0,0124	0,979526
PEL	0,0415 ± 0,0083	-	0,6242 ± 0,0117	0,3344 ± 0,0092	0,993877
PUR	0,1585 ± 0,0032	-	0,1493 ± 0,0049	0,6922 ± 0,0047	0,993895
RPT	0,1997 ± 0,0023	-	0,0866 ± 0,0043	0,7138 ± 0,0038	0,997362

A **Figura 5**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 7**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

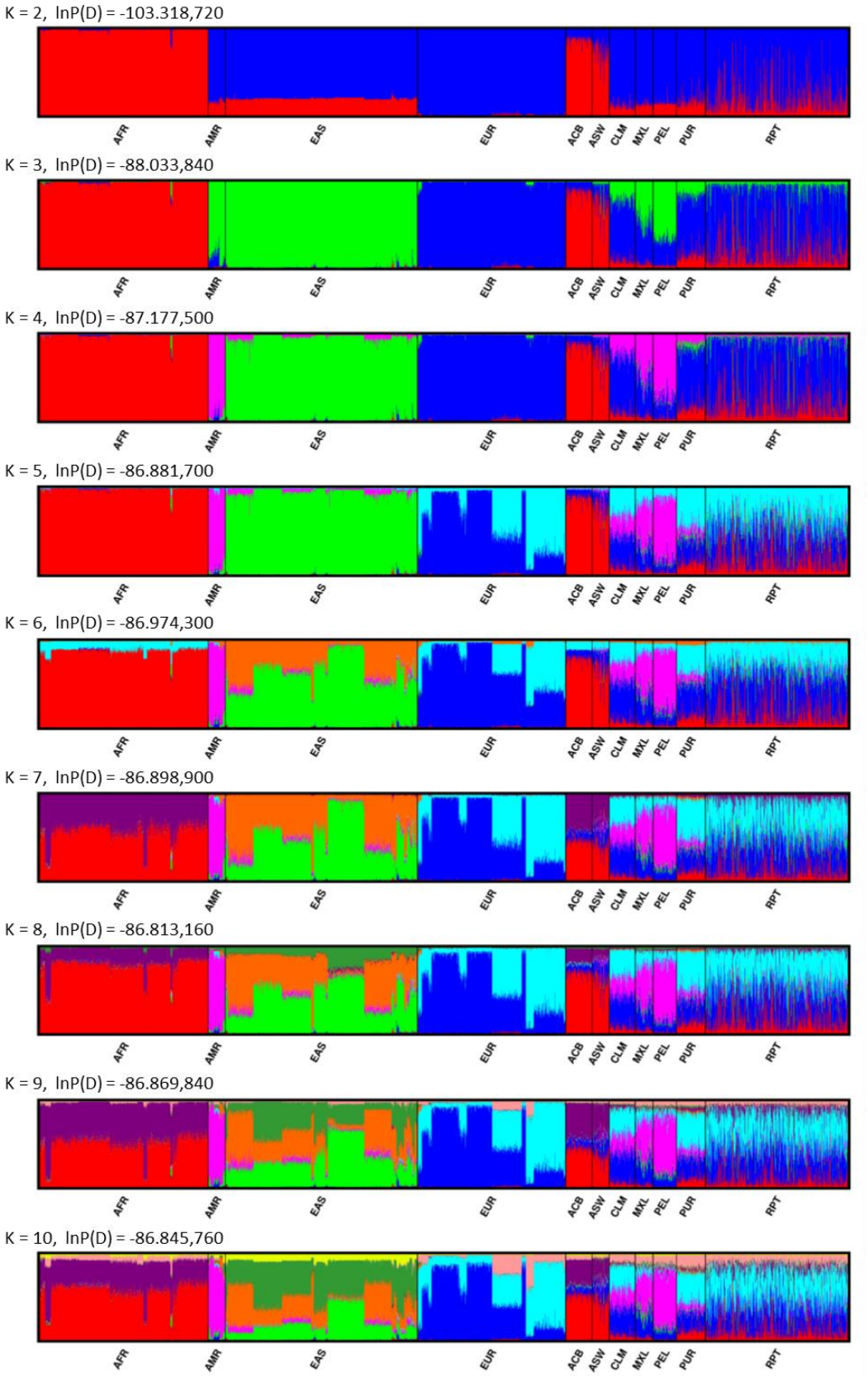


Figura 5 Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel SNP*for*ID 34-plex. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 7. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 34 marcadores que compõem o painel SNP*for*ID 34-plex. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,9895**	0,0020	0,0040	0,0046
AMR	0,0039	**0,9269**	0,0337	0,0355
EAS	0,0035	0,0308	**0,9609**	0,0047
EUR	0,0081	0,0015	0,0077	**0,9827**
ACB	0,8852	0,0044	0,0044	0,1060
ASW	0,7433	0,0416	0,0095	0,2056
CLM	0,0941	0,2372	0,0088	0,6598
MXL	0,0526	0,4715	0,0418	0,4342
PEL	0,0315	0,7255	0,0213	0,2218
PUR	0,1626	0,1014	0,0713	0,6646
RPT	0,2074	0,0279	0,0626	0,7020

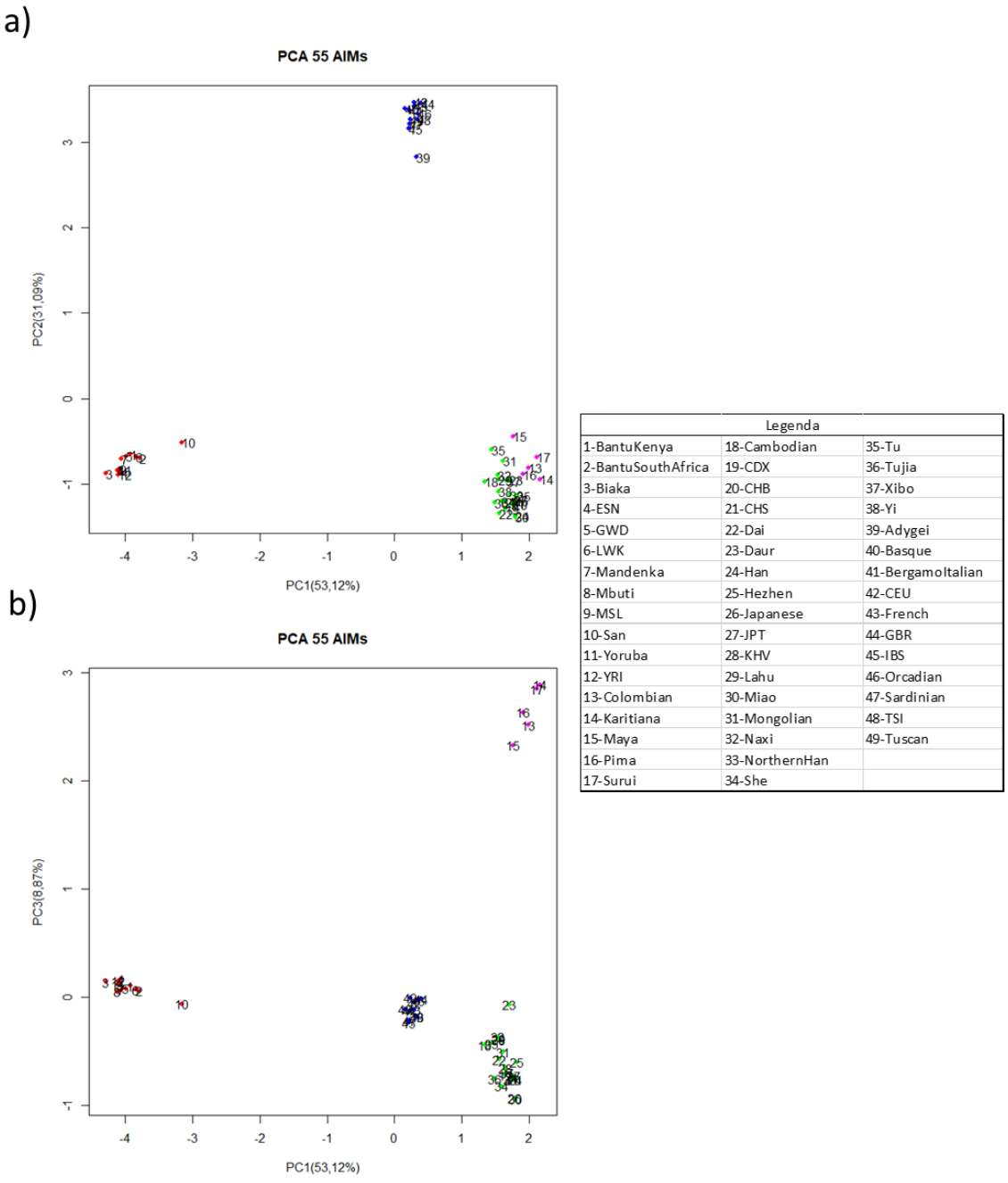
A **Tabela 8** mostra os valores de probabilidade de match combinado (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade combinado (IPC) obtidos para as populações miscigenadas presentes no âmbito do 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando o painel SNP*for*ID 34-plex.

Tabela 8. Parâmetros forenses combinados para as populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto (*n* = 343) utilizando o painel SNP*for*ID 34-plex.

População	PMC	PDC	IPC	PEC
ACB	8,3623E-11	0,999999999916377	0,000177	0,982224020
ASW	4,5626E-12	0,999999999995437	0,001885	0,993411990
CLM	3,9052E-11	0,999999999960948	0,000413	0,984489799
MXL	1,6598E-11	0,999999999983402	0,000512	0,987628585
PEL	6,1646E-10	0,999999999383540	0,000043	0,969141143
PUR	1,1444E-11	0,999999999988556	0,000113	0,966075334
RPT	1,1943E-11	0,999999999988057	0,000114	0,968195972

6.5 Kidd's Lab Set of 55 AIMs

A análise de componentes principais (**Figura 6)** resultou em três principais componentes que somados explicam 93,08% da variância, distribuída da seguinte forma: 53,12% para a PC1, 31,09% para a PC2 e de 8,87% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das três primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-BergamoItalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 6 Análise de componentes principais baseada nos 55 marcadores que compõem o painel *Kidd Lab set of 55 AISNPs*, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (53,12%) vs PC2 (31,09%); (b) PC1 vs PC3 (8,87%).

Os valores apresentados na **Tabela 9** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 9. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel *Kidd Lab set of 55 AISNPs*.

População	AFR	AMR	EAS	EUR	R^2
ACB	0,8858 ± 0,0005	0,0276 ± 0,0010	-0,0045 ± 0,0010	0,0911 ± 0,0000	0,999991
ASW	0,7780 ± 0,0014	0,0487 ± 0,0023	-0,0107 ± 0,0027	0,1839 ± 0,0000	0,999940
CLM	0,0728 ± 0,0020	0,2248 ± 0,0029	0,0258 ± 0,0037	0,6766 ± 0,0029	0,996693
MXL	0,0455 ± 0,0008	0,5114 ± 0,0018	0,0118 ± 0,0019	0,4313 ± 0,0005	0,999849
PEL	0,0264 ± 0,0003	0,7648 ± 0,0008	0,0249 ± 0,0009	0,1838 ± 0,0004	0,999992
PUR	0,1203 ± 0,0026	0,1356 ± 0,0043	0,0356 ± 0,0057	0,7085 ± 0,0047	0,992038
RPT	0,1915 ± 0,0006	0,0569 ± 0,0013	0,0404 ± 0,0016	0,7112 ± 0,0014	0,999661
ACB	0,8851 ± 0,0005	0,0240 ± 0,0007	-	0,0909 ± 0,0007	0,999985
ASW	0,7758 ± 0,0012	0,0427 ± 0,0017	-	0,1815 ± 0,0014	0,999922
CLM	0,0854 ± 0,0026	0,2386 ± 0,0019	-	0,6760 ± 0,0030	0,997065
MXL	0,0514 ± 0,0007	0,5190 ± 0,0010	-	0,4295 ± 0,0010	0,999855
PEL	0,0299 ± 0,0004	0,7807 ± 0,0005	-	0,1894 ± 0,0005	0,999983
PUR	0,1333 ± 0,0038	0,1538 ± 0,0039	-	0,7128 ± 0,0043	0,988247
RPT	0,1994 ± 0,0007	0,0833 ± 0,0011	-	0,7173 ± 0,0011	0,999528
ACB	0,8773 ± 0,0004	-	0,0213 ± 0,0006	0,1015 ± 0,0005	0,999992
ASW	0,7626 ± 0,0034	-	0,0356 ± 0,0043	0,2019 ± 0,0028	0,999733
CLM	0,0660 ± 0,0060	-	0,2492 ± 0,0075	0,6848 ± 0,0080	0,968160
MXL	0,0129 ± 0,0126	-	0,4467 ± 0,0159	0,5404 ± 0,0141	0,955621
PEL	0,0188 ± 0,0173	-	0,6384 ± 0,0257	0,3428 ± 0,0245	0,935216
PUR	0,1314 ± 0,0021	-	0,1687 ± 0,0028	0,6999 ± 0,0028	0,995505
RPT	0,1973 ± 0,0010	-	0,0931 ± 0,0018	0,7096 ± 0,0017	0,999000

A **Figura 7**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 10**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

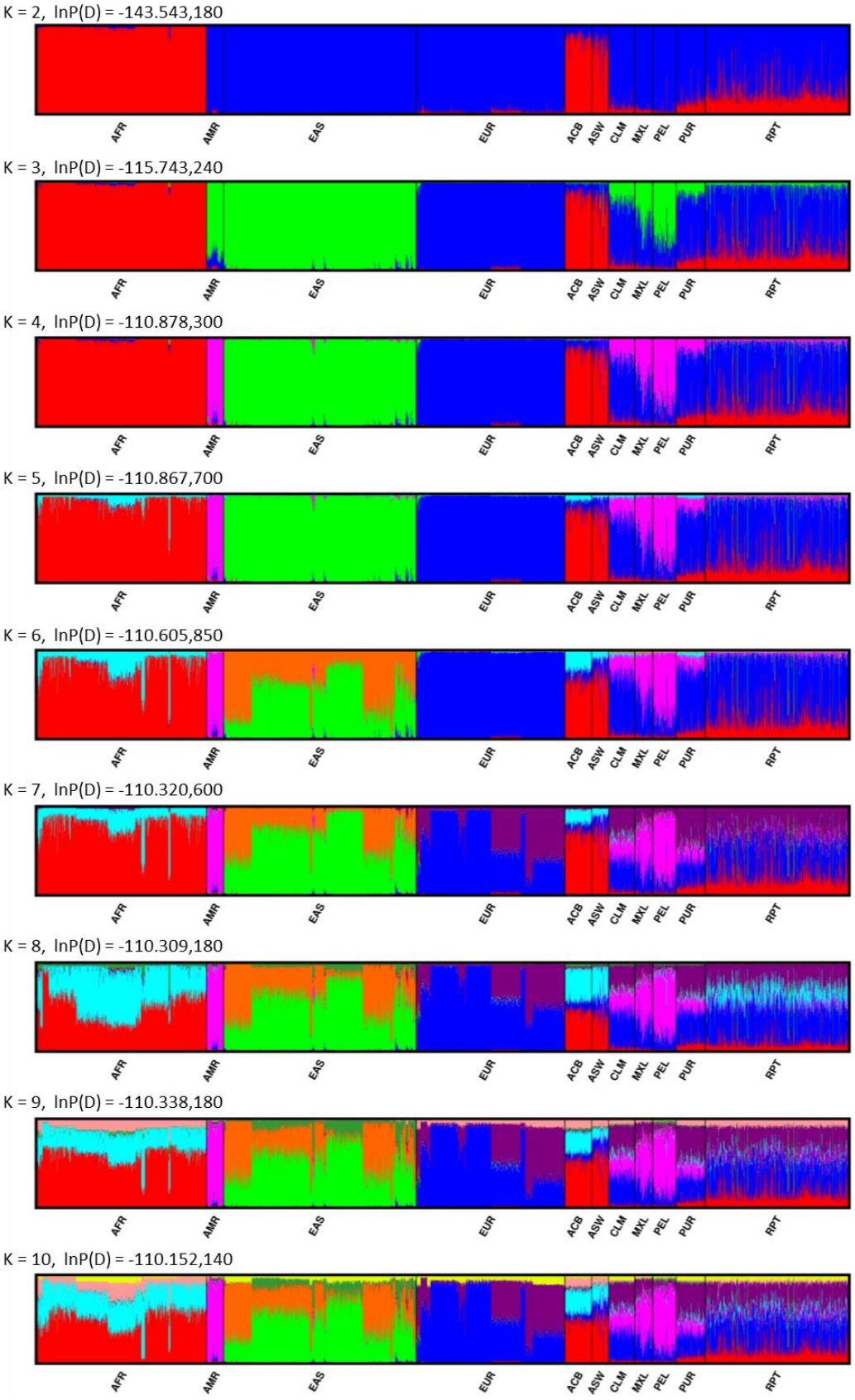


Figura 7 Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel *Kidd Lab set of 55 AISNPs*. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 10. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (K = 4), resultante da análise dos 55 marcadores que compõem o painel *Kidd Lab set of 55 AISNPs*. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (n = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,9897**	0,0007	0,0015	0,0081
AMR	0,0136	**0,9385**	0,0142	0,0337
EAS	0,0023	0,0071	**0,9830**	0,0075
EUR	0,0065	0,0015	0,0033	**0,9888**
ACB	0,8758	0,0040	0,0090	0,1112
ASW	0,7779	0,0325	0,0012	0,1884
CLM	0,0815	0,2346	0,0103	0,6736
MXL	0,0423	0,4809	0,0252	0,4516
PEL	0,0381	0,7324	0,0147	0,2148
PUR	0,1403	0,1577	0,0074	0,6946
RPT	0,2479	0,0431	0,0341	0,6749

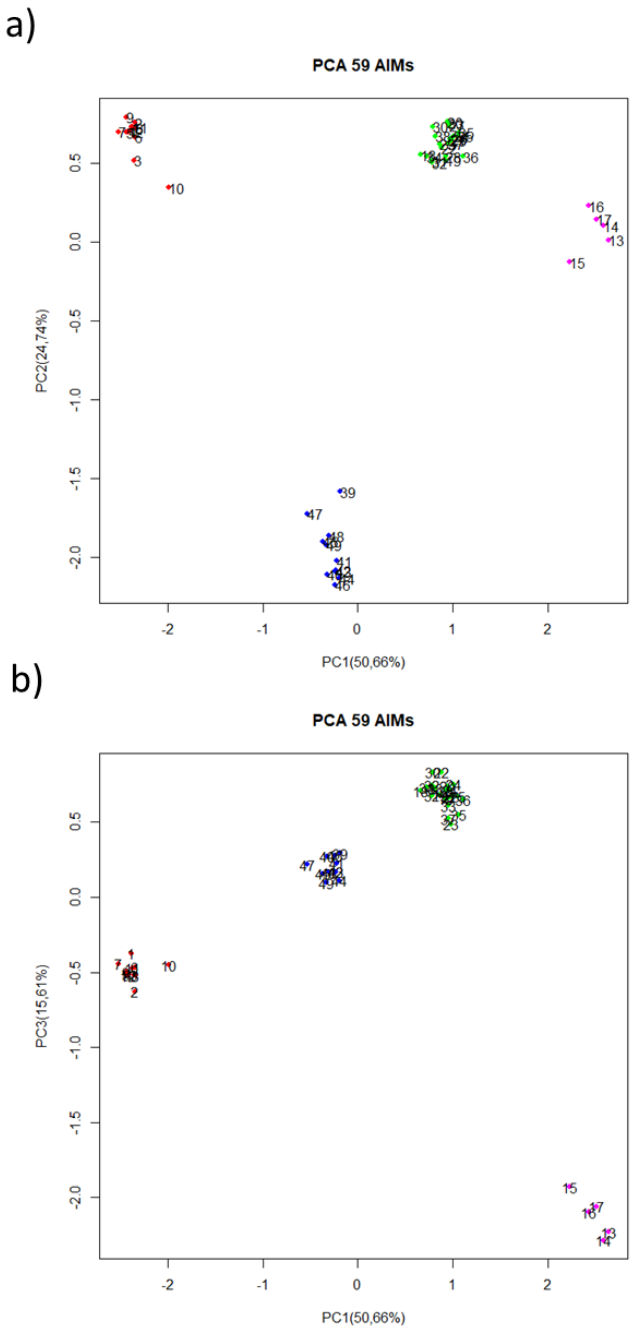
A **Tabela 11** mostra os valores de probabilidade de match combinado (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade combinado (IPC) obtidos para as populações miscigenadas presentes no âmbito do 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando o painel *Kidd Lab set of 55 AISNPs*.

Tabela 11. Parâmetros forenses combinados para as populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto (*n* = 343) utilizando o painel *Kidd Lab set of 55 AISNPs*.

População	PMC	PDC	IPC	PEC
ACB	1,1418E-10	0,999999999885823	1,6451E-11	0,926709141
ASW	2,1856E-13	0,999999999999781	8,3827E-10	0,975721415
CLM	3,0350E-16	>0,999999999999999	9,3436E-08	0,994038142
MXL	5,0880E-17	>0,999999999999999	1,5706E-07	0,995556653
PEL	9,3370E-15	0,999999999999991	3,9021E-09	0,984434778
PUR	4,2831E-16	>0,999999999999999	2,5222E-08	0,989705902
RPT	4,2874E-16	>0,999999999999999	1,6591E-08	0,988328675

6.6 PIMA + 34-plex (59 AIMs)

A análise de componentes principais (**Figura 8)** resultou em três principais componentes que somados explicam 91,01% da variância, distribuída da seguinte forma: 50,66% para a PC1, 24,74% para a PC2 e de 15,61% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das duas primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-BergamoItalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 8 Análise de componentes principais baseada nos 59 marcadores que compõem o painel PIMA + 34-plex, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (53,12%) vs PC2 (31,09%); (b) PC1 vs PC3 (8,87%).

Os valores apresentados na **Tabela 12** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 12. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel PIMA + 34-plex.

População	AFR	AMR	EAS	EUR	R^2
ACB	0,9004 ± 0,0011	0,0200 ± 0,0009	-0,0176 ± 0,0017	0,0972 ± 0,0000	0,999966
ASW	0,7532 ± 0,0002	0,0645 ± 0,0002	-0,0126 ± 0,0003	0,1950 ± 0,0000	0,999999
CLM	0,0821 ± 0,0003	0,2396 ± 0,0002	0,0165 ± 0,0004	0,6618 ± 0,0004	0,999965
MXL	0,0308 ± 0,0022	0,4716 ± 0,0027	0,0545 ± 0,0040	0,4431 ± 0,0000	0,999324
PEL	0,0318 ± 0,0009	0,7221 ± 0,0013	0,0397 ± 0,0021	0,2064 ± 0,0000	0,999957
PUR	0,1474 ± 0,0025	0,1338 ± 0,0025	0,0157 ± 0,0046	0,7031 ± 0,0044	0,990802
RPT	0,1902 ± 0,0016	0,0671 ± 0,0020	0,0244 ± 0,0038	0,7184 ± 0,0035	0,996302
ACB	0,8957 ± 0,0008	0,0129 ± 0,0008	-	0,0915 ± 0,0011	0,999530
ASW	0,7493 ± 0,0001	0,0589 ± 0,0001	-	0,1918 ± 0,0001	0,999999
CLM	0,0876 ± 0,0004	0,2466 ± 0,0002	-	0,6658 ± 0,0004	0,999962
MXL	0,0421 ± 0,0019	0,4917 ± 0,0016	-	0,4663 ± 0,0023	0,999400
PEL	0,0437 ± 0,0009	0,7402 ± 0,0007	-	0,2161 ± 0,0011	0,999947
PUR	0,1563 ± 0,0032	0,1401 ± 0,0026	-	0,7036 ± 0,0039	0,988050
RPT	0,1989 ± 0,0020	0,0774 ± 0,0023	-	0,7237 ± 0,0028	0,994230
ACB	0,8849 ± 0,0029	-	0,0076 ± 0,0036	0,1075 ± 0,0030	0,999615
ASW	0,7089 ± 0,0065	-	0,0561 ± 0,0077	0,2350 ± 0,0066	0,997102
CLM	0,0017 ± 0,0165	-	0,3009 ± 0,0228	0,6974 ± 0,0279	0,752818
MXL	-0,1294 ± 0,0279	-	0,5183 ± 0,0417	0,6110 ± 0,0369	0,734187
PEL	-0,1421 ± 0,0377	-	0,6877 ± 0,0609	0,4544 ± 0,0531	0,699305
PUR	0,1338 ± 0,0097	-	0,1963 ± 0,0128	0,6699 ± 0,0153	0,873801
RPT	0,1977 ± 0,0042	-	0,1045 ± 0,0064	0,6978 ± 0,0076	0,977849

A **Figura 9**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 13**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

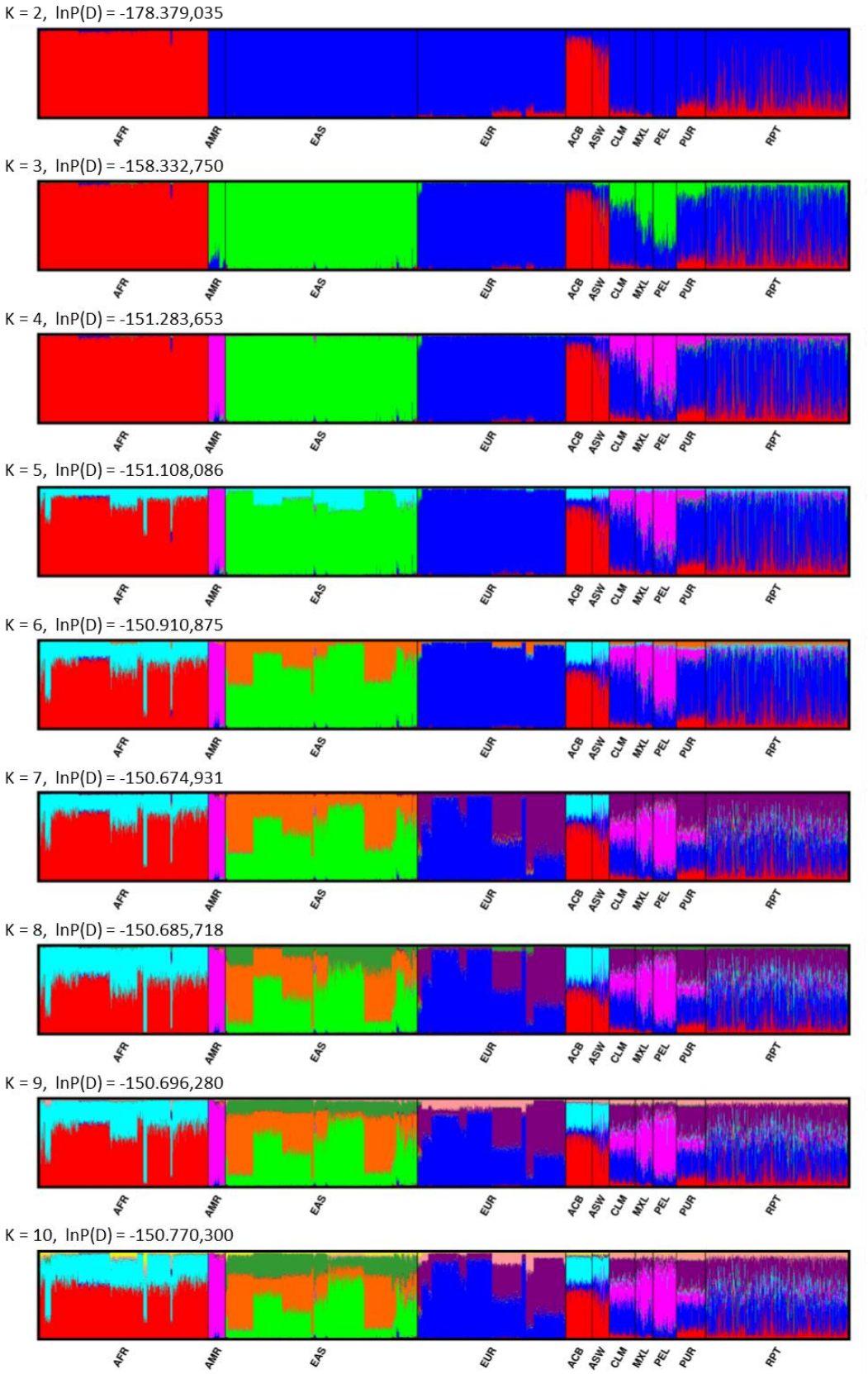


Figura 9. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel PIMA + 34-plex. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 13. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (K = 4), resultante da análise dos 59 marcadores que compõem o painel PIMA + 34-plex. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (n = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,9905**	0,0006	0,0021	0,0068
AMR	0,0026	**0,9445**	0,0158	0,0371
EAS	0,0023	0,0047	**0,9878**	0,0052
EUR	0,0090	0,0017	0,0059	**0,9833**
ACB	0,8887	0,0039	0,0027	0,1047
ASW	0,7395	0,0442	0,0058	0,2104
CLM	0,0879	0,2317	0,0170	0,6634
MXL	0,0523	0,4604	0,0312	0,4561
PEL	0,0328	0,6831	0,0452	0,2389
PUR	0,1637	0,1271	0,0353	0,6739
RPT	0,1983	0,0454	0,0547	0,7016

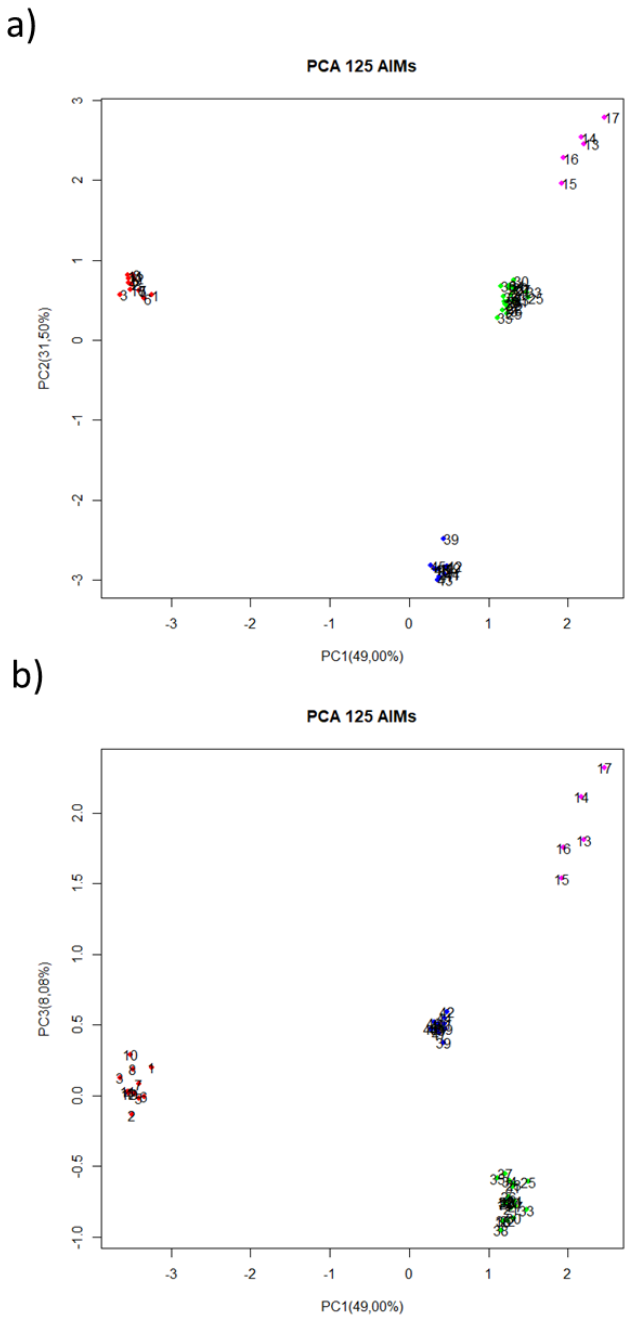
A **Tabela 14** mostra os valores de probabilidade de match combinado (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade combinado (IPC) obtidos para as populações miscigenadas presentes no âmbito do 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando o painel PIMA + 34-plex (59 AIMs).

Tabela 14. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o painel PIMA + 34-plex (59 AIMs).

População	PMC	PDC	IPC	PEC
ACB	9,0714E-15	0,999999999999991	1,11653E-09	0,993604431
ASW	6,1509E-18	>0,999999999999999	1,96344E-07	0,998788706
CLM	1,4389E-20	>0,999999999999999	3,04923E-05	0,999719161
MXL	4,9681E-21	>0,999999999999999	1,05791E-05	0,999599007
PEL	7,9635E-18	>0,999999999999999	5,52254E-08	0,997169490
PUR	3,0782E-20	>0,999999999999999	3,39590E-06	0,999211616
RPT	7,4668E-20	>0,999999999999999	4,50901E-07	0,998306489

6.7 Seldin's 128 AIMs (125 AIMs)

A análise de componentes principais (**Figura 10)** resultou em três principais componentes que somados explicam 88,58% da variância, distribuída da seguinte forma: 49,00% para a PC1, 31,50% para a PC2 e de 8,08% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das duas primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-Bergamoltalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 10. Análise de componentes principais baseada em 125 dos 128 marcadores que compõem o painel desenvolvido por Seldin, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (49,00%) vs PC2 (31,50%); (b) PC1 vs PC3 (8,08%).

Os valores apresentados na **Tabela 15** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 15. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel desenvolvido por Seldin.

População	AFR	AMR	EAS	EUR	R^2
ACB	0,8937 ± 0,0001	0,0025 ± 0,0001	0,0190 ± 0,0002	0,0848 ± 0,0000	0,999999
ASW	0,7734 ± 0,0001	0,0277 ± 0,0001	0,0194 ± 0,0003	0,1795 ± 0,0000	0,999999
CLM	0,0755 ± 0,0010	0,2441 ± 0,0014	0,0538 ± 0,0028	0,6266 ± 0,0016	0,999624
MXL	0,0307 ± 0,0013	0,4544 ± 0,0021	0,0981 ± 0,0034	0,4168 ± 0,0020	0,999405
PEL	0,0420 ± 0,0026	0,7678 ± 0,0048	0,0344 ± 0,0077	0,1558 ± 0,0025	0,999227
PUR	0,1468 ± 0,0005	0,1324 ± 0,0009	0,0183 ± 0,0018	0,7024 ± 0,0010	0,999744
RPT	0,1901 ± 0,0001	0,0715 ± 0,0001	0,0256 ± 0,0002	0,7129 ± 0,0001	0,999996
ACB	0,8967 ± 0,0001	0,0119 ± 0,0001	-	0,0914 ± 0,0001	0,999998
ASW	0,7765 ± 0,0001	0,0374 ± 0,0001	-	0,1861 ± 0,0001	0,999997
CLM	0,0805 ± 0,0009	0,2714 ± 0,0006	-	0,6481 ± 0,0007	0,999728
MXL	0,0438 ± 0,0011	0,5035 ± 0,0009	-	0,4527 ± 0,0012	0,999624
PEL	0,0450 ± 0,0024	0,7869 ± 0,0022	-	0,1681 ± 0,0025	0,999248
PUR	0,1498 ± 0,0006	0,1410 ± 0,0005	-	0,7092 ± 0,0005	0,999697
RPT	0,1931 ± 0,0001	0,0843 ± 0,0001	-	0,7226 ± 0,0001	0,999985
ACB	0,8924 ± 0,0001	-	0,0229 ± 0,0001	0,0847 ± 0,0001	1
ASW	0,7604 ± 0,0014	-	0,0649 ± 0,0017	0,1747 ± 0,0015	0,999718
CLM	0,0261 ± 0,0115	-	0,4471 ± 0,0135	0,5268 ± 0,0155	0,907635
MXL	-0,0499 ± 0,0169	-	0,6759 ± 0,0223	0,3740 ± 0,0217	0,893276
PEL	-0,0271 ± 0,0264	-	1,0085 ± 0,0186	0,0186 ± 0,0344	0,878386
PUR	0,1422 ± 0,0056	-	0,2335 ± 0,0072	0,6243 ± 0,0069	0,960388
RPT	0,1953 ± 0,0028	-	0,1463 ± 0,0042	0,6584 ± 0,0034	0,99109

A **Figura 11**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 16**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

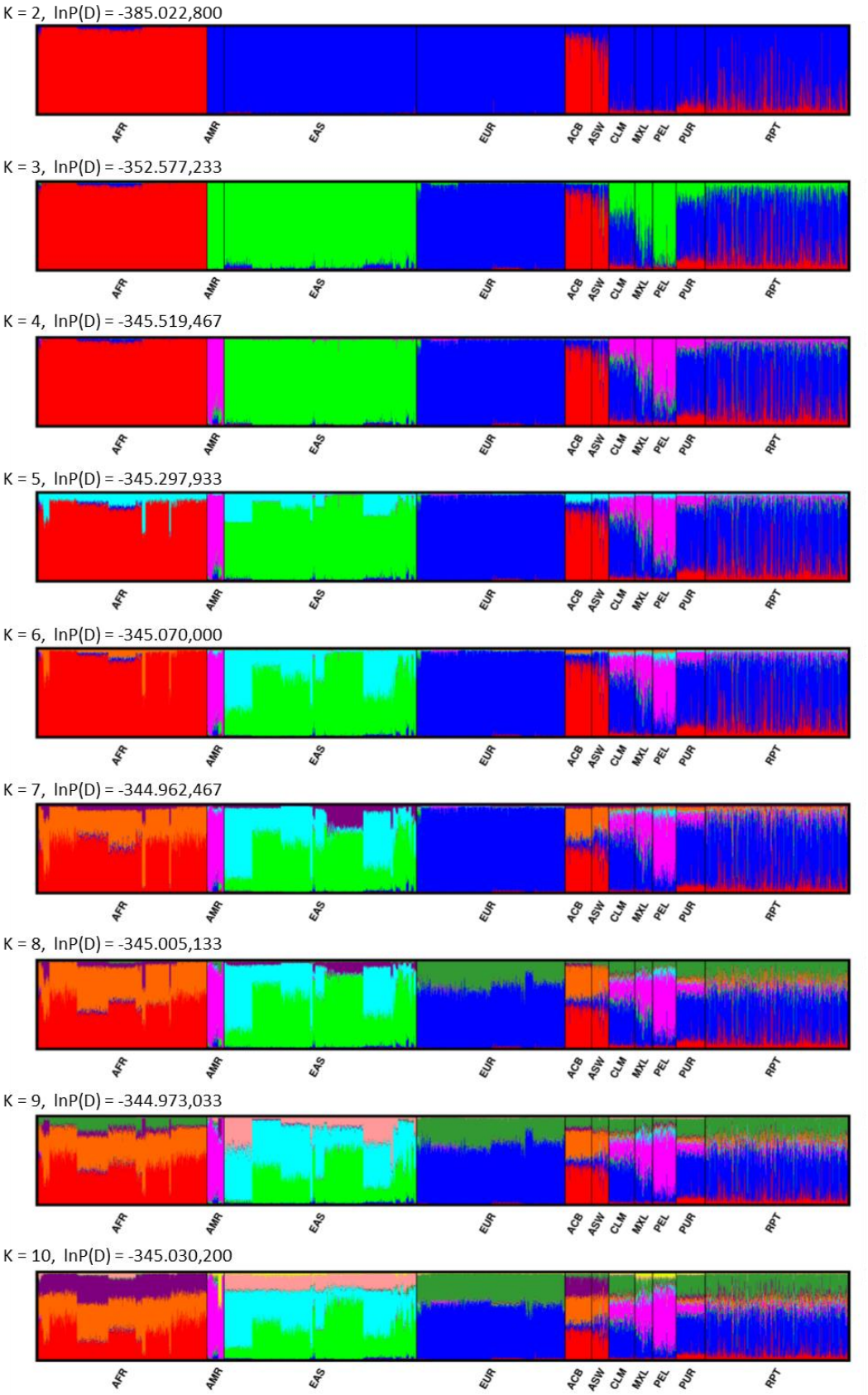


Figura 11. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel desenvolvido por Seldin. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 16. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise de 125 dos 128 marcadores que compõem o painel desenvolvido por Seldin. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,9843**	0,0011	0,0017	0,0129
AMR	0,0102	**0,8822**	0,0662	0,0415
EAS	0,0044	0,0065	**0,9764**	0,0127
EUR	0,0066	0,0065	0,0069	**0,9800**
ACB	0,8823	0,0039	0,0083	0,1056
ASW	0,7665	0,0245	0,0213	0,1877
CLM	0,0774	0,2211	0,0707	0,6309
MXL	0,0432	0,4031	0,1257	0,4280
PEL	0,0515	0,6825	0,0860	0,1800
PUR	0,1508	0,1215	0,0331	0,6946
RPT	0,1906	0,0530	0,0564	0,7000

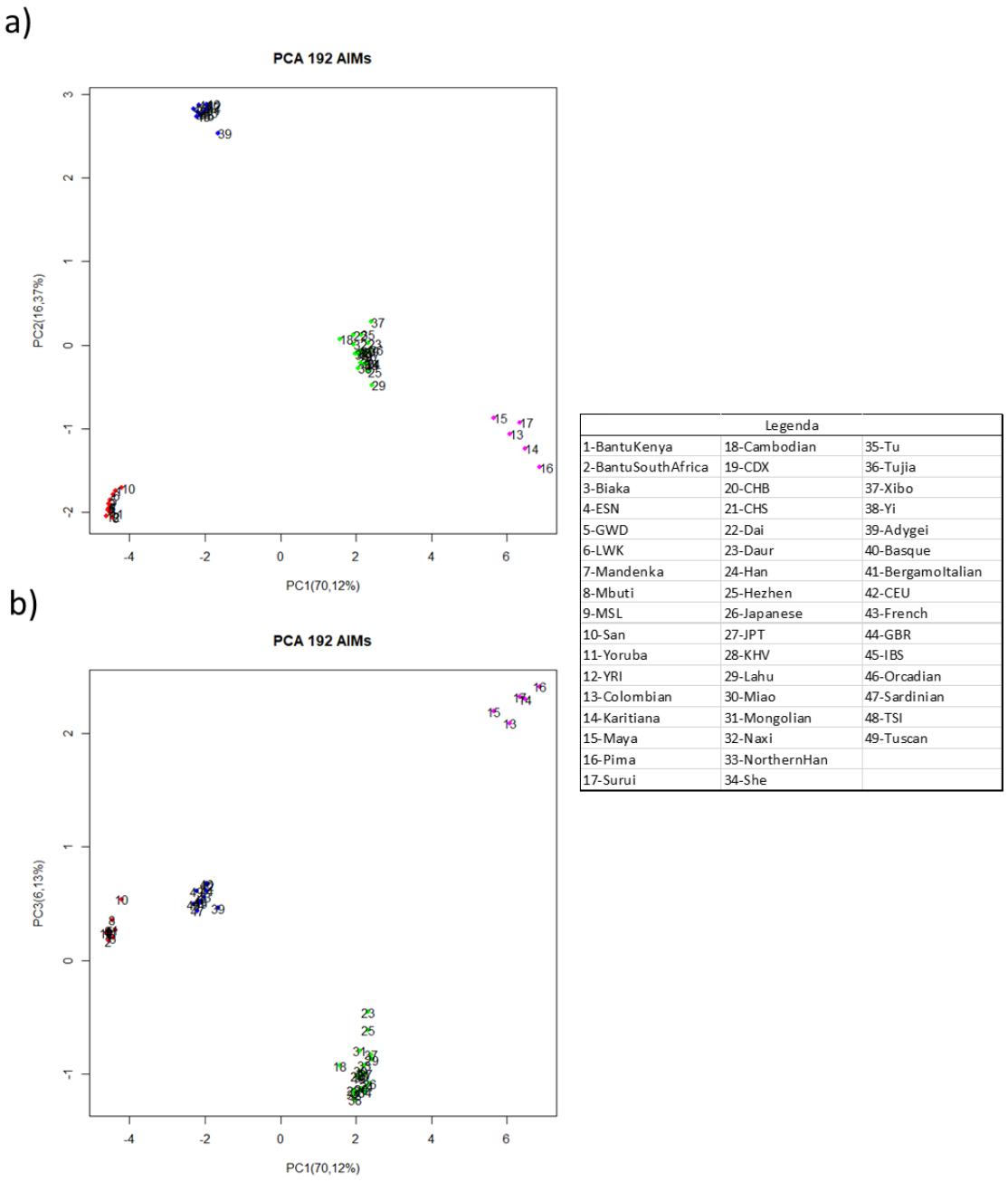
A **Tabela 17** mostra os valores de probabilidade de match combinado (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade combinado (IPC) obtidos para as populações miscigenadas presentes no âmbito do 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando o painel desenvolvido por Seldin (125 AIMs).

Tabela 17. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o painel desenvolvido por Seldin (125 AIMs).

População	PMC	PDC	IPC	PEC
ACB	3,9376E-30	>0,999999999999999	2,84693E-20	0,999906834
ASW	9,2004E-35	>0,999999999999999	1,53914E-16	0,999994621
CLM	2,4497E-44	>0,999999999999999	2,47927E-10	0,999999924
MXL	3,6897E-48	>0,999999999999999	1,47760E-08	0,999999988
PEL	2,6411E-45	>0,999999999999999	3,42758E-09	0,999999981
PUR	3,2014E-41	>0,999999999999999	9,86362E-13	0,999999485
RPT	2,2308E-40	>0,999999999999999	1,45656E-14	0,999997297

6.8 Santos's 192 AIMs (192 AIMs)

A análise de componentes principais (**Figura 12)** resultou em três principais componentes que somados explicam 92,62% da variância, distribuída da seguinte forma: 70,12% para a PC1, 16,37% para a PC2 e de 6,13% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das duas primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-Bergamoltalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 12. Análise de componentes principais baseada nos 192 marcadores que compõem o painel desenvolvido por Santos, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (70,12%) vs PC2 (16,37%); (b) PC1 vs PC3 (6,13%).

Os valores apresentados na **Tabela 18** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 18. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os marcadores genéticos presentes no painel desenvolvido por Santos

População	AFR	AMR	EAS	EUR	R^2
ACB	0,9084 ± 0,0036	0,0059 ± 0,0037	0,0173 ± 0,0071	0,0684 ± 0,0000	0,998708
ASW	0,7994 ± 0,0002	0,0425 ± 0,0002	-0,0063 ± 0,0004	0,1645 ± 0,0000	0,999997
CLM	0,1003 ± 0,0007	0,2717 ± 0,0005	-0,0196 ± 0,0011	0,6477 ± 0,0007	0,999899
MXL	0,0614 ± 0,0008	0,4823 ± 0,0012	0,0057 ± 0,0021	0,4506 ± 0,0000	0,999941
PEL	0,0745 ± 0,0011	0,7603 ± 0,0020	-0,0111 ± 0,0034	0,1763 ± 0,0000	0,999965
PUR	0,1470 ± 0,0007	0,1489 ± 0,0007	-0,0107 ± 0,0016	0,7148 ± 0,0006	0,999558
RPT	0,1902 ± 0,0002	0,0652 ± 0,0003	0,0275 ± 0,0006	0,7171 ± 0,0001	0,999900
ACB	0,8831 ± 0,0002	0,0070 ± 0,0001	-	0,1099 ± 0,0001	0,999995
ASW	0,7978 ± 0,0001	0,0393 ± 0,0001	-	0,1629 ± 0,0001	0,999996
CLM	0,1001 ± 0,0007	0,2622 ± 0,0002	-	0,6377 ± 0,0006	0,999949
MXL	0,0658 ± 0,0006	0,4866 ± 0,0003	-	0,4477 ± 0,0005	0,999944
PEL	0,0721 ± 0,0007	0,7547 ± 0,0003	-	0,1732 ± 0,0007	0,999978
PUR	0,1472 ± 0,0007	0,1432 ± 0,0002	-	0,7095 ± 0,0008	0,999632
RPT	0,2443 ± 0,0115	0,0909 ± 0,0032	-	0,6648 ± 0,0109	0,836719
ACB	0,8771 ± 0,0004	-	0,0137 ± 0,0004	0,1091 ± 0,0003	0,999979
ASW	0,7548 ± 0,0019	-	0,0672 ± 0,0018	0,1780 ± 0,0018	0,999031
CLM	-0,0142 ± 0,0113	-	0,4482 ± 0,0081	0,5660 ± 0,0093	0,960876
MXL	-0,1645 ± 0,0140	-	0,7895 ± 0,0121	0,3749 ± 0,0142	0,969650
PEL	-0,1975 ± 0,0221	-	1,2315 ± 0,0191	-0,0340 ± 0,0260	0,961115
PUR	0,1140 ± 0,0077	-	0,2642 ± 0,0050	0,6219 ± 0,0076	0,936540
RPT	0,2095 ± 0,0092	-	0,1547 ± 0,0071	0,6358 ± 0,0093	0,795764

A **Figura 13**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 19**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

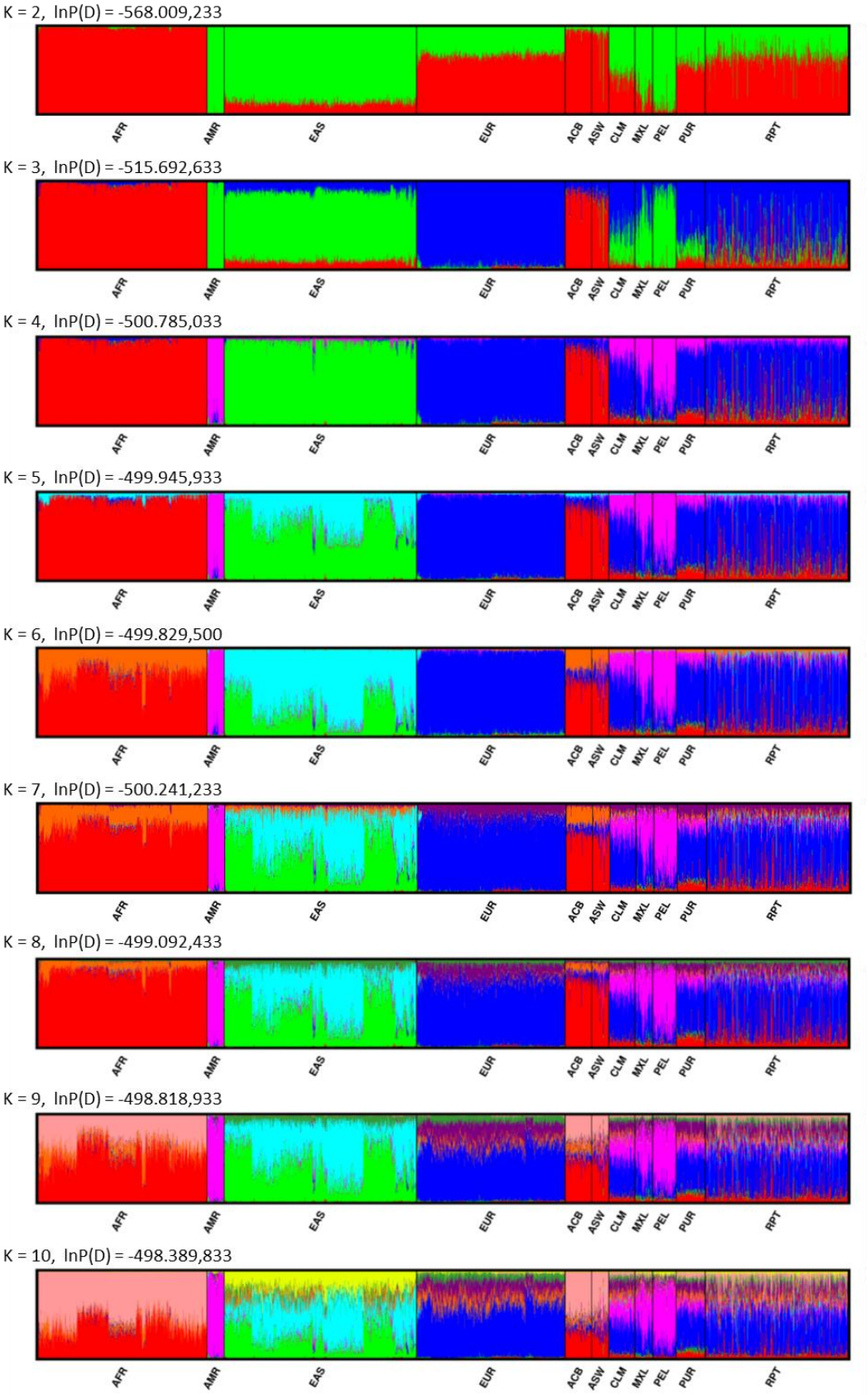


Figura 13. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o painel desenvolvido por Santos. As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 19. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 192 marcadores que compõem o painel desenvolvido por Santos. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,9821**	0,0006	0,0033	0,0141
AMR	0,0024	**0,9391**	0,0221	0,0364
EAS	0,0025	0,0284	**0,9585**	0,0105
EUR	0,0110	0,0095	0,0137	**0,9658**
ACB	0,8703	0,0055	0,0066	0,1176
ASW	0,7877	0,0376	0,0079	0,1668
CLM	0,1013	0,2512	0,0177	0,6298
MXL	0,0566	0,4496	0,0396	0,4543
PEL	0,0530	0,7050	0,0252	0,2168
PUR	0,1569	0,1386	0,0183	0,6862
RPT	0,1897	0,0573	0,0568	0,6963

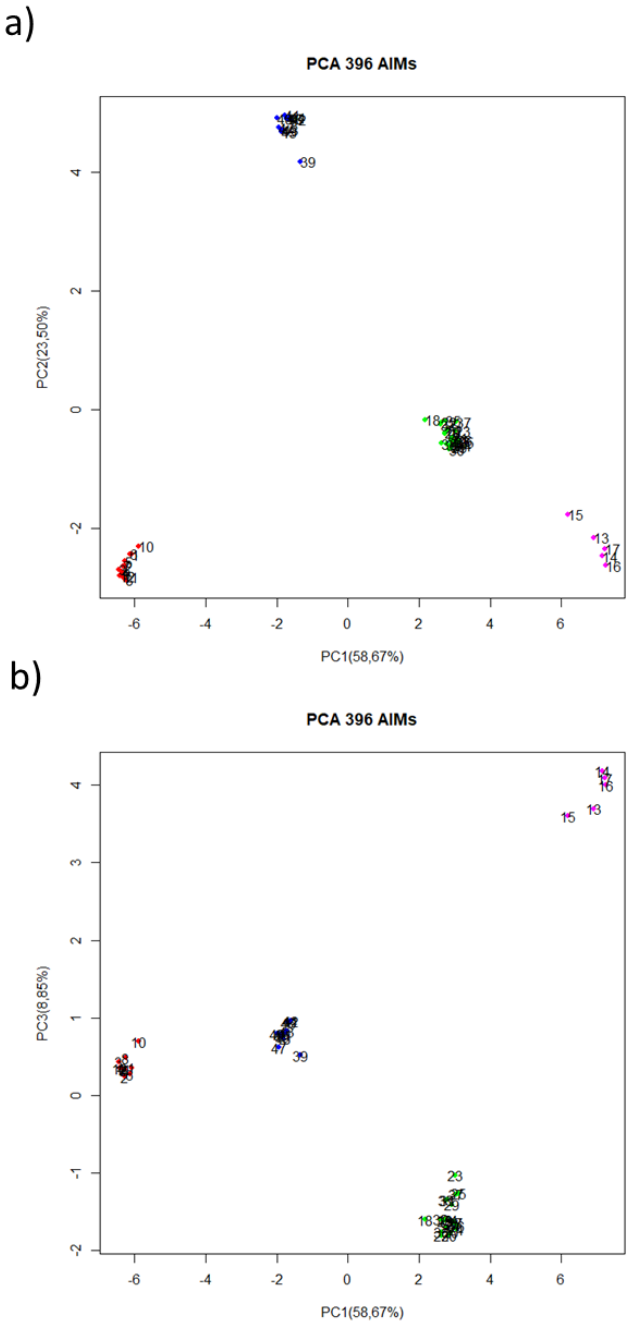
A **Tabela 20** mostra os valores de probabilidade de match combinado (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade combinado (IPC) obtidos para as populações miscigenadas presentes no âmbito do 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando o painel desenvolvido por Santos (192 AIMs).

Tabela 20. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o painel desenvolvido por Santos (192 AIMs).

População	PMC	PDC	IPC	PEC
ACB	7,1457E-34	>0,999999999999999	9,03950E-40	0,9994180190673
ASW	1,2383E-42	>0,999999999999999	2,24208E-34	0,9999777547431
CLM	7,4118E-72	>0,999999999999999	3,51835E-11	0,9999999999998
MXL	3,1714E-77	>0,999999999999999	1,22917E-12	0,9999999999994
PEL	8,8729E-63	>0,999999999999999	2,24027E-21	0,9999999978028
PUR	4,8219E-65	>0,999999999999999	9,24555E-18	0,9999999998795
RPT	1,5329E-60	>0,999999999999999	1,43798E-23	0,9999999778421

6.9 Conjunto de Todos os Painéis (396 AIMs)

A análise de componentes principais (**Figura 14)** resultou em três principais componentes que somados explicam 91,02% da variância, distribuída da seguinte forma: 58,67% para a PC1, 23,50% para a PC2 e de 8,85% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das duas primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-Bergamoltalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 14. Análise de componentes principais baseada nos 396 marcadores presentes no trabalho, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (70,12%) vs PC2 (16,37%); (b) PC1 vs PC3 (6,13%).

Os valores apresentados na **Tabela 21** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 21. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os 396 marcadores presentes no trabalho.

População	AFR	AMR	EAS	EUR	R^2
ACB	0,8897 ± 0,0001	0,0106 ± 0,0001	-0,0014 ± 0,0002	0,1010 ± 0,0000	0,999998
ASW	0,7799 ± 0,0000	0,0441 ± 0,0000	-0,0101 ± 0,0001	0,1861 ± 0,0000	1
CLM	0,0864 ± 0,0000	0,2536 ± 0,0000	0,0170 ± 0,0000	0,6430 ± 0,0000	1
MXL	0,0467 ± 0,0002	0,4730 ± 0,0004	0,0355 ± 0,0006	0,4448 ± 0,0000	0,999962
PEL	0,0546 ± 0,0004	0,7530 ± 0,0006	0,0076 ± 0,0011	0,1849 ± 0,0000	0,999965
PUR	0,1426 ± 0,0006	0,1380 ± 0,0006	0,0095 ± 0,0013	0,7099 ± 0,0008	0,998785
RPT	0,7049 ± 0,0000	0,0673 ± 0,0000	0,0255 ± 0,0000	0,7049 ± 0,0000	0,999998
ACB	0,8895 ± 0,0001	0,0100 ± 0,0001	-	0,1005 ± 0,0001	0,999997
ASW	0,7780 ± 0,0000	0,0390 ± 0,0000	-	0,1830 ± 0,0000	0,999999
CLM	0,0894 ± 0,0000	0,2622 ± 0,0000	-	0,6484 ± 0,0000	1
MXL	0,0516 ± 0,0002	0,4907 ± 0,0001	-	0,4578 ± 0,0002	0,999967
PEL	0,0555 ± 0,0003	0,7569 ± 0,0002	-	0,1876 ± 0,0004	0,999964
PUR	0,1460 ± 0,0007	0,1422 ± 0,0003	-	0,7118 ± 0,0007	0,998609
RPT	0,2066 ± 0,0000	0,0801 ± 0,0000	-	0,7133 ± 0,0000	0,999992
ACB	0,8820 ± 0,0003	-	0,0158 ± 0,0004	0,1022 ± 0,0005	0,999944
ASW	0,7469 ± 0,0012	-	0,0574 ± 0,0015	0,1957 ± 0,0018	0,998977
CLM	0,0098 ± 0,0063	-	0,4072 ± 0,0062	0,5830 ± 0,0080	0,915231
MXL	-0,1064 ± 0,0091	-	0,6802 ± 0,0115	0,4261 ± 0,0138	0,899719
PEL	-0,1205 ± 0,0144	-	1,0168 ± 0,0190	0,1037 ± 0,0210	0,879527
PUR	0,1260 ± 0,0036	-	0,2354 ± 0,0034	0,6386 ± 0,0049	0,939144
RPT	0,1959 ± 0,0021	-	0,1399 ± 0,0019	0,6642 ± 0,0027	0,975707

A **Figura 15**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 22**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

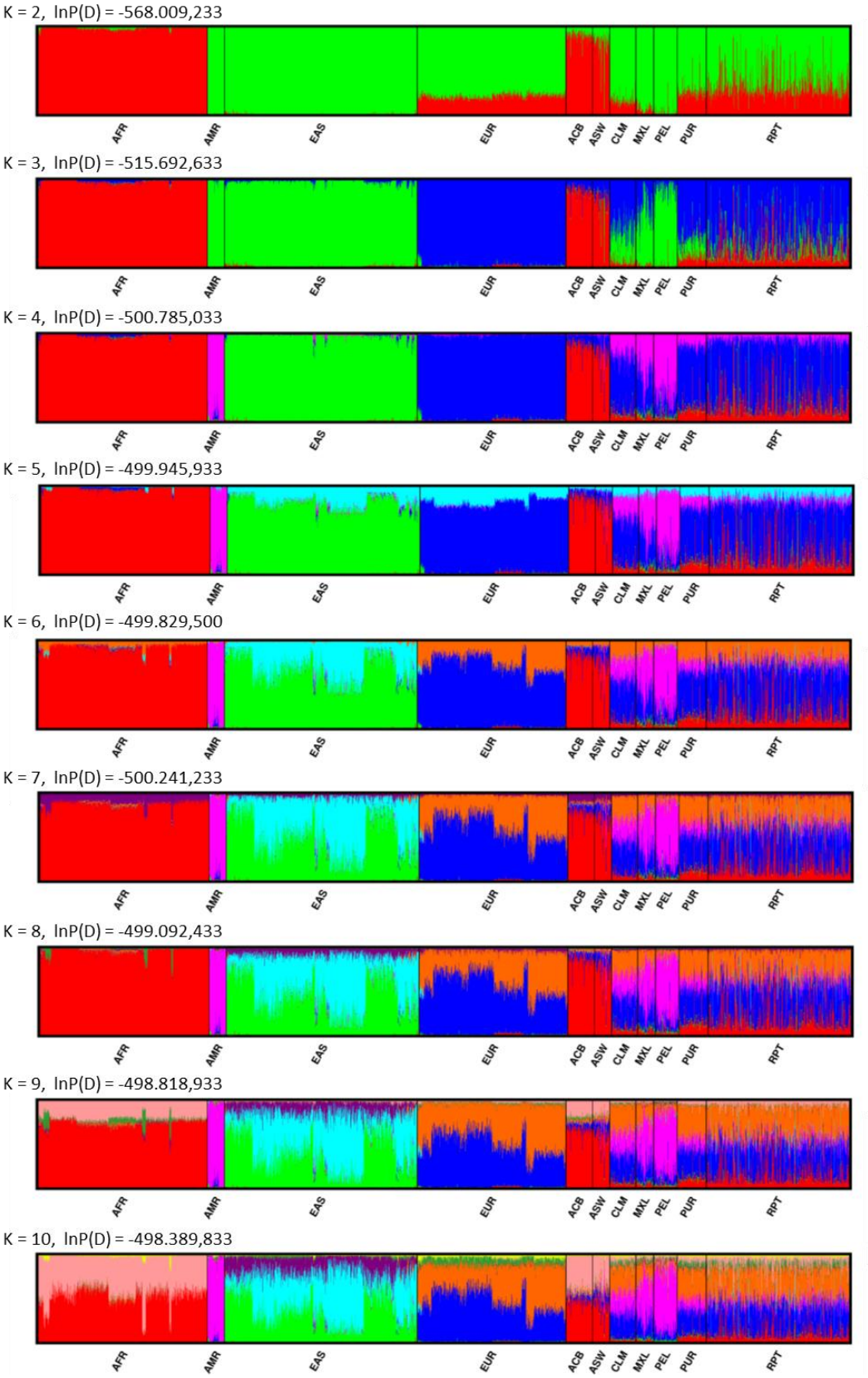


Figura 15. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o conjunto de todos os painéis do trabalho (396 AIMs). As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 22. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (K = 4), resultante da análise dos 396 marcadores que compõem o conjunto de todos os painéis do trabalho. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (n = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,9875**	0,0004	0,0019	0,0102
AMR	0,0049	**0,9454**	0,0162	0,0335
EAS	0,0032	0,0075	**0,9823**	0,0070
EUR	0,0090	0,0052	0,0060	**0,9798**
ACB	0,8799	0,0063	0,0055	0,1083
ASW	0,7702	0,0353	0,0022	0,1923
CLM	0,0914	0,2473	0,0175	0,6438
MXL	0,0544	0,4526	0,0392	0,4538
PEL	0,0467	0,7106	0,0276	0,2151
PUR	0,1534	0,1341	0,0178	0,6947
RPT	0,2107	0,0616	0,0361	0,6916

A **Tabela 23** mostra os valores de probabilidade de match combinado (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade combinado (IPC) obtidos para as populações miscigenadas presentes no âmbito do 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando o conjunto de todos os painéis presentes no trabalho (396 AIMs).

Tabela 23. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o conjunto de todos os painéis presentes no trabalho (396 AIMs).

População	PMC	PDC	IPC	PEC
ACB	1,6500E-81	>0,999999999999999	9,3485E-72	0,99999999989552
ASW	6,0250E-99	>0,999999999999999	2,9924E-60	0,99999999999996
CLM	5,4969E-139	>0,999999999999999	1,8237E-29	>0,999999999999999
MXL	6,3109E-148	>0,999999999999999	3,5167E-29	>0,999999999999999
PEL	1,8661E-127	>0,999999999999999	1,2936E-41	>0,999999999999999
PUR	2,6700E-129	>0,999999999999999	4,9621E-39	>0,999999999999999
RPT	1,7123E-124	>0,999999999999999	9,5353E-47	>0,999999999999999

6.9.1 Seleção dos Painéis

Os marcadores que melhor individualizam as populações de referência estão apresentados na **Tabela 24,** a seguir. A soma dos módulos das diferenças entre as frequências alélicas para a população escolhida para ser individualizada quando comparadas às outras populações foi chamada de Diferença Absoluta, e pode atingir um valor máximo de 3,0. Um total de 33 AISNPs foram selecionados seguindo as regras estipuladas previamente (i.e., diferença nas frequências alélicas maior do que 30%) a única exceção, rs671, apresentando diferença de 0,1746 das outras referências, foi adicionado devido ao fato de apresentar variabilidade na somente na população asiática e divido menor quantidade de marcadores que individualizam a população asiática. O painel resultante foi intitulado de "Painel Ancestralidade".

Tabela 24. Lista dos marcadores informativos de ancestralidade que melhor individualizam as populações de referência, de acordo com suas frequências alélicas.

Pop	AIM	Frequência Alelo de Referência				Diferença Absoluta
		AFR	AMR	EAS	EUR	
AFR	rs2814778	0,0124	0,9836	1,0000	0,9935	2,9400
	rs10173381	0,0173	0,9918	0,9899	0,7959	2,7258
	rs10497191	0,9531	0,0738	0,0902	0,1283	2,5672
	rs7753890	0,0847	0,9836	0,8468	0,9803	2,5566
	rs1871534	0,9250	0,0246	0,0000	0,0066	2,7440
	rs1561937	0,8964	0,0000	0,0332	0,0955	2,5604
	rs3916235	0,8758	0,0246	0,0036	0,0478	2,5515
	rs310644	0,0559	0,9508	0,9647	0,9242	2,6719
	rs2823850	0,1184	0,9918	0,9286	0,9925	2,5576
AMR	rs12402499	1,0000	0,2705	1,0000	0,9287	2,1172
	rs3737576	0,9877	0,2705	0,9156	0,9560	2,0478
	rs10179812	0,9531	0,2295	0,8030	0,9128	1,9804
	rs12498138	0,9984	0,0902	0,9132	0,9363	2,5774
	rs2471552	0,0666	0,9426	0,2016	0,1854	2,3743
	rs2051827	0,9482	0,1667	0,9256	0,9504	2,3241

Pop	AIM	Frequência Alelo de Referência				Diferença Absoluta
		AFR	AMR	EAS	EUR	
	rs17359176	0,9901	0,1475	0,8276	0,9335	2,3086
	rs1958099	0,9622	0,0820	0,9653	0,8165	2,4981
	rs4780476	0,1916	0,9426	0,2525	0,2912	2,0925
EAS	rs1229984	0,0000	0,0164	0,6655	0,0449	1,9351
	rs3811801	1,0000	1,0000	0,5123	0,9972	1,4604
	rs671	1,0000	1,0000	0,8254	1,0000	0,5238
	rs1800414	1,0000	1,0000	0,4233	0,9991	1,7292
	rs881929	0,9531	0,5984	0,1133	0,5696	1,7811
	rs1701930	0,3059	0,0246	0,8980	0,4139	1,9495
EUR	rs11804831	0,0403	0,0328	0,0145	0,8267	2,3927
	rs16891982	0,9975	0,9508	0,9921	0,0852	2,6848
	rs28777	0,8454	0,9426	0,8838	0,0646	2,4780
	rs2196051	0,0132	0,0328	0,0036	0,7069	2,0712
	rs10962599	0,9975	0,9590	0,9971	0,2865	2,0941
	rs853577	0,7327	0,8115	0,7840	0,0637	2,1372
	rs9522149	0,9695	0,9417	0,9834	0,2238	2,2232
	rs1426654	0,0519	0,0328	0,0166	0,9981	2,8931
	rs11637235	0,9638	0,9917	0,9906	0,2350	2,2411

Os marcadores que melhor discriminam os indivíduos dentro da amostra populacional de Ribeirão Preto, sem foco primário na ancestralidade, estão apresentados na **Tabela 25.** Um total de 50 AISNPs foram selecionados e o painel resultante foi intitulado de "Painel Identificação Humana". O **Apêndice A** contêm os parâmetros forenses para todos os 396 marcadores.

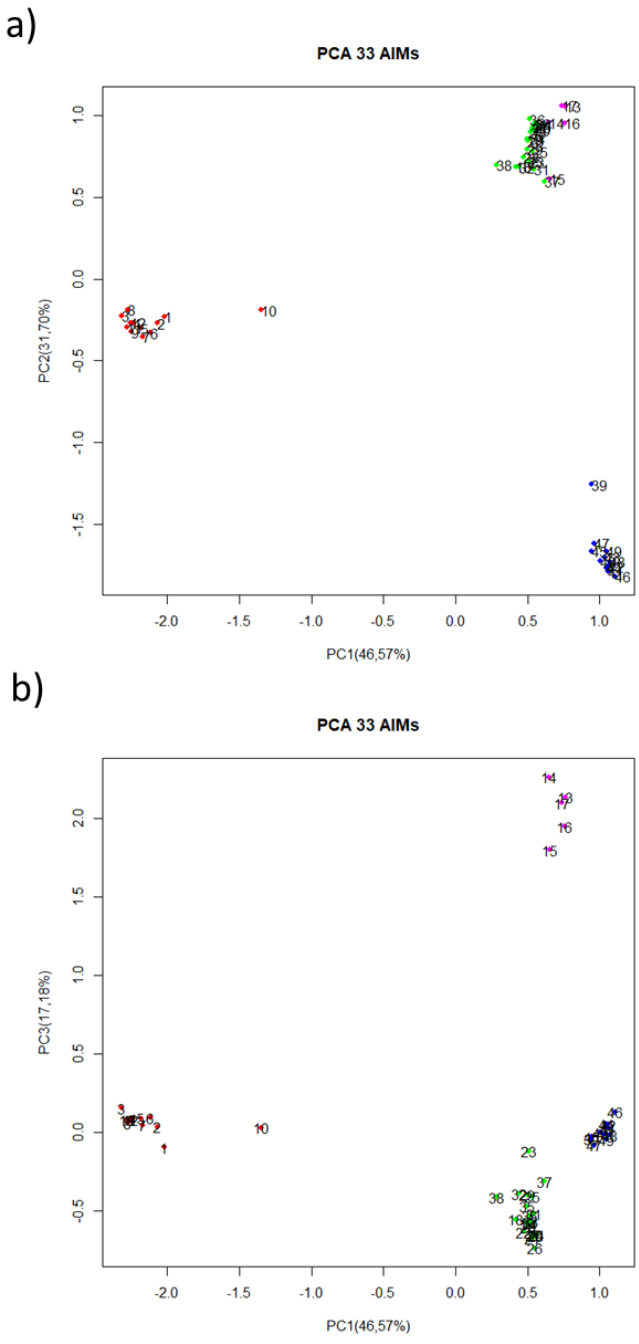
Tabela 25. Lista dos marcadores informativos de ancestralidade que melhor discriminam os indivíduos de Ribeirão Preto, de acordo com os valores de PM e PD.

AIM	Cromossomo	PM	PD	PE
rs11804831	1	0,36544	0,63456	0,11487
rs7554936	1	0,35987	0,64013	0,15635
rs1040404	1	0,36541	0,63459	0,17021
rs10127540	1	0,36030	0,63970	0,13646
rs798443	2	0,37128	0,62872	0,13246
rs7421394	2	0,36061	0,63939	0,13782
rs1863086	2	0,37073	0,62927	0,17906
rs6548616	3	0,37456	0,62544	0,17829
rs9845457	3	0,35091	0,64909	0,13979
rs1498444	3	0,37260	0,62740	0,12219
rs4540055	4	0,21776	0,78224	0,24411
rs385194	4	0,36127	0,63873	0,16240
rs316598	5	0,37328	0,62672	0,16706
rs16891982	5	0,35014	0,64986	0,09835
rs10079352	5	0,36459	0,63541	0,13512
rs6880706	5	0,35343	0,64657	0,14665
rs2504853	6	0,36892	0,63108	0,14902
rs4463276	6	0,35126	0,64874	0,12983
rs6570644	6	0,34577	0,65423	0,12853
rs1871428	6	0,36874	0,63126	0,16313
rs2330442	7	0,36446	0,63554	0,15339
rs705308	7	0,36315	0,63685	0,12344
rs7803075	7	0,34734	0,65266	0,13246
rs12544346	8	0,36188	0,63812	0,16240
rs2196051	8	0,36914	0,63086	0,17665
rs7844723	8	0,37059	0,62941	0,16394
rs6993205	8	0,35442	0,64558	0,13378
rs10962599	9	0,35742	0,64258	0,15486
rs201381710	10	0,25135	0,74865	0,32040

AIM	Cromossomo	PM	PD	PE
rs7902158	10	0,36140	0,63860	0,13246
rs3814614	10	0,36947	0,63053	0,16773
rs10839880	11	0,36290	0,63710	0,16394
rs5030240	11	0,27308	0,72692	0,23886
rs2238151	12	0,35596	0,64404	0,13782
rs1335873	13	0,37219	0,62781	0,16549
rs1886510	13	0,37419	0,62581	0,15339
rs1409842	13	0,35926	0,64074	0,15635
rs9522149	13	0,36622	0,63378	0,15046
rs1950993	14	0,36239	0,63761	0,16394
rs3784230	14	0,36626	0,63374	0,17180
rs12913832	15	0,36663	0,63337	0,08869
rs11637235	15	0,35245	0,64755	0,13570
rs2052074	17	0,37484	0,62516	0,11607
rs2125345	17	0,36215	0,63785	0,15936
rs874299	18	0,36708	0,63292	0,15339
rs1024116	18	0,36307	0,63693	0,16549
rs8113143	19	0,36945	0,63055	0,17341
rs2532060	19	0,35980	0,64020	0,15192
rs2269160	21	0,36663	0,63337	0,13538
rs4821004	22	0,35388	0,64612	0,14758

6.10 Painel Ancestralidade (33 AIMs)

A análise de componentes principais (**Figura 16)** resultou em três principais componentes que somados explicam 90,36% da variância, distribuída da seguinte forma: 45,19% para a PC1, 40,36% para a PC2 e de 4,81% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das três primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-Bergamoltalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 16 Análise de componentes principais baseada nos 33 marcadores presentes no Painel Ancestralidade, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (45,19%) vs PC2 (40,36%); (b) PC1 vs PC3 (4,81%).

Os valores apresentados na **Tabela 26** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 26. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os 33 marcadores presentes no Painel Ancestralidade.

População	AFR	AMR	EAS	EUR	R^2
ACB	0,8979 ± 0,0004	0,0078 ± 0,0004	-0,0132 ± 0,0006	0,1075 ± 0,0002	0,999999
ASW	0,7819 ± 0,0001	0,0439 ± 0,0001	-0,0127 ± 0,0002	0,1869 ± 0,0000	1
CLM	0,0722 ± 0,0007	0,2291 ± 0,0004	0,0354 ± 0,0011	0,6633 ± 0,0010	0,999918
MXL	0,0455 ± 0,0024	0,4519 ± 0,0027	0,0523 ± 0,0035	0,4503 ± 0,0015	0,999254
PEL	0,0485 ± 0,0008	0,7482 ± 0,0009	0,0073 ± 0,0013	0,1961 ± 0,0000	0,999996
PUR	0,1437 ± 0,0000	0,1376 ± 0,0000	0,0179 ± 0,0000	0,7009 ± 0,0000	1
RPT	0,1916 ± 0,0002	0,0631 ± 0,0004	0,0423 ± 0,0005	0,7030 ± 0,0005	0,999980
ACB	0,8951 ± 0,0003	0,0020 ± 0,0005	-	0,1029 ± 0,0004	0,999998
ASW	0,7786 ± 0,0001	0,0390 ± 0,0002	-	0,1825 ± 0,0001	1
CLM	0,0786 ± 0,0011	0,2444 ± 0,0007	-	0,6770 ± 0,0011	0,999786
MXL	0,0709 ± 0,0016	0,4683 ± 0,0021	-	0,4608 ± 0,0018	0,999641
PEL	0,0488 ± 0,0010	0,7521 ± 0,0007	-	0,1991 ± 0,0007	0,999989
PUR	0,1485 ± 0,0000	0,1449 ± 0,0000	-	0,7066 ± 0,0000	1
RPT	0,2036 ± 0,0003	0,0812 ± 0,0004	-	0,7153 ± 0,0003	0,999969
ACB	0,8927 ± 0,0012	-	-0,0035 ± 0,0019	0,1108 ± 0,0011	0,999981
ASW	0,7599 ± 0,0047	-	0,0327 ± 0,0062	0,2074 ± 0,0029	0,999757
CLM	0,0654 ± 0,0163	-	0,2134 ± 0,0267	0,7212 ± 0,0215	0,854104
MXL	0,0468 ± 0,0231	-	0,3754 ± 0,0273	0,5778 ± 0,0097	0,985436
PEL	0,1060 ± 0,0302	-	0,5331 ± 0,0333	0,3608 ± 0,0222	0,964434
PUR	0,1542 ± 0,0104	-	0,1375 ± 0,0193	0,7084 ± 0,0176	0,934727
RPT	0,2020 ± 0,0045	-	0,1060 ± 0,0089	0,6920 ± 0,0087	0,988860

A **Figura 17**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 27**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

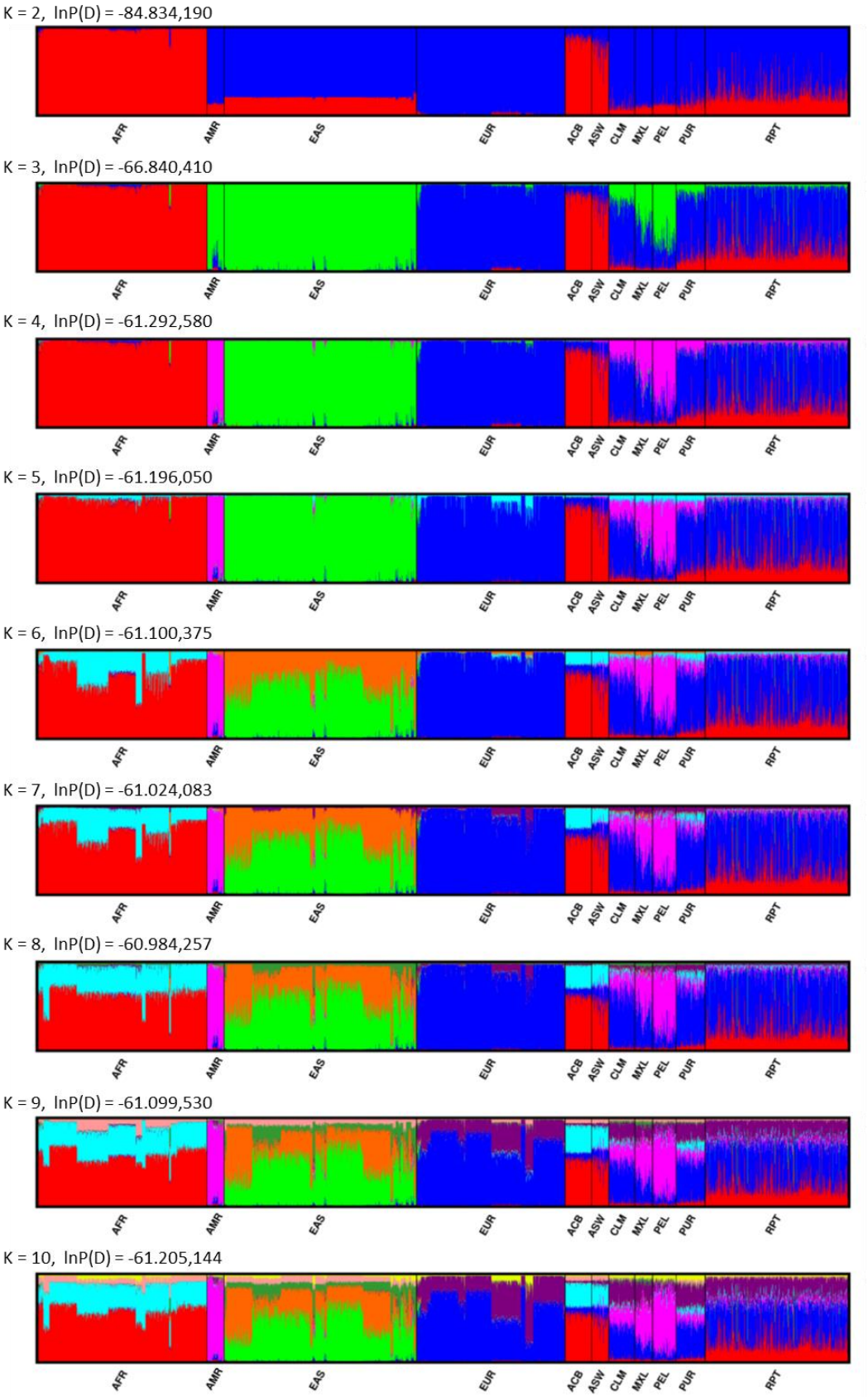


Figura 17. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o Painel Ancestralidade (33 AIMs). As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 27. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (*K* = 4), resultante da análise dos 33 marcadores que compõem o Painel Ancestralidade. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,984**	0,002	0,004	0,011
AMR	0,014	**0,922**	0,020	0,044
EAS	0,002	0,007	**0,980**	0,010
EUR	0,008	0,004	0,012	**0,977**
ACB	0,883	0,005	0,007	0,105
ASW	0,773	0,033	0,002	0,192
CLM	0,087	0,225	0,011	0,678
MXL	0,050	0,436	0,038	0,475
PEL	0,057	0,706	0,009	0,228
PUR	0,159	0,133	0,006	0,702
RPT	0,199	0,055	0,042	0,703

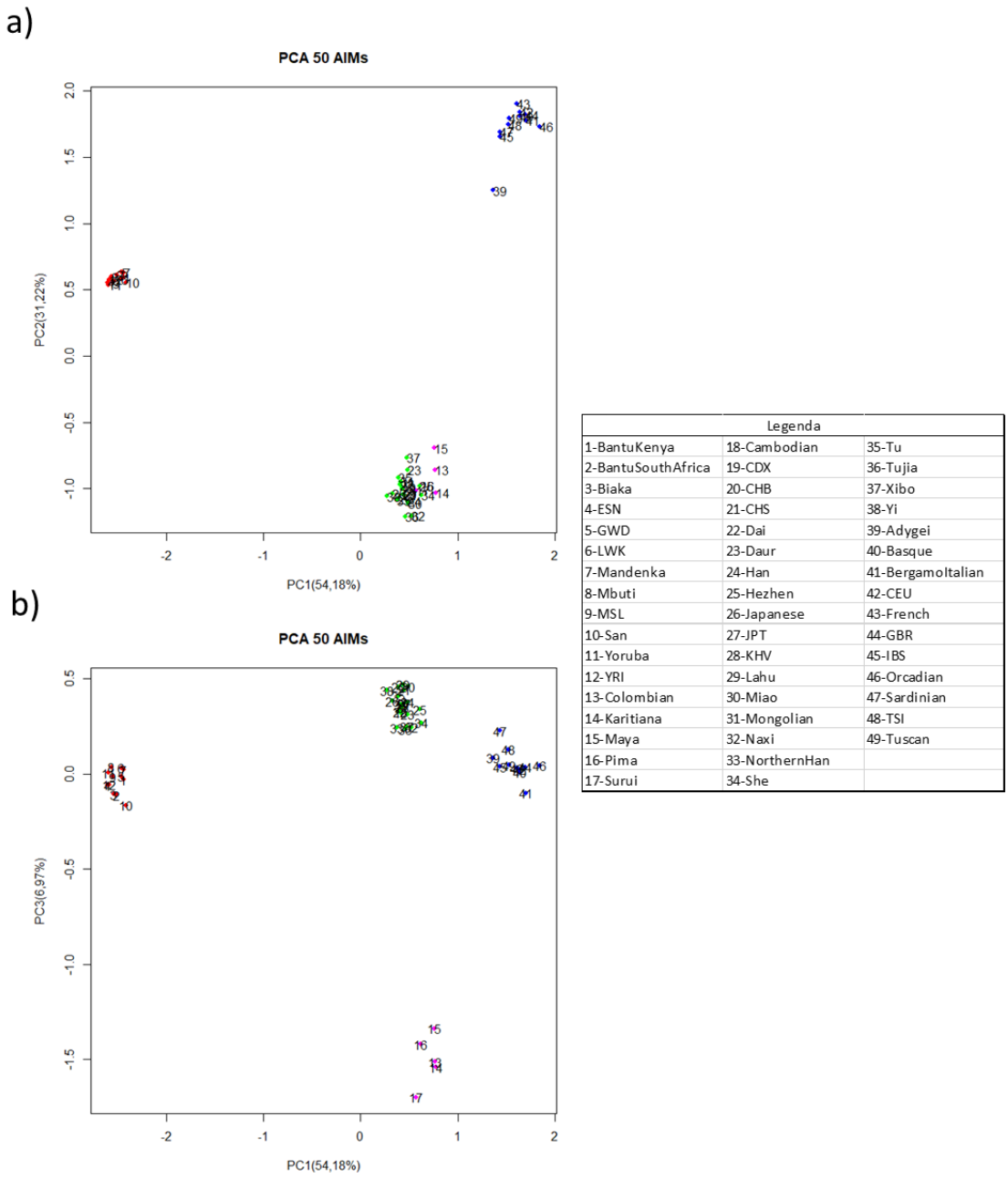
A **Tabela 28** mostra os valores de probabilidade de match combinado (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade combinado (IPC) obtidos para as populações miscigenadas presentes no âmbito do 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando o Painel Ancestralidade (33 AIMs).

Tabela 28. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando o Painel Ancestralidade (33 AIMs).

População	PMC	PDC	IPC	PEC
ACB	5,5907E-06	0,99999440926578	1,3710E-07	0,733691608
ASW	4,0681E-08	0,99999995931919	4,2456E-06	0,899854086
CLM	3,1416E-10	0,99999999968584	7,5046E-05	0,960001571
MXL	2,2988E-10	0,99999999977012	6,0502E-05	0,960813659
PEL	1,5845E-08	0,99999998415499	4,6604E-06	0,901613254
PUR	3,4852E-10	0,99999999965148	5,2902E-05	0,952244714
RPT	4,1314E-10	0,99999999958686	2,3296E-05	0,933190463

6.11 Painel Identificação Humana (50 AIMs)

A análise de componentes principais (**Figura 18)** resultou em três principais componentes que somados explicam 92,37% da variância, distribuída da seguinte forma: 54,18% para a PC1, 31,22% para a PC2 e de 6,97% para a PC3. De modo geral, as populações dos quatro grupos principais foram separadas com sucesso por meio das três primeiras componentes.



Legenda		
1-BantuKenya	18-Cambodian	35-Tu
2-BantuSouthAfrica	19-CDX	36-Tujia
3-Biaka	20-CHB	37-Xibo
4-ESN	21-CHS	38-Yi
5-GWD	22-Dai	39-Adygei
6-LWK	23-Daur	40-Basque
7-Mandenka	24-Han	41-Bergamoltalian
8-Mbuti	25-Hezhen	42-CEU
9-MSL	26-Japanese	43-French
10-San	27-JPT	44-GBR
11-Yoruba	28-KHV	45-IBS
12-YRI	29-Lahu	46-Orcadian
13-Colombian	30-Miao	47-Sardinian
14-Karitiana	31-Mongolian	48-TSI
15-Maya	32-Naxi	49-Tuscan
16-Pima	33-NorthernHan	
17-Surui	34-She	

Figura 18. Análise de componentes principais baseada nos 50 marcadores presentes no Painel Identificação Humana, considerando populações europeias (azul), africanas (vermelho), leste-asiáticas (verde) e ameríndias (magenta) presente nos bancos de dados 1KGEN e HGDP. (a) PC1 (54,18%) vs PC2 (31,22%); (b) PC1 vs PC3 (6,97%).

Os valores apresentados na **Tabela 29** consistem nos valores inferidos para a ancestralidade populacional das populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto, estimadas utilizando o software Admix 95 e modelos tetra- e tri-híbridos (neste último caso, variando entre as populações de referência AMR e EAS).

Tabela 29. Ancestralidade populacional inferida pelo software Admix 95 para as populações miscigenadas analisadas pelo projeto 1KGEN e para a população de Ribeirão Preto (*n* = 513), utilizando modelos tetra- e tri-híbridos, e os 50 marcadores presentes no Painel Identificação Humana.

População	AFR	AMR	EAS	EUR	R^2
ACB	0,9020 ± 0,0003	0,0019 ± 0,0005	-0,0081 ± 0,0006	0,1042 ± 0,0002	1
ASW	0,7687 ± 0,0022	0,1255 ± 0,0032	-0,0809 ± 0,0041	0,1867 ± 0,0012	0,999977
CLM	0,0944 ± 0,0001	0,2337 ± 0,0002	0,0362 ± 0,0002	0,6356 ± 0,0001	0,999994
MXL	0,0555 ± 0,0003	0,4347 ± 0,0009	0,0854 ± 0,0009	0,4244 ± 0,0004	0,999986
PEL	0,0825 ± 0,0020	0,7787 ± 0,0046	-0,0566 ± 0,0056	0,1955 ± 0,0016	0,999873
PUR	0,1689 ± 0,0004	0,1492 ± 0,0011	0,0136 ± 0,0013	0,6683 ± 0,0007	0,99976
RPT	0,2110 ± 0,0006	0,0706 ± 0,0011	0,0129 ± 0,0014	0,7055 ± 0,0010	0,999596
ACB	0,9004 ± 0,0002	-0,0032 ± 0,0004	-	0,1028 ± 0,0002	1
ASW	0,7528 ± 0,0013	0,0737 ± 0,0022	-	0,1735 ± 0,0011	0,999982
CLM	0,1014 ± 0,0000	0,2569 ± 0,0000	-	0,6417 ± 0,0000	1
MXL	0,0719 ± 0,0005	0,4901 ± 0,0009	-	0,4380 ± 0,0005	0,999978
PEL	0,0707 ± 0,0017	0,7427 ± 0,0020	-	0,1866 ± 0,0011	0,99991
PUR	0,1719 ± 0,0003	0,1566 ± 0,0008	-	0,6716 ± 0,0008	0,999858
RPT	0,2131 ± 0,0005	0,0799 ± 0,0011	-	0,7070 ± 0,0011	0,999693
ACB	0,9020 ± 0,0004	-	-0,0066 ± 0,0006	0,1046 ± 0,0003	0,999999
ASW	0,7666 ± 0,0056	-	0,0249 ± 0,0087	0,2085 ± 0,0038	0,999733
CLM	0,0923 ± 0,0026	-	0,2289 ± 0,0066	0,6788 ± 0,0049	0,995209
MXL	0,0524 ± 0,0072	-	0,4660 ± 0,0114	0,4819 ± 0,0059	0,995946
PEL	0,0924 ± 0,0165	-	0,6130 ± 0,0234	0,2946 ± 0,0143	0,981877
PUR	0,1676 ± 0,0011	-	0,1369 ± 0,0033	0,6955 ± 0,0029	0,998902
RPT	0,2096 ± 0,0011	-	0,0726 ± 0,0031	0,7180 ± 0,0030	0,998942

A **Figura 19**, apresenta gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três corridas do STRUCTURE que obtiveram o maior valor possível para lnP(D) para um mesmo valor de *K*. Cada barra vertical representa um indivíduo e cada cor representa um *cluster* e não necessariamente uma população de referência. A **Tabela 30**, demostra o quanto de cada população está inserida dentro de cada cluster para $K = 4$.

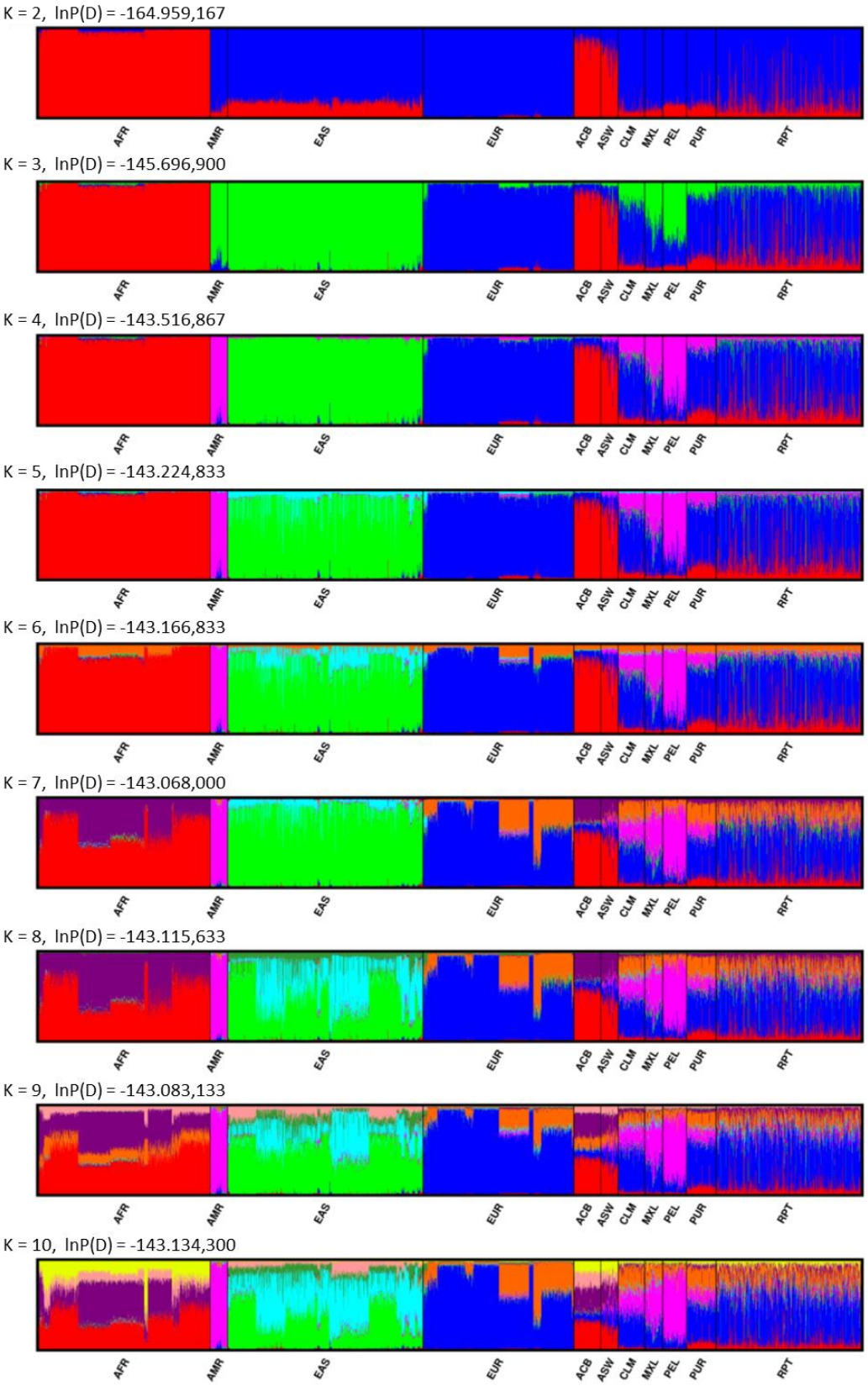


Figura 19. Gráficos de barras gerados pelo programa Clumpp, cada um baseado nas três melhores corridas do STRUCTURE para um mesmo valor de *K* (variando de 2 a 10), utilizando o Painel Identificação Humana (50 AIMs). As populações ACB, ASW, CLM, MXL, PEL e PUR provém do projeto 1KGEN e RPT refere-se à população de Ribeirão Preto (*n* = 513).

Tabela 30. Proporção da ancestralidade de cada população atribuída a cada um dos quatro clusters (K = 4), resultante da análise dos 50 marcadores que compõem o Painel Identificação Humana. AFR, AMR, EAS e EUR referem-se aos grupos de populações analisadas pelos projetos 1KGEN e do HGDP consideradas como parentais na presente análise. ACB, ASW, CLM, MXL, PEL e PUR são as populações americanas miscigenadas presentes no 1KGEN e RPT refere-se à população de Ribeirão Preto (n = 513).

POP	Cluster			
	1	2	3	4
AFR	**0,9848**	0,0021	0,0059	0,0072
AMR	0,0017	**0,9557**	0,0080	0,0345
EAS	0,0065	0,0128	**0,9721**	0,0086
EUR	0,0132	0,0147	0,0196	**0,9525**
ACB	0,8859	0,0037	0,0079	0,1025
ASW	0,7459	0,0737	0,0032	0,1772
CLM	0,1037	0,2291	0,0501	0,6172
MXL	0,0538	0,4441	0,0778	0,4242
PEL	0,0571	0,7316	0,0092	0,2021
PUR	0,1755	0,1530	0,0237	0,6479
RPT	0,2146	0,0468	0,0613	0,6773

A **Tabela 31** mostra os valores de probabilidade de match combinado (PMC), poder de exclusão combinado (PEC), poder de discriminação combinado (PDC) e do índice de paternidade combinado (IPC) obtidos para as populações miscigenadas presentes no âmbito do 1KGEN e nos indivíduos coletados randomicamente da população de Ribeirão Preto (n = 343) utilizando Painel Identificação Humana (50 AIMs).

Tabela 31. Parâmetros forenses combinados das populações miscigenadas presentes no 1KGEN e na amostra populacional de Ribeirão Preto utilizando Painel Identificação Humana (50 AIMs).

População	PMC	PDC	IPC	PEC
ACB	1,1133E-10	0,999999999888671	5,7437E-10	0,935883113
ASW	6,4976E-15	0,999999999999993	5,8765E-07	0,993799214
CLM	2,1434E-21	>0,999999999999999	1,1038E-01	0,999920208
MXL	4,0511E-20	>0,999999999999999	2,0685E-03	0,999563951
PEL	4,0831E-16	>0,999999999999999	1,7491E-05	0,997579910
PUR	6,4698E-22	>0,999999999999999	3,4618E-01	0,999956166
RPT	6,1818E-23	>0,999999999999999	6,3803E-02	0,999884034

Discussão

7. Discussão

7.1 Imputação

Os sistemas de imputação *Sanger Imputation Service* (SIS) e *Michigan Imputation Server* (MIS), são disponibilizados online, onde apenas o preparo dos arquivos VCF é necessário para upload ao site onde as posteriores análises serão realizadas. Apenas um arquivo por cromossomo é necessário, onde os próximos passos, tais como seleção das populações de referência para comparação e escolha de parâmetros, são todos feitos nos próprios sites dos servidores.

O Impute2 é um software disponível para *download* em sistema operacional Linux. Para a imputação de regiões específicas ele pode vir a ser mais prático, porém como a imputação foi realizada de maneira a obtermos a maior quantidade possível de marcadores para cada um dos cromossomos, o tempo tomado para a realização do processo foi elevado, levando cerca de um mês. O processo pode ser automatizado a partir de linhas de comando, porém, de acordo com o manual de boas práticas disponibilizados no site do software, o programa apresenta menor número de erros de imputação quando realizado de 5 em 5 milhões de pares de base, o que levou a um total de 800 linhas de comando, 400 para o pré-faseamento, passo obrigatório, e 400 linhas para a imputação em si.

A partir da **Tabela 2,** observa-se a alta compatibilidade entre as posições sequenciadas a partir do NGS e as posições genotipadas a partir do MEGA, com apenas 0,975% das posições sendo divergentes, cerca de 222 *loci* dos 29.705 comparados, demonstrando a qualidade das metodologias. Embora o sistema de imputação Impute2 não tenha atribuído valores para 6,76% das posições (3.676 *loci*), as posições que foram imputadas apresentaram alta compatibilidade com os outros sistemas de imputação (>98,3%) e para as posições compartilhadas com o NGS a compatibilidade de 90,79% foi observada.

Após a união das três metodologias foram excluídos 1,26% dos genótipos imputados (692 *loci* de 54.891), sendo que as posições mantidas apresentaram 99,734% de compatibilidade com o sequenciamento de nova geração, demonstrando uma alta qualidade para que as próximas análises sejam realizadas.

7.2 Desvios do Equilíbrio de Hardy-Weinberg

O princípio de Hardy-Weinberg segue algumas regras, dentre elas: (i) casamento aleatórios, (ii) a ausência de seleção natural, (iii) uma população grande o bastante para que a deriva genética seja insignificante, (iv) ausência de fluxo genético ou migração, (v) ausência de mutações e (vi) ao se trabalhar com uma amostra, sua coleta deverá ser realizada de maneira aleatória. (Lachance, 2016).

Análises GWAS utilizam testes de qualidade e filtragem onde amostras com alto índice de *missings* ou SNPs em desvio do EHW são removidos, no entanto, tais filtros são difíceis de serem definidos em amostragens miscigenadas e podem acabar removendo marcadores importantes para definição da estrutura populacional (Li et al., 2020). Neste trabalho, para estudos de parâmetros forenses, apenas os indivíduos coletados de maneira randômica foram utilizados, enquanto que para a ancestralidade, foi utilizada a amostragem completa.

Ao se trabalhar com a amostra completa (513 indivíduos) estamos violando uma das regras essenciais para o cálculo do equilíbrio de Hardy-Weinberg, a amostragem aleatória, já que 170 indivíduos não foram coletados de maneira randômica, devido ao objetivo de aumentar a quantidade de fenótipos raros, como a presença de sardas, ruivos e portadores de enfermidades relacionadas à pigmentação da pele.

Com a remoção desses indivíduos, 46 marcadores apresentaram desvio (11,62%). Esta quantidade de desvios não pode ser simplesmente atribuída ao acaso ($\chi^2 = 36{,}4934$; $p = 1{,}53 \times 10^{-9}$), podendo então estar relacionada a eventos histórico-demográficos diversos (afinal trata-se de uma população miscigenada ainda estruturada), seleção natural (visto que vários dos marcadores consistem em SNPs alvo de seleção) e até mesmo a erros de genotipagem, sequenciamento ou imputação (15 desses marcadores foram genotipados pelo MEGA, 7 desses marcadores foram analisados por NGS e 16 marcadores foram genotipados e sequenciados) ou de imputação (8 desses marcadores não estavam presentes no NGS ou no MEGA). Entretanto, considerando que em 44 destes 46 desvios observamos deficiência de heterozigotos, o que também não pode ser atribuído ao acaso ($\chi^2 = 42{,}0870$; $p = 8{,}73 \times 10^{-11}$), podemos com segurança associar estes desvios à estruturação populacional (efeito Wahlund). Não obstante, com a correção de Bonferroni para múltiplas comparações (Armstrong, 2014) sendo aplicada ($\alpha_{BONFERRONI} = 0{,}05/396 = 0{,}000126$), se a amostragem inteira for considerada

para o cálculo de equilíbrio, 18 marcadores apresentam desvio do EHW, mas quando se é tomado em conta apenas os indivíduos coletados de maneira randômica, somente 1 dos marcadores encontrou-se em desequilíbrio, o rs2572307, marcador que estava presente apenas no NGS e que faz parte do conjunto SNP*for*ID 34-plex. Tal marcador não é polimórfico nas populações asiáticas, europeias e ameríndias, mas é na população Africana, apresentando MAF de 47,9%. O fato de a ancestralidade europeia ter participado de eventos de miscigenação mais recentes na região sudeste do país (Kehdy et al., 2015), um maior número de homozigotos tendem a ser observados para o rs2572307, levando ao aumento da divergência entre heterozigoses observadas e esperadas em nossa amostra.

7.3 PCAs

A análise de componentes principais é uma ferramenta matemática importante para várias disciplinas de estudo, onde um conjunto muito grande de variáveis correlatas são transformadas em uma quantidade menor de componentes, explicando uma porcentagem relativamente alta da variância total observada. É importante tomarmos algumas notas de precaução: (i) a utilização de duas componentes pode trazer uma variância alta quando usadas em conjunto, porém, as outras componentes geradas não podem ser negligenciadas pois podem trazer informações importantes para os seus dados; (ii) podemos ter conhecimento prévio de nossa amostragem, mas não podemos utilizar esse preceito ao definir qual a porcentagem de variância deve ser considerado significante e (iii) uma pequena variação de porcentagem pode nos trazer informações importantes sobre um estudo (Johnson et al., 2007). Todas as análises apresentaram uma explicação da variância total acima de 90% com as três primeiras componentes, com exceção do painel de Seldin, com 88,58% (**Tabela 32**).

O painel PIMA não foi desenvolvido para ser utilizado sozinho, e sim como um complemento ao SNP*for*ID 34-plex, de maneira a melhorar a performance deste último quando aplicado em populações que apresentam miscigenação com populações nativo americanas (Gontijo et al., 2020). Tendo esse conhecimento, essa separação é percebida nas PCAs, **Figura 2,** onde a PC1 (65,92% da variância total) separa as populações AFR, EAS e EUR da população AMR, a PC2 (21,49%) separa as populações EUR, AFR e AMR da população EAS, e a PC3 (5,73%) acaba por espalhar as populações EAS, o que confirma a premissa de separar a população AMR das outras populações de referência.

Tabela 32. Variância observada para cada componente principal em todos os painéis analisados aplicados às populações de referência (1KGEN + HGDP).

Painel	PC1	PC2	PC3	PC1+PC2	Total
PIMA	65,92	21,49	5,73	87,41	93,14
34-plex	46,83	38,42	5,11	85,25	90,36
Kidd Lab set of 55 AISNPs	53,12	31,09	8,87	84,21	93,08
PIMA + 34-plex	50,66	24,74	15,61	75,40	91,01
***Seldin's* 125 AIMs**	49,00	31,50	8,08	80,50	88,58
***Santos's* 192 AIMs**	70,12	16,37	6,13	86,49	92,62
396 AIMs	58,67	23,50	8,85	82,17	91,02
Painel Ancestralidade	46,57	31,70	17,18	78,27	95,45
Painel Identificação Humana	54,18	31,22	6,97	85,40	92,37

Para o SNP*for*ID 34-plex, painel montado com o foco na separação entre as populações AFR, EAS e EUR (Phillips et al., 2007), na **Figura 4a** onde a PC1 (46,83%) e a PC2 (38,42%) estão plotadas, são observados apenas três agrupamentos, onde em um desses agrupamentos, as populações EAS e AMR estão sobrepostas. A PC1 também dispõe no mesmo ponto três populações, incluindo então a população EUR, quem vem a ser separada pela PC2. A PC3 (5,11%) individualiza o agrupamento formado por populações AMR (**Figura 4b**), mas observa-se que os grupos EUR e EAS se sobrepõem, porém com um valor explicado da variância total bem pequeno.

Os painéis PIMA e SNP*for*ID 34-plex quando utilizados em conjunto nos traz um resultado melhorado quando comparado aos PCAs anteriores. Na **Figura 8a** vemos que a PC1 (50,66%) sozinha já é capaz de separar as quatro populações de referência em seu eixo, enquanto sua origem separa as populações europeia e africana das populações ameríndia e asiática. A PC2 (24,74%) isola a população europeia das outras populações de referência. Já a **Figura 8b** mostra que a PC3 (15,61%) separa em sua origem as populações africanas e ameríndias das populações asiáticas e europeias. Em nenhuma das duas situações existe sobreposição de populações.

O painel *Kidd Lab set of 55 AISNPs* foi montado com o objetivo de separar as seis grandes populações do planeta (Kidd et al., 2014). Visualiza-se na **Figura 6a,** três grandes agrupamentos com ausência de separação entre AMR e EAS, com a população europeia separada pela origem da PC2 (31,09%), e a população africana isolada pela origem da PC1. A

PC3 (8,87%) separa a população AMR a partir da origem das outras populações de referência e, se comparado ao 34-plex, a **Figura 6b** não agrupa as populações asiáticas e europeias.

O painel de Seldin foi desenvolvido para utilização nos Estados Unidos da América, tomando as quatro populações utilizadas para a realização deste projeto como referência (Kosoy et al., 2009) e apresenta mais do que o dobro de marcadores presente nos painéis discutidos anteriormente. Tal painel também não apresenta sobreposições de agrupamentos em nenhuma das plotagens da **Figura 10**. Assim como na **Figura 8a,** observa-se na **Figura 10a** um isolamento da população africana no eixo da PC1 (49,00%), ainda mais expressivo que o observado no painel anterior, mas é importante observar que a PC2 (31,50%) atribui o mesmo valor para as populações africanas e asiáticas. A primeira componente também já é capaz de separar todas as populações de referência em seu eixo. A PC3 (8,08%) separa na origem as populações asiáticas das populações ameríndias, e separa também as populações AFR das populações EAS, estando a segunda população aquém da origem do eixo. Todas as subpopulações apresentam-se mais agrupadas do que quando comparadas aos painéis anteriores, com exceção das ameríndias, que apresenta melhor agrupamento nos painéis *Kidd Lab set of 55 AISNPs* e PIMA + 34-plex.

O painel de Santos, desenvolvido como um número mínimo de marcadores para a inferência acurada da ancestralidade da população brasileira (Santos et al., 2016) apresenta o maior valor para explicação da variabilidade total em uma única componente, PC1 (70,12%), onde cada uma das populações encontra-se bem distantes uma das outras. A origem separa as populações africanas e europeias das populações asiáticas e ameríndias. A **Figura 12a** mostra também que a segunda componente principal (16,37%) também não agrupa nenhuma das populações em um mesmo ponto, isolando a população europeia através da origem do seu eixo. A PC3 (6,13%), **Figura 12b**, isola a população asiática das outras populações de referência, apresentando-se na porção negativa do seu eixo, porém deixando as populações europeias e africanas mais próximas. Este painel apresenta então um grande poder de separação das populações referências, onde as duas primeiras componentes explicam 86,49% da variância total da amostragem, cerca de 2% de diferença da variância explicada pelas três primeiras componentes obtidas para o painel de Seldin, e que pode estar diretamente relacionada à quantidade de marcadores utilizadas neste painel.

O resultado observado quando todos os 396 marcadores foram utilizados em conjunto foi muito próximo do resultado observado para o painel de Santos (192 AIMs), que representa

48,5% do total de SNPs. Na **Figura 14a** é observado o isolamento da população europeia a partir da origem na PC2 (23,50) de maneira mais expressiva quando comparada à **Figura 12a.** A PC1 (58,67%) separa a partir de sua origem as populações europeias e africanas, na porção negativa, das populações asiáticas e ameríndias, nas porções positivas, deixando todas elas bem distantes uma das outras em seu eixo. Na **Figura 12b**, nota-se novamente a proximidade das populações europeias e africanas no eixo da PC3 (8,85%), com a separação a partir de sua origem das populações asiáticas das outras populações de referência. Nota-se então que o aumento do número de marcadores não traz diferenças relativas com relação ao agrupamento das populações de referência nas PCAs.

O Painel Ancestralidade (33 AIMs) traz na **Figura 16a** o isolamento das populações africanas através da origem da PC1 (46,57%) das outras populações de referência. As populações asiáticas, europeias e ameríndias encontram-se próximas umas das outras, porém é possível ver que se agrupam separadamente, a PC2 (31,70%) separa, através de sua origem, as populações africanas e europeias, em sua porção negativa, e as populações asiáticas e ameríndias, em sua porção positiva. A PC3 (17,18%), **Figura 16b,** separa as populações ameríndias, na porção positiva, das populações asiáticas na porção negativa. As populações africanas e europeias estão dispostas na origem da terceira componente principal, sendo que em todas as análises, é a PC3 que apresenta a maior porcentagem da variância total quando comparada às outras análises. Embora a subpopulação San apresente-se como *outlier* com a primeira componente, as PC2 e PC3 agrupam muito bem a população africana. Os agrupamentos das populações de referência são bem definidos, mesmo com uma quantidade pequena de marcadores.

O Painel Identificação Humana (50 AIMs) traz na **Figura 18a,** o isolamento da população africana através da origem no eixo da PC1 (54,18%) das outras populações de referência. A PC2 (31,22%) separa as populações AFR e EUR das populações EAS e AMR através de sua origem. As populações asiáticas e ameríndias estão dispostas muito próximas uma da outra, porém a PC3 (6,97%), **Figura 18b** separa através da origem, com a população asiática, na porção positiva, e a população ameríndia na porção negativa, enquanto que as populações africanas e europeias se encontram na origem. Mesmo que este painel tenha sido desenvolvido visando a informatividade forense e não a diferenciação entre as populações de referência, os marcadores presentes no Painel Identificação Humana apresentaram um

agrupamento bem definido das populações de referência, o que demonstra que existe também a possibilidade de sua utilização, com resultados expressivos, para testes de ancestralidade.

7.4 Inferência de Ancestralidade

Tendo conhecimento de sua complexa composição ancestral, estudos envolvendo populações brasileiras aplicam metodologias tri- (i.e., africana, europeia e ameríndia) ou tetra-híbridas (i.e., africana, europeia, ameríndia e oriental) para o estudo de inferência de ancestralidade biogeográfica (Gomes et al., 2018; Naslavsky et al., 2021). Como alguns painéis não apresentam poder estatístico para a separação entre as referências ameríndias e leste asiáticas, a atribuição de inferência de ancestralidade torna-se idêntica, em sua terceira componente, para ambos os modelos tri-híbridos. Isso torna possível a utilização da população EAS como referência quando indivíduos nativo americanos não estão disponíveis (Oliveira et al., 2018; Valle-Silva et al., 2019). Nesse contexto, Valle-Silva et al. (2019) determinaram a ascendência tri-parental de Ribeirão Preto usando o painel SNP*for*ID 52-plex e inferiram, variando ameríndios e asiáticos orientais como terceiro componente, uma faixa de 0,74-0,79 ascendência europeia, 0,16-0,21 ascendência africana e 0,04-0,05 ameríndia ou ascendência do leste asiático.

O principal evento fundador que originou os ameríndios ocorreu quando pequenos grupos da Ásia cruzaram a Beringia para as Américas, tornando essas duas populações geneticamente relacionadas (Ribeiro-dos-Santos et al., 2020; Wood et al., 2019). Embora a estreita relação genética entre as duas populações esteja presente, estudos forenses usam marcadores do mtDNA (Álvarez-Iglesias et al., 2006; Wood et al., 2019) e do cromossomo Y (Páez-Riberos et al., 2006; Su et al., 1999) para estabelecer haplogrupos ameríndios e asiáticos. No entanto, marcadores autossômicos revelam gradientes de diversidade genética humana em vez de descontinuidades entre ameríndios e asiáticos orientais (Serre & Pääbo, 2004), exigindo o uso de dados SNP em larga escala ou a seleção de marcadores específicos para diferenciar as populações AMR e EAS. Posto isto, Andrade et al., (2018) avaliou o poder estatístico de um conjunto de 61 AIMs (InDels) capazes de atribuir ascendência asiática na cidade de São Paulo, como o quarto componente de um modelo tetra-híbrido (Andrade et al., 2018).

É importante diferenciar esses dois grupos populacionais quando se trata de uma população altamente miscigenada, pois as ancestralidades ameríndias e asiáticas estão correlacionadas com diferentes fenótipos de interesse médico ou forense. Por exemplo, a ancestralidade ameríndia está relacionada a proteção contra a doença de Alzheimer (Marca-Ysabel et al., 2021) ou à asma e rinite (Salam et al., 2015) quando comparadas à ancestralidade asiática. A avaliação precisa das proporções de ancestralidade AMR ou EAS é de extrema importância para evitar erros de interpretação em estudos de associação sobre o impacto dessas ancestralidades genômicas na resistência ou suscetibilidade de certas doenças (Cardena et al., 2014; Chua et al., 2015; Lee et al., 2022; Marca-Ysabel et al., 2021).

De acordo com as análises utilizando Admix95, as populações miscigenadas presentes no 1KGEN apresentaram resultados muito semelhantes quando são utilizados modelos tetra-híbridos e tri-hibridos com a remoção da referência asiática. Porém, a utilização da população EAS como substituta de AMR levou a grandes distorções nos resultados, como a atribuição de ancestralidades negativas e maiores do que 100% como observado na população PEL para os painéis de Seldin e Santos e valores de R^2 muito baixos, os tornando inutilizáveis devido ao modelo não ser representativo da variabilidade dos dados. Essa situação pode decorrer do fato de que as populações asiáticas e ameríndias são melhores definidas pelos painéis onde tal ocorrência é observada, (PIMA, PIMA + 34-plex, os painéis de Seldin e Santos e o Painel Ancestralidade), ocorrendo erro ao compelir a atribuição a uma ancestralidade não relacionada com as populações que apresentam maior ancestralidade AMR, como as populações PEL e MXL.

A amostra de população brasileira apresentou melhores resultados na inferência de ancestralidade populacional quando um modelo tetra-parental foi utilizado. Esse achado é corroborado pelo trabalho de Andrade et al. (2018), onde através da análise de 61 INDEL-AIMs de 1050 indivíduos no estado de São Paulo, foi demonstrado que existe a possibilidade de utilização das parentais AMR e EAS de maneira permutativa, apresentando valores de 12% e 14% respectivamente; porém, quando as duas populações foram utilizadas na análise além das populações EUR e AFR, observou-se que esse fator de ancestralidade era atribuído de maneira diferente para cada população, 6,6% para a população ameríndia e 9,8% para a população asiática (Andrade et al., 2018). Na presente amostra, o valor esperado é de uma ancestralidade AMR maior do que a EAS, de acordo com o resultado obtido em Marcorin, (2021), porém a utilização da população EAS como substituta adequada para a população

AMR também foi visualizada em nossa amostragem, com exceção dos painéis desenvolvidos por Santos e por Seldin.

A partir da **Tabela 33** é possível observar quais painéis obtiveram os resultados mais próximos de suas referências para cada população. Para a população ACB, os melhores painéis foram o painel desenvolvido por Santos (192 AIMs) com 1,4% de diferença absoluta e o Painel Ancestralidade (33 AIMs), desenvolvido para esse projeto, com 3,0% de diferença. Para a população ASW, cinco painéis apresentaram a mesma diferença de 9,6%, sendo eles: *Kidd Lab set of 55 AISNPs*, o painel de Seldin, o painel de Santos, o Painel Ancestralidade e o conjunto dos 396 AIMs do projeto. Para a população CLM, os painéis PIMA e SNP*for*ID 34-plex quando utilizados em conjunto (59 AIMs) apresentou uma diferença de 2,0%, enquanto que o Painel Ancestralidade apresentou diferença de 2,4%. A população MXL teve como melhores painéis, o PIMA + 34-plex, com diferença de 4,6% e o Painel Ancestralidade com 5,6%. Para a população PEL, os painéis SNP*for*ID 34-plex e o *Kidd Lab set of 55 AISNPs* se apresentaram como os melhores painéis, com diferenças de 1,8% e 2,5%, respectivamente. Na população PUR, os painéis de Santos e de Seldin demonstraram-se com as menores diferenças, para painéis já desenvolvidos, com valores de 4,5% e 4,6%, respectivamente. Em Ribeirão Preto, os painéis que apresentaram resultados mais próximos do esperado foram o PIMA + 34-plex (1,5%) e o de Seldin (2,3%). Todos os resultados obtidos para as populações miscigenadas do 1KGEN foram muito próximos do esperado. Para a população de Ribeirão Preto todos os painéis apresentaram resultados muito próximos do esperado, e demonstraram a ancestralidade AMR maior do que a ancestralidade EAS.

Tabela 33. Diferença absoluta entre as inferências de ancestralidade populacional utilizando o software Admix 95 e os parâmetros de referência (Martin et al., 2017 e Marcorin, 2021). As abreviações estão presentes no Apêndice B.

Pop	Painel	AFR	AMR	EAS	EUR	Diferença Absoluta
ACB	**PIMA**	0,924	0,025	-	0,051	0,131
	34-plex	0,889	0,016	-	0,094	0,045
	Kidd Lab set of 55 AISNPs	0,885	0,024	-	0,091	0,052
	PIMA + 34-plex	0,896	0,013	-	0,092	0,051
	***Seldin's* 125 AIMs**	0,897	0,012	-	0,091	0,051
	***Santos's* 192 AIMs**	0,883	0,007	-	0,110	0,014

Pop	Painel	AFR	AMR	EAS	EUR	Diferença Absoluta
	396 AIMs	0,890	0,010	-	0,101	0,033
	Painel Ancestralidade	0,895	0,002	-	0,103	0,030
	Painel Identificação Humana	0,889	0,016	-	0,094	0,045
	WGS	0,880	0,003	-	0,117	-
ASW	**PIMA**	0,753	0,071	-	0,176	0,080
	34-plex	0,749	0,052	-	0,198	0,043
	Kidd Lab set of 55 AISNPs	0,776	0,043	-	0,182	0,063
	PIMA + 34-plex	0,749	0,059	-	0,192	0,056
	***Seldin's* 125 AIMs**	0,777	0,037	-	0,186	0,054
	***Santos's* 192 AIMs**	0,798	0,039	-	0,163	0,100
	396 AIMs	0,778	0,039	-	0,183	0,060
	Painel Ancestralidade	0,779	0,039	-	0,183	0,061
	Painel Identificação Humana	0,753	0,074	-	0,174	0,085
	WGS	0,756	0,031	-	0,213	-
CLM	**PIMA**	0,062	0,241	-	0,698	0,064
	34-plex	0,093	0,241	-	0,666	0,032
	Kidd Lab set of 55 AISNPs	0,085	0,239	-	0,676	0,036
	PIMA + 34-plex	0,088	0,247	-	0,666	0,020
	***Seldin's* 125 AIMs**	0,081	0,271	-	0,648	0,035
	***Santos's* 192 AIMs**	0,100	0,262	-	0,638	0,056
	396 AIMs	0,089	0,262	-	0,648	0,034
	Painel Ancestralidade	0,079	0,244	-	0,677	0,024
	Painel Identificação Humana	0,101	0,257	-	0,642	0,048
	WGS	0,078	0,257	-	0,666	-
MXL	**PIMA**	0,005	0,466	-	0,529	0,083
	34-plex	0,025	0,524	-	0,450	0,108
	Kidd Lab set of 55 AISNPs	0,051	0,519	-	0,430	0,115
	PIMA + 34-plex	0,042	0,492	-	0,466	0,043
	***Seldin's* 125 AIMs**	0,044	0,504	-	0,453	0,069
	***Santos's* 192 AIMs**	0,066	0,487	-	0,448	0,079
	396 AIMs	0,052	0,491	-	0,458	0,059
	Painel Ancestralidade	0,071	0,468	-	0,461	0,056
	Painel Identificação Humana	0,072	0,490	-	0,438	0,098
	WGS	0,043	0,470	-	0,487	-
PEL	**PIMA**	0,039	0,713	-	0,248	0,119
	34-plex	0,019	0,782	-	0,200	0,018

Pop	Painel	AFR	AMR	EAS	EUR	Diferença Absoluta
	Kidd Lab set of 55 AISNPs	0,030	0,781	-	0,189	0,025
	PIMA + 34-plex	0,044	0,740	-	0,216	0,066
	***Seldin's* 125 AIMs**	0,045	0,787	-	0,168	0,068
	***Santos's* 192 AIMs**	0,072	0,755	-	0,173	0,094
	396 AIMs	0,056	0,757	-	0,188	0,061
	Painel Ancestralidade	0,049	0,752	-	0,199	0,048
	Painel Identificação Humana	0,071	0,743	-	0,187	0,091
	WGS	0,025	0,773	-	0,202	-
PUR	**PIMA**	0,155	0,142	-	0,703	0,058
	34-plex	0,160	0,142	-	0,698	0,069
	Kidd Lab set of 55 AISNPs	0,133	0,154	-	0,713	0,050
	PIMA + 34-plex	0,156	0,140	-	0,704	0,057
	***Seldin's* 125 AIMs**	0,150	0,141	-	0,709	0,046
	***Santos's* 192 AIMs**	0,147	0,143	-	0,710	0,045
	396 AIMs	0,146	0,142	-	0,712	0,040
	Painel Ancestralidade	0,149	0,145	-	0,707	0,051
	Painel Identificação Humana	0,172	0,157	-	0,672	0,121
	WGS	0,139	0,129	-	0,732	-
RPT	**PIMA**	0,194	0,066	0,011	0,729	0,032
	34-plex	0,189	0,072	0,034	0,705	0,038
	Kidd Lab set of 55 AISNPs	0,192	0,057	0,040	0,711	0,046
	PIMA + 34-plex	0,190	0,067	0,024	0,718	0,015
	***Seldin's* 125 AIMs**	0,190	0,072	0,026	0,713	0,023
	***Santos's* 192 AIMs**	0,190	0,065	0,028	0,717	0,017
	396 AIMs	0,202	0,067	0,026	0,705	0,039
	Painel Ancestralidade	0,192	0,063	0,042	0,703	0,050
	Painel Identificação Humana	0,211	0,071	0,013	0,706	0,063
	MEGA	0,183	0,067	0,026	0,724	-

É observado também que os painéis Ancestralidade e Identificação Humana apresentaram resultados mais acurados do que painéis com o dobro, ou mais, do número de marcadores em sua composição, como nas populações ACB, CLM, MXL e PEL. Ao se fazer a média das "Diferenças Absolutas" para cada painel nas 7 populações em que foram aplicados, obtêm-se os seguintes resultados: PIMA = 0,0810, SNP*for*ID 34-plex = 0,0504, *Kidd Lab set of 55 AISNPs* = 0,0553, PIMA + 34-plex = 0,0440, Seldin = 0,0494, Santos =

0,0579, Painel Ancestralidade = 0,0457, Painel Identificação Humana = 0,0787 e 396 AIMs = 0,0466, demonstrando que o painel PIMA + 34-plex apresenta o menor desvio médio das ancestralidades atribuídas pelas referências, seguido do Painel Ancestralidade, com pouco mais da metade do número de marcadores.

O software STRUCTURE atribui as ancestralidades individuais a clusters, os quais devem refletir majoritariamente um único grupo populacional de referência para que a inferência tenha qualidade estatística válida. STRUCTURE e outros softwares de inferência de ancestralidade usam do Equilíbrio de Hardy-Weinberg (EHW) para os seus cálculos, buscando clusters caracterizados pelo EHW bem como ausência de desequilíbrio de ligação (Padhukasahasram, 2014). Para se avaliar a eficiência dos diferentes painéis, a ancestralidade populacional para k = 4 foi utilizada como padrão, onde a média das ancestralidades atribuídas a cada indivíduo é obtida e atribuída à população a qual este indivíduo pertence. Para as populações miscigenadas analisadas no contexto do 1KGEN o conjunto de todos os 396 AIMs fora utilizado como gabarito para suas análises (**Tabela 34**). Dessa maneira, o método STRUCTURE demonstra a ancestralidade de uma população a partir da ancestralidade atribuída a cada indivíduo a ela pertencentes, ao invés de fazer uso das frequências alélicas de cada marcador.

Tabela 34. Diferença absoluta entre as inferências de ancestralidade populacional utilizando o software STRUCTURE e os parâmetros de referência (396 SNPs ou resultados de Marcorin, 2021).

Pop	Painel	AFR	AMR	EAS	EUR	Diferença Absoluta
ACB	**PIMA**	0,886	0,005	0,004	0,105	0,013
	34-plex	0,885	0,004	0,004	0,106	0,011
	Kidd Lab set of 55 AISNPs	0,876	0,004	0,009	0,111	0,013
	PIMA + 34-plex	0,889	0,004	0,003	0,105	0,018
	***Seldin's* 125 AIMs**	0,882	0,004	0,008	0,106	0,010
	***Santos's* 192 AIMs**	0,870	0,006	0,007	0,118	0,021
	Painel Ancestralidade	0,884	0,005	0,007	0,104	0,012
	Painel Identificação Humana	0,886	0,004	0,008	0,102	0,017
	396 SNPs	0,880	0,006	0,006	0,108	-
ASW	**PIMA**	0,729	0,047	0,010	0,214	0,082
	34-plex	0,743	0,042	0,010	0,206	0,054

Pop	Painel	AFR	AMR	EAS	EUR	Diferença Absoluta
	Kidd Lab set of 55 AISNPs	0,778	0,033	0,001	0,188	0,016
	PIMA + 34-plex	0,740	0,044	0,006	0,210	0,061
	***Seldin's* 125 AIMs**	0,767	0,024	0,021	0,188	0,038
	***Santos's* 192 AIMs**	0,788	0,038	0,008	0,167	0,051
	Painel Ancestralidade	0,774	0,034	0,002	0,191	0,007
	Painel Identificação Humana	0,746	0,074	0,003	0,177	0,079
	396 SNPs	0,770	0,035	0,002	0,192	-
	PIMA	0,058	0,217	0,068	0,658	0,128
	34-plex	0,094	0,237	0,009	0,660	0,038
	Kidd Lab set of 55 AISNPs	0,081	0,235	0,010	0,674	0,060
	PIMA + 34-plex	0,088	0,232	0,017	0,663	0,039
CLM	***Seldin's* 125 AIMs**	0,077	0,221	0,071	0,631	0,106
	***Santos's* 192 AIMs**	0,101	0,251	0,018	0,630	0,028
	Painel Ancestralidade	0,087	0,226	0,010	0,677	0,067
	Painel Identificação Humana	0,104	0,229	0,050	0,617	0,090
	396 SNPs	0,091	0,247	0,018	0,644	-
	PIMA	0,030	0,442	0,022	0,506	0,105
	34-plex	0,053	0,471	0,042	0,434	0,043
	Kidd Lab set of 55 AISNPs	0,042	0,481	0,025	0,452	0,057
	PIMA + 34-plex	0,052	0,460	0,031	0,456	0,020
MXL	***Seldin's* 125 AIMs**	0,043	0,403	0,126	0,428	0,173
	***Santos's* 192 AIMs**	0,057	0,450	0,040	0,454	0,006
	Painel Ancestralidade	0,050	0,439	0,039	0,473	0,038
	Painel Identificação Humana	0,054	0,444	0,078	0,424	0,077
	396 SNPs	0,054	0,453	0,039	0,454	-
	PIMA	0,027	0,666	0,037	0,271	0,130
	34-plex	0,031	0,726	0,021	0,222	0,043
	Kidd Lab set of 55 AISNPs	0,038	0,732	0,015	0,215	0,044
	PIMA + 34-plex	0,033	0,683	0,045	0,239	0,083
PEL	***Seldin's* 125 AIMs**	0,052	0,682	0,086	0,180	0,126
	***Santos's* 192 AIMs**	0,053	0,705	0,025	0,217	0,016
	Painel Ancestralidade	0,057	0,709	0,008	0,227	0,043
	Painel Identificação Humana	0,057	0,732	0,009	0,202	0,063
	396 SNPs	0,047	0,711	0,028	0,215	-
	PIMA	0,162	0,132	0,018	0,687	0,018
PUR	**34-plex**	0,163	0,101	0,071	0,665	0,126
	Kidd Lab set of 55 AISNPs	0,140	0,158	0,007	0,695	0,047
	PIMA + 34-plex	0,164	0,127	0,035	0,674	0,056

Pop	Painel	AFR	AMR	EAS	EUR	Diferença Absoluta
	***Seldin's* 125 AIMs**	0,151	0,121	0,033	0,695	0,031
	***Santos's* 192 AIMs**	0,157	0,139	0,018	0,686	0,017
	Painel Ancestralidade	0,159	0,131	0,010	0,701	0,023
	Painel Identificação Humana	0,175	0,153	0,024	0,648	0,094
	396 SNPs	0,153	0,134	0,018	0,695	-
RPT	**PIMA**	0,201	0,055	0,069	0,674	0,123
	34-plex	0,207	0,028	0,063	0,702	0,122
	Kidd Lab set of 55 AISNPs	0,197	0,047	0,041	0,715	0,058
	PIMA + 34-plex	0,198	0,045	0,055	0,702	0,088
	***Seldin's* 125 AIMs**	0,191	0,053	0,056	0,700	0,076
	***Santos's* 192 AIMs**	0,190	0,057	0,057	0,696	0,075
	396	0,211	0,062	0,036	0,692	0,076
	Painel Ancestralidade	0,199	0,055	0,042	0,703	0,065
	Painel Identificação Humana	0,215	0,047	0,061	0,677	0,134
	MEGA SNPs	0,183	0,067	0,026	0,724	-

Para a população ACB, não se verifica muita diferença entre as atribuições de ancestralidade populacional, independente de qual painel foi utilizado, com os painéis de Seldin e SNP*for*ID 34-plex apresentando diferenças de 1,0% e 1,1% da referência. Todas as outras populações apresentam maiores variações de valores para diferença absoluta da atribuição de ancestralidade. Para a população ASW, existe uma variação entre 0,7% de diferença com o Painel Ancestralidade até 8,2% com o painel PIMA. A população CLM apresenta variação entre 2,8% com painel de Santos e 12,8% com o painel PIMA. Na população MXL, de 0,6% para o painel desenvolvido por Santos até 17,3% com painel de Seldin. O painel de Seldin apresenta também uma diferença alta (12,6%) na população PEL, mas é superado pelo painel PIMA com 13,0%, o painel com menor porcentagem de diferença é o desenvolvido por Santos, com 1,6%. Para a população PUR, o painel de Santos demonstra-se novamente com a menor diferença da referência (1,7%), seguido do Painel Ancestralidade (2,3%). Nessa população, o painel SNP*for*ID 34-plex apresenta-se mais distante da referência, apresentando 12,6% de diferença.

Na população de Ribeirão Preto dois painéis apresentaram resultados muito próximos entre si e do resultado obtido por 336.112 SNPs, com diferenças de 7,5% e 7,6% respectivamente. Porém, apenas quando utilizado todos os 396 AIMs, ou os painéis *Kidd Lab set of 55 AISNPs* e Ancestralidade, a ancestralidade ameríndia foi apresentada como superior

à ancestralidade asiática. A **Tabela 35** traz os desvios do que seria a clusterização ideal das populações de referência (1,000).

Tabela 35. Desvio da clusterização ideal das populações de referência utilizando o software STRUCTURE. Os clusters são atribuídos como: 1- AFR, 2- AMR, 3- EAS e 4- EUR.

Painéis	Cluster				Desvio do Ideal
	1	2	3	4	
PIMA	0,9804	0,9359	0,9745	0,9377	0,1716
SNP*for*ID 34-plex	0,9895	0,9269	0,9609	0,9827	0,1400
Kidd Lab set of 55 AISNPs	0,9897	0,9385	0,9830	0,9888	0,0985
PIMA + 34-plex	0,9905	0,9445	0,9878	0,9833	0,0939
Seldin's 125 AIMs	0,9843	0,8822	0,9764	0,9800	0,1771
Santos's 192 AIMs	0,9821	0,9391	0,9585	0,9658	0,1545
396 AIMs	0,9875	0,9454	0,9823	0,9798	0,1049
Painel Ancestralidade	0,9862	0,9255	0,9802	0,9757	0,1324
Painel Identificação Humana	0,9848	0,9557	0,9721	0,9525	0,1349

O painel PIMA + 34-plex demonstra o resultado mais próximo do que seria a clusterização ideal, seguidos pelo *Kidd Lab set of 55 AISNPs* e os dois conjuntos desenvolvidos nesse trabalho. Em conjunto com as médias das diferenças absolutas observadas na **Tabela 34**, (PIMA = 0,0856, SNP*for*ID 34-plex = 0,0622, *Kidd Lab set of 55 AISNPs* = 0,0421, PIMA + 34-plex = 0,0521, Seldin = 0,0801, Santos = 0,0306, Painel Ancestralidade = 0,0364 e Painel Identificação Humana = 0,0790), observa-se que embora o painel desenvolvido por Santos apresente a menor diferença absoluta média, seu desvio da clusterização ideal é um dos maiores, sendo assim, os painéis *Kidd Lab set of 55 AISNPs* e PIMA + 34-plex demonstram-se como os melhores sistemas a serem utilizados em amostras de populações miscigenadas, porém, embora o Painel Ancestralidade tenha ficado em terceiro colocação, cabe reiterar a informação de que este painel apresenta apenas 33 marcadores, com resultados mais próximos do gabarito do que os apresentados pelos painéis de Seldin e Santos, que possuem, respectivamente, 92 e 159 SNPs a mais.

Com relação as atribuições de ancestralidade individual, o painel SNPforID 34-plex, quando aplicado à população de Ribeirão Preto, atribuiu um aumento na ancestralidade EAS para indivíduos que deveriam apresentar ancestralidade AMR, de acordo com a referência

(Marcorin, 2021). Para as populações miscigenadas, o painel de Seldin apresentou subestimação de ancestralidade AMR para populações com maior parcela de ancestralidade ameríndias, levando a uma superestimação da ancestralidade EAS, o que está relacionado com o menor poder de atribuição dos clusters AMR por esses painéis.

7.5 Parâmetros Forenses

O presente estudo reitera que os painéis de inferência de ancestralidade podem muitas vezes apresentar resultados importantes para o processo de identificação humana e estabelecimento de vínculos genéticos. No entanto, a informatividade forense de AIMs é altamente dependente da história demográfica que caracteriza cada população. Populações miscigenadas com grande porcentagem de ancestralidade provenientes de uma única população, como ACB e ASW, para a ancestralidade africana, apresentaram resultados menos informativos, devido a redução da variabilidade genética que os AIMs apresentam em populações autóctones; sendo assim, a busca pela utilização de SNPs informativos de ancestralidade para tais populações no âmbito forense é altamente prejudicada (Oliveira et al., 2009). A população PEL que apresenta predomínio de ancestralidade AMR apresentou uma situação semelhante, o que nos leva a entender que o fator miscigenação é de grande importância para que um painel de AIMs nos traga ambas as funções em uma amostra populacional.

A análise de parâmetros forenses para os sistemas CODIS (13 e 20 STRs) na mesma amostra populacional estudada neste projeto (Valle-Silva, 2019), além das estimativas obtidas para outros dois painéis de SNPs focados na identificação humana, o painel SNP*for*ID 52-plex (51 SNPs) (Sanchez et al., 2006) e o *II*SNP (92 SNPs) (Pakstis et al., 2010), serão utilizados para comparação (**Tabela 36**).

Tabela 36. Parâmetros forenses obtidos para os 343 indivíduos da população de Ribeirão Preto coletados de maneira randômica.

Painel	PMC	PEC	Referência
SNP*for*ID 52-plex (51 SNPs)	6,48E-21	0,9997	Valle-Silva, 2018
***II*SNPs (92 SNPs)**	4,91E-38	0,99999997	Valle-Silva, 2018
13 STR CODIS	3,95E-16	0,9993	Valle-Silva, 2018
20 STR CODIS	3,99E-25	0,99997	Valle-Silva, 2018
PIMA	6,25E-09	0,946751696	Presente Trabalho
SNP*for*ID 34-plex	1,19E-11	0,968195972	Presente Trabalho
Kidd Lab set of 55 AISNPs	4,29E-16	0,988328675	Presente Trabalho
PIMA + 34-plex	7,47E-20	0,998306489	Presente Trabalho
Seldin's 125 AIMs	2,23E-40	0,999997297	Presente Trabalho
Santos's 192 AIMs	1,53E-60	0,9999999778421	Presente Trabalho
396 AIMs	1,71E-124	>0,999999999999999	Presente Trabalho
Painel Ancestralidade	4,13E-10	0,933190463	Presente Trabalho
Painel Identificação Humana	6,18E-23	0,99988	Presente Trabalho

Os painéis Ancestralidade e SNP*for*ID 34-plex apresentaram um número muito pequeno de marcadores para que seus valores de informatividade fossem aptos à utilização forense, já o Painel Identificação Humana apresentou resultados importantes onde, dentro da população de Ribeirão Preto, sua informatividade superou o painel SNP*for*ID 52-plex e os 13 STR do CODIS, tanto para a probabilidade de *match* quanto para o poder de exclusão combinados, ficando próximo da informatividade demonstrada pelos 20 STRs do CODIS. Este mesmo painel também apresentou resultados expressivos para as populações PUR, CLM e MXL. Os Painéis *Kidd Lab set of 55 AISNPs* e PIMA + 34-plex apresentam mais marcadores que o Painel Identificação Humana, porém seus valores de probabilidade de *match* e de exclusão combinados foram menos informativos do que o obtido para os 50 AIMs. Para o *Kidd Lab set of 55 AISNPs*, observou-se valores equiparados aos 13 STR do CODIS, com reduzido valor de PEC. Para o painel PIMA + 34-plex, o valor de PMC foi mais informativo que os 13 STR do CODIS, mas também apresentou menor PEC. Como esperado, todos os painéis de ancestralidade com mais de 100 marcadores apresentaram resultados altamente informativos tanto para PMC, quanto para PEC, onde se tem o conhecimento de que são necessários de 3 a 5 AIMs para se obtermos a informatividade de apenas um STR (Oliveira et al., 2009).

Análises através de eletroforese capilar associada aos tamanhos dos STRs são considerados o “Padrão Ouro” em casos de genética forense há décadas, devido à alta informatividade desses marcadores, porém problemas como elevada taxa de mutação, *stutters* e o grande tamanho associado aos *amplicons*, levam a dificuldades de análise em amostras degradadas; tais problemas podem ser contornadas pelos SNPs (Avila et al., 2019). Sendo assim, os resultados aqui apresentados sugerem que os SNPs informativos de ancestralidade podem ser utilizados em casos forenses, tanto para a identificação humana, onde a análise de amostras degradadas são mais facilmente realizadas com a redução do tamanho do *amplicon*, quanto para a inferência da ancestralidade genética, o que pode contribuir para reduzir o número de suspeitos a serem investigados (Avila et al., 2019; Kidd et al., 2014; Oliveira et al., 2009; Pakstis et al., 2010).

Conclusões

8. Conclusões

Este trabalho foi o primeiro a analisar simultaneamente os painéis PIMA (25 AIMs), SNP*for*ID 34-plex, *Seldin's 128* AIMs (125 AIMs), *Kidd Lab set of 55 AISNPs* e *Santos's 192 AIMs* em uma amostra de população brasileira e nas populações miscigenadas do 1KGEN, para utilização em inferências de ancestralidade e para obtenção de parâmetros forenses.

A partir das análises de ancestralidade populacional e individual, a amostra brasileira apresentou-se como uma população tetra-híbrida, caracterizada pela miscigenação entre as populações europeias, africanas, ameríndias e asiáticas. As demais populações latino-americanas mostraram-se melhor avaliadas quando analisadas usando modelos tri-híbridos, devido a não apresentarem grande contribuição proveniente de populações asiáticas. A combinação dos painéis PIMA e SNP*for*ID 34-plex, mesmo apresentando 59 AIMs, apresentaram resultados muito próximos aos painéis de Santos (192 AIMs) e Seldin (125 AIMs) em todas as populações analisadas, demonstrando que a seleção dos marcadores informativos de ancestralidade é mais importante do que a quantidade utilizada para o ensaio.

O Painel Ancestralidade, com 33 marcadores informativos de ancestralidade, mostrou resultados muito próximos dos obtidos através da análise GWAS para as populações miscigenadas estudadas no âmbito do 1KGEN, tanto para análises de inferências de ancestralidade populacionais, quanto para análises de ancestralidade individual. Para a população de Ribeirão Preto seus resultados para ancestralidade populacional foram muito próximos da referência com seus valores superados apenas pelo PIMA + 34-plex. Sua utilização na América Latina demonstra-se viável, mesmo com a menor quantidade de marcadores.

Os painéis *Kidd Lab set of 55 AISNPs* e PIMA + 34-plex apresentaram resultados muito próximos dos obtidos pelos 13 marcadores originais do CODIS para obtenção de parâmetros forenses na amostra brasileira e para as populações miscigenadas avaliadas pelo 1KGEN, com exceção das populações ACB, ASW e PEL, devido à grande proporção de ancestralidade proveniente de uma única população de referência, o que reduz a informatividade forense dos AIMs. Já os painéis de Santos (192 AIMs) e Seldin (125 AIMs) se mostraram mais informativos do que os 20 marcadores atuais do CODIS, principalmente devido ao número de marcadores.

Por fim, o Painel Identificação Humana, contendo 50 AIMs, apresentou valores mais informativos do que os observados para os 13 STRs originais do sistema CODIS e para o SNP*for*ID 52-plex quando aplicados na população de Ribeirão Preto. A informatividade quase atingiu os níveis observados para os 20 atuais CODIS. Tal painel também apresentou valores informativos para as populações CLM, MXL, PEL e PUR. Sendo assim, espera-se que este painel apresente valores informativos de parâmetros forenses em outras populações que tenham se originado a partir da miscigenação entre as parentais AFR, AMR, EAS e EUR em outras regiões geográficas.

Referências

9. Referências

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*. https://doi.org/10.1186/1471-2105-12-246

Álvarez-Iglesias, V., Salas, A., Cerezo, M., Ramos-Luis, E., Jaime, J. C., Lareu, M. V., & Carracedo, A. (2006). Genotyping coding region mtDNA SNPs for Asian and Native American haplogroup assignation. *International Congress Series*, *1288*, 4–6. https://doi.org/10.1016/j.ics.2005.11.036

Ambers, A., Bus, M. M., King, J. L., Jones, B., Durst, J., Bruseth, J. E., Gill-King, H., & Budowle, B. (2020). Forensic genetic investigation of human skeletal remains recovered from the La Belle shipwreck. *Forensic Science International*, *306*, 110050. https://doi.org/10.1016/j.forsciint.2019.110050

Amigo, J., Phillips, C., Lareu, M., & Carracedo, Á. (2008). The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. *International Journal of Legal Medicine*, *122*(5), 435–440. https://doi.org/10.1007/s00414-008-0233-7

Andrade, R. B., Amador, M. A. T., Cavalcante, G. C., Leitão, L. P. C., Fernandes, M. R., Modesto, A. A. C., Moreira, F. C., Khayat, A. S., Assumpção, P. P., Ribeiro-dos-Santos, Â., Santos, S., & Santos, N. P. C. (2018). Estimating Asian Contribution to the Brazilian Population: A New Application of a Validated Set of 61 Ancestry Informative Markers. *G3 Genes|Genomes|Genetics*, *8*(11), 3577–3582. https://doi.org/10.1534/g3.118.200650

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502–508. https://doi.org/10.1111/opo.12131

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Avila, E., Felkl, A. B., Graebin, P., Nunes, C. P., & Alho, C. S. (2019). Forensic characterization of Brazilian regional populations through massive parallel sequencing of 124 SNPs included in HID ion Ampliseq Identity Panel. *Forensic Science International: Genetics*, *40*, 74–84. https://doi.org/10.1016/j.fsigen.2019.02.012

Barnholtz-Sloan, J. S., Pfaff, C. L., Chakraborty, R., & Long, J. C. (2005). Informativeness of the CODIS STR Loci for Admixture Analysis. *Journal of Forensic Sciences*, *50*(6), 1–5. https://doi.org/10.1520/JFS2003404

Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., Blanché, H., Deleuze, J. F., Cann, H., Mallick, S., Reich, D., Sandhu, M. S., Skoglund, P., Scally, A., Xue, Y., … Tyler-Smith, C. (2020).

Insights into human genetic variation and population history from 929 diverse genomes. *Science*, *367*(6484). https://doi.org/10.1126/science.aay5012

Brookes, A. J. (1999). The essence of SNPs. *Gene*, *234*(2), 177–186. https://doi.org/10.1016/S0378-1119(99)00219-X

Butler, J. M., Coble, M. D., & Vallone, P. M. (2007). STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine, and Pathology*, *3*(3), 200–205. https://doi.org/10.1007/s12024-007-0018-1

Butler, J. M., & Hill, C. R. (2012). Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis. *Forensic Science Review*, *24*(1), 15–26.

Cardena, M. M. S. G., Ribeiro-dos-Santos, Â., Santos, S., Mansur, A. J., Pereira, A. C., & Fridman, C. (2014). Amerindian genetic ancestry is associated with higher survival rates compared to African and European ancestry in Brazilian patients with heart failure. *International Journal of Cardiology*, *176*(2), 527–528. https://doi.org/10.1016/j.ijcard.2014.07.039

Carvalho Gontijo, C., Porras-Hurtado, L. G., Freire-Aradas, A., Fondevila, M., Santos, C., Salas, A., Henao, J., Isaza, C., Beltrán, L., Nogueira Silbiger, V., Castillo, A., Ibarra, A., Moreno Chavez, F., Söchtig, J., Ruiz, Y., Barreto, G., Rondon, F., Zabala, W., Borjas, L., … Phillips, C. (2020). PIMA: A population informative multiplex for the Americas. *Forensic Science International: Genetics*, *44*, 102200. https://doi.org/10.1016/j.fsigen.2019.102200

Chakraborty, R. (1986). Gene admixture in human populations: Models and predictions. *American Journal of Physical Anthropology*, *29*(S7), 1–43. https://doi.org/10.1002/ajpa.1330290502

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1). https://doi.org/10.1186/s13742-015-0047-8

Cheung, E. Y. Y., Phillips, C., Eduardoff, M., Lareu, M. V., & McNevin, D. (2019). Performance of ancestry-informative SNP and microhaplotype markers. *Forensic Science International: Genetics*, *43*, 102141. https://doi.org/10.1016/j.fsigen.2019.102141

Chua, J., Koh, J. Y., Tan, A. G., Zhao, W., Lamoureux, E., Mitchell, P., Wang, J. J., Wong, T. Y., & Cheng, C.-Y. (2015). Ancestry, Socioeconomic Status, and Age-Related Cataract in Asians. *Ophthalmology*, *122*(11), 2169–2178. https://doi.org/10.1016/j.ophtha.2015.06.052

Craig, J. D., & Faria, A. B. (2021). Immigrant nationality and human capital formation in Brazil. *International Journal of Educational Development*, *80*. https://doi.org/10.1016/j.ijedudev.2020.102260

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., … Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, *48*(10), 1284–1287. https://doi.org/10.1038/ng.3656

Durães, R. O., Berardinelli, G. N., da Costa, A. M., Scapulatempo-Neto, C., Pereira, R., Oliveira, M. A., Guimarães, D. P., & Reis, R. M. (2020). Role of Genetic Ancestry in 1,002 Brazilian Colorectal Cancer Patients From Barretos Cancer Hospital. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.00145

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Ferreira, L. B., Mendes-Junior, C. T., Wiezel, C. E. V., Luizon, M. R., & Simões, A. L. (2006). Genomic ancestry of a sample population from the State of São Paulo, Brazil. *American Journal of Human Biology*, *18*(5), 702–705. https://doi.org/10.1002/ajhb.20474

Gettings, K. B., Kiesler, K. M., & Vallone, P. M. (2015). Performance of a next generation sequencing SNP assay on degraded DNA. *Forensic Science International: Genetics*, *19*, 1–9. https://doi.org/10.1016/j.fsigen.2015.04.010

Gomes, M. B., Gabrielli, A. B., Santos, D. C., Pizarro, M. H., Barros, B. S. V., Negrato, C. A., Dib, S. A., Porto, L. C., & Silva, D. A. (2018). Self-reported color-race and genomic ancestry in an admixed population: A contribution of a nationwide survey in patients with type 1 diabetes in Brazil. *Diabetes Research and Clinical Practice*, *140*, 245–252. https://doi.org/10.1016/j.diabres.2018.03.021

Goodwin, W. (2015). DNA profiling: The first 30years. *Science & Justice*, *55*(6), 375–376. https://doi.org/10.1016/j.scijus.2015.11.001

Gouy, A., & Zieger, M. (2017). STRAF—A convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International: Genetics*, *30*, 148–151. https://doi.org/10.1016/j.fsigen.2017.07.007

Grover, A., & Sharma, P. C. (2016). Development and use of molecular markers: past and present. *Critical Reviews in Biotechnology*, *36*(2), 290–302. https://doi.org/10.3109/07388551.2014.959891

Hagelberg, E., Gray, I. C., & Jeffreys, A. J. (1991). Identification of the skeletal remains of a murder victim by DNA analysis. *Nature*, *352*(6334), 427–429. https://doi.org/10.1038/352427a0

Hardy, D. S., Garvin, J. T., Mersha, T. B., & Racette, S. B. (2020). Ancestry specific associations of FTO gene variant and metabolic syndrome. *Medicine*, *99*(6), e18820. https://doi.org/10.1097/MD.0000000000018820

Howie, B., Marchini, J., & Stephens, M. (2011). Genotype Imputation with Thousands of Genomes. *G3 Genes|Genomes|Genetics*, *1*(6), 457–470. https://doi.org/10.1534/g3.111.001198

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genetics*, *5*(6), e1000529. https://doi.org/10.1371/journal.pgen.1000529

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jeffreys, A. J., Allen, M. J., Hagelberg, E., & Sonnberg, A. (1992). Identification of the skeletal remains of josef mengele by DNA analysis. *Forensic Science International*, *56*(1), 65–76. https://doi.org/10.1016/0379-0738(92)90148-P

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Individual-specific 'fingerprints' of human DNA. *Nature*, *316*(6023), 76–79. https://doi.org/10.1038/316076a0

Johnson, G. W., Ehrlich, R., Full, W., & Ramos, S. (2007). PRINCIPAL COMPONENTS ANALYSIS AND RECEPTOR MODELS IN ENVIRONMENTAL FORENSICS. In *Introduction to Environmental Forensics* (pp. 207–272). Elsevier. https://doi.org/10.1016/B978-012369522-2/50008-7

Kehdy, F. S. G., Gouveia, M. H., Machado, M., Magalhães, W. C. S., Horimoto, A. R., Horta, B. L., Moreira, R. G., Leal, T. P., Scliar, M. O., Soares-Souza, G. B., Rodrigues-Soares, F., Araújo, G. S., Zamudio, R., Sant Anna, H. P., Santos, H. C., Duarte, N. E., Fiaccone, R. L., Figueiredo, C. A., Silva, T. M., … Zolini, C. (2015). Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the National Academy of Sciences*, *112*(28), 8696–8701. https://doi.org/10.1073/pnas.1504447112

Kidd, K. K., Speed, W. C., Pakstis, A. J., Furtado, M. R., Fang, R., Madbouly, A., Maiers, M., Middha, M., Friedlaender, F. R., & Kidd, J. R. (2014). Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Science International: Genetics*, *10*(1), 23–32. https://doi.org/10.1016/j.fsigen.2014.01.002

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). <scp>Clumpak</scp> : a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kosoy, R., Nassir, R., Tian, C., White, P. A., Butler, L. M., Silva, G., Kittles, R., Alarcon-Riquelme, M. E., Gregersen, P. K., Belmont, J. W., de La Vega, F. M., & Seldin, M. F. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation*, *30*(1), 69–78. https://doi.org/10.1002/humu.20822

Ku, C. S., Loy, E. Y., Salim, A., Pawitan, Y., & Chia, K. S. (2010). The discovery of human genetic variations and their use as disease markers: past, present and future. *Journal of Human Genetics*, *55*(7), 403–415. https://doi.org/10.1038/jhg.2010.55

Kyle, R. A., & Shampo, M. A. (2001). Karl Landsteiner--discoverer of the major human blood groups. *Mayo Clinic Proceedings. Mayo Clinic*, *76*(8), 830. https://doi.org/10.1016/S0025-6196(11)63228-X

Lachance, J. (2016). Hardy–Weinberg Equilibrium and Random Mating. In *Encyclopedia of Evolutionary Biology* (pp. 208–211). Elsevier. https://doi.org/10.1016/B978-0-12-800049-6.00022-6

Lanza, A. L., & Lamounier, M. L. (2015). Café, imigrantes e empresas no nordeste de São Paulo (Ribeirão Preto, 1880-1930). *História Econômica & História de Empresas*, *17*(2). https://doi.org/10.29182/hehe.v17i2.239

Latham, K. E., & Miller, J. J. (2019). DNA recovery and analysis from skeletal material in modern forensic contexts. *Forensic Sciences Research*, *4*(1), 51–59. https://doi.org/10.1080/20961790.2018.1515594

Lee, M. H., Koh, J. W. H., Ng, C. H., Lim, S. H. H., Cho, J., Lateef, A., Mak, A., & Tay, S. H. (2022). A meta-analysis of clinical manifestations in asian systemic lupus erythematous: The effects of ancestry, ethnicity and gender. *Seminars in Arthritis and Rheumatism*, *52*, 151932. https://doi.org/10.1016/j.semarthrit.2021.12.004

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, J., Zarzar, T. G., White, J. D., Indencleef, K., Hoskens, H., Matthews, H., Nauwelaers, N., Zaidi, A., Eller, R. J., Herrick, N., Günther, T., Svensson, E. M., Jakobsson, M., Walsh, S., van Steen, K., Shriver, M. D., & Claes, P. (2020). Robust genome-wide ancestry inference for heterogeneous datasets: illustrated using the 1,000 genome project with 3D facial images. *Scientific Reports*, *10*(1), 11850. https://doi.org/10.1038/s41598-020-68259-w

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., … Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206. https://doi.org/10.1038/nature18964

Marca-Ysabel, M. V., Rajabli, F., Cornejo-Olivas, M., Whitehead, P. G., Hofmann, N. K., Illanes Manrique, M. Z., Veliz Otani, D. M., Milla Neyra, A. K., Castro Suarez, S., Meza Vega, M., Adams, L. D., Mena, P. R., Rosario, I., Cuccaro, M. L., Vance, J. M., Beecham, G. W., Custodio, N., Montesinos, R., Mazzetti Soler, P. E., & Pericak-Vance, M. A. (2021). Dissecting the role of Amerindian genetic ancestry and the ApoE ε4 allele on Alzheimer disease in an admixed Peruvian population. *Neurobiology of Aging*, *101*, 298.e11-298.e15. https://doi.org/10.1016/j.neurobiolaging.2020.10.003

Marcorin, L. (2021). *Influence of ancestry and genetic variability of the MITF transcription factor and its targets TYR, TYRP1 and DCT on the development of vitiligo in a Brazilian population sample* [Universidade de São Paulo]. https://doi.org/10.11606/T.17.2021.tde-06122021-122111

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., Daly, M. J., Bustamante, C. D., & Kenny, E. E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics*, *100*(4), 635–649. https://doi.org/10.1016/j.ajhg.2017.03.004

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*(3), 1215–1215. https://doi.org/10.1093/nar/16.3.1215

Naslavsky, M. S., Scliar, M. O., Nunes, K., Wang, J. Y. T., Yamamoto, G. L., Guio, H., Tarazona-Santos, E., Duarte, Y. A. O., Passos-Bueno, M. R., Meyer, D., & Zatz, M. (2021). Biased pathogenic assertions of loss of function variants challenge molecular diagnosis of admixed individuals. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics*, *187*(3), 357–363. https://doi.org/10.1002/ajmg.c.31931

Oliveira, M. L. G. de, Veiga-Castelli, L. C., Marcorin, L., Debortoli, G., Pereira, A. L. E., Fracasso, N. C. de A., Silva, G. do V., Souza, A. S., Massaro, J. D., Simões, A. L., Sabbagh, A., Donadi, E. A., Castelli, E. C., & Mendes-Junior, C. T. (2018). Extended HLA-G genetic diversity and ancestry composition in a Brazilian admixed population sample: Implications for HLA-G transcriptional control and for case-control association studies. *Human Immunology*, *79*(11), 790–799. https://doi.org/10.1016/j.humimm.2018.08.005

Oliveira, S. F., Godinho, N. M., Amorim, C. E. G., Gontijo, C. C., Falcão-Alencar, G., Barcelos, R. S. S., Luizon, M. R., Klautau-Guimarães, M. N., Simões, A. L., & Mendes-Junior, C. T. (2009). STRs and AIMs informativeness for forensic purposes in an admixed Brazilian population. *Forensic Science International: Genetics Supplement Series*, *2*(1), 475–476. https://doi.org/10.1016/j.fsigss.2009.10.003

Padhukasahasram, B. (2014). Inferring ancestry from population genomic data and its applications. *Frontiers in Genetics*, *5*. https://doi.org/10.3389/fgene.2014.00204

Páez-Riberos, L. A., Muñoz-Valle, J. F., Figuera, L. E., Nuño-Arana, I., Sandoval-Ramírez, L., González-Martín, A., Ibarra, B., & Rangel-Villalobos, H. (2006). Y-linked haplotypes in Amerindian chromosomes from Mexican populations: Genetic evidence to the dual origin of the Huichol tribe. *Legal Medicine*, *8*(4), 220–225. https://doi.org/10.1016/j.legalmed.2006.02.003

Pakstis, A. J., Gurkan, C., Dogan, M., Balkaya, H. E., Dogan, S., Neophytou, P. I., Cherni, L., Boussetta, S., Khodjet-El-Khil, H., ben Ammar ElGaaied, A., Salvo, N. M., Janssen, K., Olsen, G. H., Hadi, S., Almohammed, E. K., Pereira, V., Truelsen, D. M., Bulbul, O., Soundararajan, U., … Kidd, K. K. (2019). Genetic relationships of European, Mediterranean, and SW Asian populations using a panel of 55 AISNPs. *European Journal of Human Genetics*, *27*(12), 1885–1893. https://doi.org/10.1038/s41431-019-0466-6

Pakstis, A. J., Haigh, E., Cherni, L., ElGaaied, A. B. A., Barton, A., Evsanaa, B., Togtokh, A., Brissenden, J., Roscoe, J., Bulbul, O., Filoglu, G., Gurkan, C., Meiklejohn, K. A., Robertson, J. M., Li, C.-X., Wei, Y.-L., Li, H., Soundararajan, U., Rajeevan, H., … Kidd, K. K. (2015). 52 additional reference population samples for the 55 AISNP panel. *Forensic Science International: Genetics*, *19*, 269–271. https://doi.org/10.1016/j.fsigen.2015.08.003

Pakstis, A. J., Kang, L., Liu, L., Zhang, Z., Jin, T., Grigorenko, E. L., Wendt, F. R., Budowle, B., Hadi, S., al Qahtani, M. S., Morling, N., Mogensen, H. S., Themudo, G. E., Soundararajan, U., Rajeevan, H., Kidd, J. R., & Kidd, K. K. (2017). Increasing the reference populations for the 55 AISNP panel: the need and benefits. *International Journal of Legal Medicine*, *131*(4), 913–917. https://doi.org/10.1007/s00414-016-1524-z

Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics*, *127*(3), 315–324. https://doi.org/10.1007/s00439-009-0771-1

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, *28*(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pereira, V., Mogensen, H. S., Børsting, C., & Morling, N. (2017). Evaluation of the Precision ID Ancestry Panel for crime case work: A SNP typing assay developed for typing of 165 ancestral informative markers. *Forensic Science International: Genetics*, *28*, 138–145. https://doi.org/10.1016/j.fsigen.2017.02.013

Pereira, V., Santangelo, R., Børsting, C., Tvedebrink, T., Almeida, A. P. F., Carvalho, E. F., Morling, N., & Gusmão, L. (2020). Evaluation of the Precision of Ancestry Inferences in South American Admixed Populations. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00966

Pérez-Mayoral, J., Soto-Salgado, M., Shah, E., Kittles, R., Stern, M. C., Olivera, M. I., Gonzalez-Pons, M., Rodriguez-Quilichinni, S., Torres, M., Reyes, J. S., Tous, L., López, N., Chevere, V. C., & Cruz-Correa, M. (2019). Association of genetic ancestry with colorectal tumor location in Puerto Rican Latinos. *Human Genomics*, *13*(1), 12. https://doi.org/10.1186/s40246-019-0196-4

Pfaffelhuber, P., Grundner-Culemann, F., Lipphardt, V., & Baumdicker, F. (2020). How to choose sets of ancestry informative markers: A supervised feature selection approach. *Forensic Science International: Genetics*, *46*. https://doi.org/10.1016/j.fsigen.2020.102259

Phillips, C., Fondevila, M., García-Magariños, M., Rodriguez, A., Salas, A., Carracedo, Á., & Lareu, M. V. (2008). Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International: Genetics*, *2*(3), 198–204. https://doi.org/10.1016/j.fsigen.2008.02.002

Phillips, C., McNevin, D., Kidd, K. K., Lagacé, R., Wootton, S., de la Puente, M., Freire-Aradas, A., Mosquera-Miguel, A., Eduardoff, M., Gross, T., Dagostino, L., Power, D., Olson, S., Hashiyada, M., Oz, C., Parson, W., Schneider, P. M., Lareu, M. V., & Daniel, R. (2019). MAPlex - A massively parallel sequencing ancestry analysis multiplex for Asia-Pacific populations. *Forensic Science International: Genetics*, *42*, 213–226. https://doi.org/10.1016/j.fsigen.2019.06.022

Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, A., Álvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., Lareu, M. v., & Carracedo, Á. (2007). Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics*, *1*(3–4), 273–280. https://doi.org/10.1016/j.fsigen.2007.06.008

Pontes, M. L., Fondevila, M., Laréu, M. V., & Medeiros, R. (2015). SNP Markers as Additional Information to Resolve Complex Kinship Cases. *Transfusion Medicine and Hemotherapy*, *42*(6), 385–388. https://doi.org/10.1159/000440832

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. v. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, *4*. https://doi.org/10.3389/fgene.2013.00098

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Ribeiro-dos-Santos, A. M., Vidal, A. F., Vinasco-Sandoval, T., Guerreiro, J., Santos, S., Ribeiro-dos-Santos, Â., & de Souza, S. J. (2020). Exome Sequencing of Native Populations From the Amazon Reveals Patterns on the Peopling of South America. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.548507

Rosenberg, N. A. (2003). distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Salam, M. T., Avoundjian, T., Knight, W. M., & Gilliland, F. D. (2015). Genetic Ancestry and Asthma and Rhinitis Occurrence in Hispanic Children: Findings from the Southern California Children's Health Study. *PLOS ONE*, *10*(8), e0135384. https://doi.org/10.1371/journal.pone.0135384

Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., & Morling, N. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *ELECTROPHORESIS*, *27*(9), 1713–1724. https://doi.org/10.1002/elps.200500671

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Santos, H. C., Horimoto, A. V. R., Tarazona-Santos, E., Rodrigues-Soares, F., Barreto, M. L., Horta, B. L., Lima-Costa, M. F., Gouveia, M. H., Machado, M., Silva, T. M., Sanches, J. M., Esteban, N., Magalhaes, W. C. S., Rodrigues, M. R., Kehdy, F. S. G., & Pereira, A. C. (2016). A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: The Brazilian set. *European Journal of Human Genetics*, *24*(5), 725–731. https://doi.org/10.1038/ejhg.2015.187

Say, Y.-H. (2017). The association of insertions/deletions (INDELs) and variable number tandem repeats (VNTRs) with obesity and its related traits and complications. *Journal of Physiological Anthropology*, *36*(1), 25. https://doi.org/10.1186/s40101-017-0142-x

Serre, D., & Pääbo, S. (2004). Evidence for Gradients of Human Genetic Diversity Within and Among Continents. *Genome Research*, *14*(9), 1679–1685. https://doi.org/10.1101/gr.2529604

Silva, G. do V. (2019). *Análise de marcadores forenses (STRs e SNPs) rotineiramente empregados na identificação humana utilizando sequenciamento de nova geração* [Universidade de São Paulo]. https://doi.org/10.11606/D.59.2019.tde-23112018-095204

Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J., Huang, W., Shen, D., Lu, D., Luo, J., Chu, J., Tan, J., Shen, P., Davis, R., Cavalli-Sforza, L., Chakraborty, R., Xiong, M., Du, R., Oefner, P., … Jin, L. (1999). Y-Chromosome Evidence for a Northward Migration of Modern Humans into Eastern Asia during the Last Ice Age. *The American Journal of Human Genetics*, *65*(6), 1718–1724. https://doi.org/10.1086/302680

The 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–1073. https://doi.org/10.1038/nature09534

Tozzo, P., Delicati, A., Frigo, A. C., & Caenazzo, L. (2021). Comparison of the Allelic Alterations between InDel and STR Markers in Tumoral Tissues Used for Forensic Purposes. *Medicina*, *57*(3), 226. https://doi.org/10.3390/medicina57030226

Tvedebrink, T., & Eriksen, P. S. (2019). Inference of admixed ancestry with Ancestry Informative Markers. *Forensic Science International: Genetics*, *42*, 147–153. https://doi.org/10.1016/j.fsigen.2019.06.013

Valle-Silva, G. do, Souza, F. D. N. de, Marcorin, L., Pereira, A. L. E., Carratto, T. M. T., Debortoli, G., Oliveira, M. L. G. de, Fracasso, N. C. de A., Andrade, E. S. de, Donadi, E. A., Norton, H. L., Parra, E. J., Simões, A. L., Castelli, E. C., & Mendes-Junior, C. T. (2019). Applicability of the SNPforID 52-plex panel for human identification and ancestry evaluation in a Brazilian population sample by next-generation sequencing. *Forensic Science International: Genetics*, *40*, 201–209. https://doi.org/10.1016/j.fsigen.2019.03.003

Wood, M. R., Sturk-Andreaggi, K., Ring, J. D., Huber, N., Bodner, M., Crawford, M. H., Parson, W., & Marshall, C. (2019). Resolving mitochondrial haplogroups B2 and B4 with next-generation mitogenome sequencing to distinguish Native American from Asian haplotypes. *Forensic Science International: Genetics*, *43*, 102143. https://doi.org/10.1016/j.fsigen.2019.102143

Zhang, Q., Zhou, Z., Wang, L., Quan, C., Liu, Q., Tang, Z., Liu, L., Liu, Y., & Wang, S. (2020). Pairwise kinship testing with a combination of STR and SNP loci. *Forensic Science International: Genetics*, *46*, 102265. https://doi.org/10.1016/j.fsigen.2020.102265

Apêndices

Apêndices

Apêndice A

Tabela A1. Identificação dos 399 marcadores analisados. Probabilidade de *match* (PM), poder de discriminação (PD), poder de exclusão (PE), índice de paternidade típico, heterozigoses observadas e esperadas (O(HET) e E(HET)), e P-*value* obtido para o HWE para os 513 indivíduos de Ribeirão Preto. Os valores estatisticamente significantes ao nível de 5% estão indicados em itálico, e aquele que permaneceu significante após a aplicação da correção de Bonferroni para múltiplos testes ($\alpha_{BONFERRONI} = 0.05/396 = 0.000126$) se encontra em negrito.

Cromossomo	POS	ID	Conjunto	REF	ALT	PM	PD	PE	IPT	O(HET)	E(HET)	P-*value*
1	1259424	rs11804831	SANTOS	T	C	0,3654	0,6346	0,1149	0,8355	0,4016	0,4749	*0,000562*
1	3372960	rs12135987	SANTOS	G	A	0,4011	0,5989	0,1187	0,8442	0,4077	0,4405	0,102000
1	3953894	rs12119556	SANTOS	G	A	0,6301	0,3699	0,0389	0,6506	0,2315	0,2188	0,277300
1	6490316	rs2986742	SELDIN	T	C	0,4736	0,5264	0,0743	0,7405	0,3248	0,3598	*0,035300*
1	12548144	rs6541030	SELDIN	G	A	0,6420	0,3580	0,0316	0,6302	0,2066	0,2155	0,311400
1	17844391	rs647325	SELDIN	A	G	0,3905	0,6095	0,1392	0,8906	0,4386	0,4595	0,292400
1	27605187	rs4908343	SELDIN	G	A	0,3763	0,6237	0,1328	0,8763	0,4294	0,4709	*0,048380*
1	35180524	rs10493066	SANTOS	C	T	0,4889	0,5111	0,0652	0,7185	0,3041	0,3457	*0,007311*
1	36302599	rs1573020	34PLEX	G	A	0,7184	0,2816	0,0177	0,5885	0,1504	0,1651	0,055360
1	41894599	rs1325502	SELDIN	G	A	0,4472	0,5528	0,0948	0,7892	0,3665	0,3878	0,211400
1	44021608	rs6429644	SANTOS	T	C	0,5003	0,4997	0,0636	0,7145	0,3002	0,3347	*0,024240*
1	50472176	rs3827730	SANTOS	T	C	0,4342	0,5658	0,1045	0,8117	0,3840	0,4030	0,275300
1	55197699	rs12130799	SELDIN	A	G	0,7600	0,2400	0,0141	0,5764	0,1326	0,1338	0,739800
1	68384004	rs3118378	SELDIN	A	G	0,4035	0,5965	0,1909	1,0079	0,5039	0,4734	0,161900
1	101063398	rs12402499	PIMA	G	A	0,7178	0,2822	0,0193	0,5937	0,1578	0,1618	0,578500
1	101244007	rs3737576	SELDIN|KIDD	T	C	0,7384	0,2616	0,0152	0,5803	0,1384	0,1520	0,070410

1	151150013	rs7554936	SELDIN|KIDD	C	T	0,3599	0,6401	0,1564	0,9293	0,4620	0,4983	0,110500
1	159204893	rs2814778	KIDD|34Plex	T	C	0,5305	0,4695	0,0474	0,6732	0,2573	0,3115	*0,000181*
1	168190652	rs1040404	SELDIN	G	A	0,3654	0,6346	0,1702	0,9607	0,4795	0,5000	0,377100
1	185716412	rs10127540	PIMA	T	C	0,3603	0,6397	0,1365	0,8845	0,4347	0,4884	*0,014530*
1	186179900	rs1407434	SELDIN	G	A	0,5546	0,4454	0,0530	0,6877	0,2729	0,2842	0,352800
1	204821849	rs2065160	34PLEX	A	G	0,4817	0,5183	0,0867	0,7703	0,3509	0,3521	0,900700
1	208169341	rs4844413	SANTOS	C	T	0,5011	0,4989	0,0788	0,7515	0,3346	0,3322	1,000000
1	212613541	rs4951629	SELDIN	T	C	0,6325	0,3675	0,0416	0,6577	0,2398	0,2140	*0,003215*
1	234566114	rs726570	SANTOS	T	C	0,4169	0,5831	0,1163	0,8389	0,4040	0,4239	0,294400
1	234580760	rs955613	SANTOS	T	C	0,4096	0,5904	0,1023	0,8066	0,3801	0,4270	*0,013190*
1	242179202	rs316873	SELDIN	C	T	0,6405	0,3595	0,0327	0,6333	0,2105	0,2155	0,542300
2	7009024	rs896788	34PLEX	C	T	0,4866	0,5134	0,0764	0,7456	0,3294	0,3468	0,253000
2	7828144	rs798443	SELDIN|KIDD	G	A	0,3713	0,6287	0,1325	0,8754	0,4288	0,4757	*0,026050*
2	14616225	rs7421394	SELDIN	A	G	0,3606	0,6394	0,1378	0,8875	0,4366	0,4887	*0,018670*
2	16634624	rs11096686	SANTOS	T	C	0,4086	0,5914	0,0980	0,7966	0,3723	0,4270	*0,003840*
2	17181301	rs1876482	KIDD	G	A	0,6887	0,3113	0,0211	0,5993	0,1657	0,1868	*0,016030*
2	17720218	rs1834619	KIDD|PIMA	G	A	0,6310	0,3690	0,0321	0,6318	0,2086	0,2257	0,113400
2	29315545	rs4666200	SELDIN	G	A	0,3844	0,6156	0,1415	0,8958	0,4418	0,4666	0,253100
2	37714253	rs4670767	SELDIN	G	T	0,5610	0,4390	0,0559	0,6951	0,2807	0,2765	0,873100
2	45993749	rs2711302	SANTOS	A	G	0,5524	0,4476	0,0484	0,6759	0,2603	0,2889	*0,031000*
2	72317001	rs10173381	SANTOS	A	G	0,3960	0,6040	0,1384	0,8889	0,4375	0,4536	0,435900
2	76867664	rs721390	SANTOS	G	T	0,4378	0,5622	0,0927	0,7844	0,3626	0,3969	0,058040
2	79637797	rs13400937	SELDIN	G	T							
2	82283778	rs12714093	SANTOS	G	A	0,3809	0,6191	0,1580	0,9331	0,4641	0,4781	0,514200
2	96700566	rs2280355	SANTOS	A	G	0,4797	0,5203	0,0926	0,7842	0,3624	0,3549	0,707600
2	96945659	rs10202644	SANTOS	T	C	0,5362	0,4638	0,0647	0,7171	0,3027	0,2985	0,882300
2	97717987	rs11123869	SANTOS	A	G	0,4742	0,5258	0,0992	0,7994	0,3745	0,3617	0,461800
2	97748787	rs1466020	SANTOS	G	A	0,4639	0,5361	0,1149	0,8355	0,4016	0,3765	0,158900
2	98126667	rs2084713	SANTOS	G	A	0,4383	0,5617	0,1058	0,8147	0,3863	0,3992	0,505300

2	98158835	rs4851839	SANTOS	G	A,T	0,4195	0,5805	0,1036	0,8097	0,3825	0,4174	0,068650
2	100933381	rs3754675	SANTOS	T	C	0,6664	0,3336	0,0266	0,6159	0,1882	0,1985	0,260200
2	108897145	rs3827760	KIDD|34Plex	A	G	0,7719	0,2281	0,0102	0,5625	0,1111	0,1355	*0,000639*
2	108946039	rs260714	SANTOS	C	T	0,3941	0,6059	0,1166	0,8393	0,4043	0,4468	*0,037440*
2	108963282	rs260690	SELDIN|KIDD|SANTOS	C	A	0,4679	0,5321	0,0746	0,7413	0,3255	0,3656	*0,015410*
2	115503413	rs4241135	SANTOS	A	G	0,4987	0,5013	0,0810	0,7566	0,3392	0,3347	0,894900
2	115558939	rs17044697	SANTOS	T	C	0,4869	0,5131	0,0835	0,7627	0,3444	0,3466	0,898500
2	115605118	rs10205952	SANTOS	T	G	0,4762	0,5238	0,0838	0,7634	0,3450	0,3574	0,458100
2	135859184	rs182549	34PLEX	C	T	0,4950	0,5050	0,0628	0,7125	0,2982	0,3402	*0,006288*
2	135950412	rs6754311	KIDD	T	C	0,4976	0,5024	0,0612	0,7086	0,2943	0,3380	*0,005681*
2	145012376	rs10496971	SELDIN	T	G	0,6512	0,3488	0,0274	0,6181	0,1910	0,2125	*0,034260*
2	154043812	rs1863086	PIMA	G	T	0,3707	0,6293	0,1791	0,9808	0,4902	0,4995	0,722800
2	157810705	rs10497191	KIDD|SANTOS	T	C	0,4263	0,5737	0,0845	0,7650	0,3464	0,4073	*0,001049*
2	174630583	rs1991600	SANTOS	C	T	0,4220	0,5780	0,1175	0,8416	0,4059	0,4192	0,461200
2	174661054	rs12470024	SANTOS	G	A	0,4664	0,5336	0,0948	0,7892	0,3665	0,3686	0,904700
2	178741811	rs2627037	SELDIN	G	A	0,5167	0,4833	0,0607	0,7072	0,2930	0,3191	0,071060
2	178911166	rs10179812	SANTOS	T	G	0,6324	0,3676	0,0364	0,6436	0,2231	0,2191	0,840500
2	200157231	rs1569175	SELDIN	T	C	0,6110	0,3890	0,0448	0,6662	0,2495	0,2328	0,128800
2	239061654	rs4852013	SANTOS	G	A	0,5552	0,4448	0,0612	0,7086	0,2943	0,2804	0,343400
2	239360513	rs12471054	SANTOS	C	T	0,6085	0,3915	0,0409	0,6560	0,2378	0,2385	1,000000
2	239373918	rs2176046	SANTOS	G	A	0,6233	0,3767	0,0382	0,6485	0,2290	0,2264	1,000000
3	2167148	rs10510228	SELDIN	A	G	0,6413	0,3587	0,0332	0,6349	0,2125	0,2140	0,836600
3	18354833	rs7432238	SANTOS	T	G	0,6486	0,3514	0,0306	0,6275	0,2032	0,2102	0,402300
3	21268289	rs17008327	SANTOS	G	A	0,6746	0,3254	0,0269	0,6166	0,1891	0,1899	0,818300
3	30374120	rs4955316	SELDIN	T	G	0,6157	0,3843	0,0379	0,6477	0,2281	0,2342	0,570200
3	39104938	rs9809104	SELDIN	T	C	0,3847	0,6153	0,1001	0,8016	0,3762	0,4514	*0,000193*
3	49520424	rs3870336	SANTOS	G	A	0,5933	0,4067	0,0405	0,6547	0,2363	0,2542	0,117000
3	54875169	rs4955903	SANTOS	T	C	0,4011	0,5989	0,1351	0,8814	0,4327	0,4470	0,489400
3	68922709	rs9871910	SANTOS	G	A,T	0,4492	0,5508	0,0963	0,7928	0,3694	0,3860	0,357800

3	79350425	rs6548616	SELDIN	T	C	0,3746	0,6254	0,1783	0,9790	0,4893	0,4953	0,789400
3	109148281	rs4855697	SANTOS	A	G	0,4120	0,5880	0,2024	1,0343	0,5166	0,4717	*0,039240*
3	120803869	rs12629908	SELDIN	G	A	0,6327	0,3673	0,0355	0,6413	0,2203	0,2198	1,000000
3	121740742	rs12498138	KIDD|PIMA	G	A	0,7106	0,2894	0,0211	0,5993	0,1657	0,1649	1,000000
3	123653126	rs702032	SANTOS	C	T	0,4172	0,5828	0,0835	0,7627	0,3444	0,4164	**0,000122**
3	136195634	rs9845457	SELDIN	G	A	0,3509	0,6491	0,1398	0,8920	0,4395	0,4993	*0,007834*
3	139391293	rs10935323	SANTOS	T	G	0,4563	0,5437	0,0857	0,7680	0,3489	0,3774	0,100700
3	141658488	rs868767	SANTOS	A	G	0,5835	0,4165	0,0428	0,6611	0,2437	0,2619	0,127400
3	148032568	rs734873	SELDIN	G	A	0,6442	0,3558	0,0310	0,6287	0,2047	0,2140	0,303700
3	168927247	rs1498444	34PLEX	T	G	0,3726	0,6274	0,1222	0,8522	0,4133	0,4703	*0,006443*
3	180246939	rs2030763	SELDIN	G	A	0,5570	0,4430	0,0509	0,6822	0,2671	0,2829	0,211200
3	188857208	rs1513181	SELDIN	G	A	0,4824	0,5176	0,0819	0,7589	0,3411	0,3510	0,529600
3	188976960	rs710493	SANTOS	G	A	0,5428	0,4572	0,0409	0,6560	0,2378	0,3042	**0,000004**
4	5388910	rs9291090	SELDIN	A	C	0,7926	0,2074	0,0105	0,5637	0,1131	0,1136	0,705900
4	18627902	rs4698702	SANTOS	C	A	0,5285	0,4715	0,0694	0,7287	0,3138	0,3054	0,664100
4	18635692	rs10012227	PIMA	G	A	0,5285	0,4715	0,0694	0,7287	0,3138	0,3054	0,664100
4	18655760	rs2874414	SANTOS	A	G	0,4441	0,5559	0,1088	0,8215	0,3914	0,3943	0,910600
4	38801634	rs4540055	34PLEX	A	T,C	0,2178	0,7822	0,2441	1,1327	0,4318	0,4764	0,076830
4	38813881	rs4833103	KIDD	A	C	0,4157	0,5843	0,0907	0,7796	0,3587	0,4186	*0,001491*
4	41102115	rs1876495	SANTOS	C	T	0,4867	0,5133	0,0746	0,7413	0,3255	0,3468	0,162500
4	41552347	rs10007810	SELDIN	G	A	0,3873	0,6127	0,1406	0,8937	0,4405	0,4633	0,294200
4	72379474	rs1369093	SELDIN	T	C	0,5892	0,4108	0,0409	0,6560	0,2378	0,2579	0,085720
4	72390023	rs7660290	SANTOS	G	T	0,4009	0,5991	0,1227	0,8533	0,4141	0,4423	0,161500
4	84387925	rs385194	SELDIN	A	G	0,3613	0,6387	0,1624	0,9430	0,4698	0,5000	0,185200
4	99318162	rs1229984	KIDD	T	C	0,7025	0,2975	0,0198	0,5951	0,1598	0,1759	*0,044300*
4	99323162	rs3811801	KIDD	G	A	0,9466	0,0534	0,0004	0,5099	0,0195	0,0345	**0,000004**
4	100100370	rs4532240	SANTOS	G	T	0,4668	0,5332	0,0969	0,7941	0,3704	0,3686	1,000000
4	104454266	rs7657799	SELDIN|KIDD	T	G	0,5895	0,4105	0,0391	0,6510	0,2320	0,2592	*0,025070*
4	158260811	rs2026721	34PLEX	T	C	0,5906	0,4094	0,0441	0,6645	0,2476	0,2538	0,600600

4	164363422	rs4431170	SANTOS	A	G	0,6861	0,3139	0,0239	0,6078	0,1774	0,1837	0,466200
4	178356420	rs2063393	SANTOS	T	C	0,4014	0,5986	0,1311	0,8724	0,4269	0,4451	0,372000
4	178478369	rs2702414	SELDIN	G	A	0,5647	0,4353	0,0467	0,6715	0,2554	0,2778	0,078760
4	184629610	rs4647693	SANTOS	T	C	0,4263	0,5737	0,1146	0,8350	0,4012	0,4139	0,521100
5	2364512	rs316598	SELDIN	T	C	0,3733	0,6267	0,1671	0,9535	0,4756	0,4904	0,528400
5	6844922	rs870347	SELDIN|KIDD	A	C	0,6483	0,3517	0,0274	0,6181	0,1910	0,2155	*0,013740*
5	22475991	rs10473594	SANTOS	T	C	0,4552	0,5448	0,0848	0,7657	0,3470	0,3784	0,062100
5	22557515	rs13357733	SANTOS	G	A	0,5328	0,4672	0,0495	0,6786	0,2632	0,3079	*0,001499*
5	31895737	rs12522914	SANTOS	C	T	0,6288	0,3712	0,0379	0,6478	0,2282	0,2211	0,678800
5	33951588	rs16891982	KIDD|34Plex	C	G	0,3501	0,6499	0,0984	0,7975	0,3730	0,4856	**1,83E-07**
5	33958854	rs28777	SANTOS	C	A	0,3868	0,6132	0,0867	0,7703	0,3509	0,4470	**0,000002**
5	35037010	rs37369	SELDIN	C	T	0,4823	0,5177	0,0764	0,7456	0,3294	0,3510	0,166700
5	43711276	rs6451722	SELDIN	G	A	0,4051	0,5949	0,1338	0,8784	0,4308	0,4425	0,550100
5	79789903	rs12657828	SELDIN	A	G	0,5023	0,4977	0,0728	0,7371	0,3216	0,3313	0,505800
5	103862724	rs1895218	SANTOS	C	A,T	0,4857	0,5143	0,0850	0,7663	0,3475	0,3479	1,000000
5	118158945	rs10079352	SANTOS	G	A	0,3646	0,6354	0,1351	0,8814	0,4327	0,4836	*0,017670*
5	118313780	rs6880706	PIMA	A	G	0,3534	0,6466	0,1467	0,9075	0,4490	0,5000	*0,021440*
5	150685982	rs252155	PIMA	T	C	0,3941	0,6059	0,2035	1,0367	0,5177	0,4902	0,239300
5	153494989	rs4145160	SANTOS	G	A	0,6206	0,3794	0,0373	0,6461	0,2261	0,2300	0,700400
5	154118220	rs1366220	SANTOS	G	A	0,3995	0,6005	0,1023	0,8066	0,3801	0,4371	*0,003386*
5	155934106	rs17053011	SANTOS	G	T	0,5887	0,4113	0,0448	0,6662	0,2495	0,2551	0,603800
5	156044704	rs6556352	SELDIN	C	T	0,3799	0,6201	0,1850	0,9942	0,4971	0,4936	0,928800
5	166312977	rs1500127	SELDIN	C	T	0,5422	0,4578	0,0552	0,6932	0,2788	0,2956	0,230300
5	167792263	rs10454965	SANTOS	A	G	0,6166	0,3834	0,0401	0,6538	0,2353	0,2311	0,848100
5	168740376	rs7728435	SANTOS	T	C	0,4026	0,5974	0,1423	0,8976	0,4430	0,4487	0,834600
5	170750576	rs2042314	SANTOS	C	T	0,5066	0,4934	0,0612	0,7086	0,2943	0,3290	*0,021780*
5	170775980	rs7722456	KIDD	C	T	0,4038	0,5962	0,1102	0,8248	0,3938	0,4350	*0,033120*
5	173034240	rs4868237	SANTOS	C	T	0,5610	0,4390	0,0559	0,6951	0,2807	0,2765	0,873100
5	178436082	rs6422347	SELDIN	T	C	0,4598	0,5402	0,0782	0,7500	0,3333	0,3735	*0,017700*

5	179470589	rs4701170	SANTOS	G	A	0,5380	0,4620	0,0635	0,7143	0,3000	0,2970	1,000000
6	4746925	rs1040045	SELDIN	G	A	0,3971	0,6029	0,1541	0,9242	0,4590	0,4599	1,000000
6	12534879	rs2504853	SELDIN	T	C	0,3689	0,6311	0,1490	0,9128	0,4522	0,4856	0,122500
6	21911385	rs7745461	SELDIN	A	G	0,4823	0,5177	0,0764	0,7456	0,3294	0,3510	0,166700
6	51746672	rs2397060	SELDIN	T	C	0,4357	0,5643	0,0819	0,7589	0,3411	0,3977	*0,001775*
6	89808559	rs192655	SELDIN|KIDD	G	A	0,5127	0,4873	0,0644	0,7165	0,3021	0,3221	0,169900
6	108717130	rs11153123	SANTOS	A	G	0,4430	0,5570	0,1161	0,8382	0,4035	0,3977	0,824400
6	134435989	rs4896046	SANTOS	C	T	0,4589	0,5411	0,0911	0,7807	0,3595	0,3755	0,345200
6	136161589	rs3823159	KIDD|SANTOS	A	G	0,5308	0,4693	0,0385	0,6494	0,2300	0,3186	**3,49E-09**
6	136195119	rs7753890	SANTOS	T	C	0,5689	0,4311	0,0349	0,6397	0,2183	0,2842	**0,000001**
6	144734195	rs4463276	SELDIN	G	A	0,3513	0,6487	0,1298	0,8695	0,4250	0,4947	*0,001746*
6	144769895	rs6570644	SANTOS	A	G	0,3458	0,6542	0,1285	0,8666	0,4230	0,4996	*0,000557*
6	155223263	rs4143683	SANTOS	T	C	0,4448	0,5552	0,1045	0,8117	0,3840	0,3924	0,652600
6	158084056	rs1744173	SANTOS	A	G	0,5686	0,4314	0,0537	0,6895	0,2749	0,2699	0,869700
6	158088718	rs350289	SANTOS	A	G	0,5773	0,4227	0,0523	0,6858	0,2710	0,2619	0,503400
6	161647400	rs10945764	SANTOS	A	G	0,4187	0,5813	0,1102	0,8248	0,3938	0,4202	0,171400
6	161669949	rs16888746	SANTOS	G	A,C	0,6196	0,3804	0,0342	0,6376	0,2158	0,2345	0,085250
6	161672108	rs10484813	SANTOS	A	G	0,5562	0,4438	0,0523	0,6858	0,2710	0,2829	0,348300
6	162800760	rs4458655	SELDIN	T	C	0,5047	0,4953	0,0711	0,7329	0,3177	0,3290	0,422400
6	168265080	rs1871428	SELDIN	G	A	0,3687	0,6313	0,1631	0,9446	0,4707	0,4929	0,324000
7	4417372	rs917118	34PLEX	C	T	0,3902	0,6098	0,1609	0,9396	0,4678	0,4703	0,925200
7	12494110	rs1550392	SANTOS	C	T	0,4808	0,5192	0,1001	0,8016	0,3762	0,3553	0,214100
7	12504168	rs17166119	SANTOS	A	G	0,4393	0,5607	0,1344	0,8797	0,4316	0,4085	0,234000
7	12629626	rs731257	SELDIN|SANTOS	G	A	0,5761	0,4239	0,0495	0,6786	0,2632	0,2646	0,867900
7	28132967	rs917115	KIDD	T	C	0,3824	0,6176	0,1490	0,9128	0,4522	0,4722	0,350000
7	32139512	rs32314	SELDIN	C	T	0,4204	0,5796	0,1351	0,8814	0,4327	0,4278	0,836900
7	38721910	rs12701660	SANTOS	T	C	0,3985	0,6015	0,1530	0,9218	0,4576	0,4580	1,000000
7	42340472	rs2330442	SELDIN	A	G	0,3645	0,6355	0,1534	0,9227	0,4581	0,4922	0,127100
7	45937574	rs2471552	PIMA|SANTOS	C	T	0,5043	0,4957	0,0764	0,7456	0,3294	0,3290	1,000000

7	74039869	rs4717865	SELDIN	G	A	0,7162	0,2838	0,0181	0,5897	0,1520	0,1664	0,059460
7	83903731	rs10954737	SELDIN	C	T							
7	98066051	rs705308	SELDIN	C	A	0,3632	0,6368	0,1234	0,8550	0,4152	0,4802	*0,002381*
7	127024098	rs978874	SANTOS	C	T	0,3946	0,6054	0,1511	0,9176	0,4551	0,4610	0,774000
7	127135144	rs2040502	SANTOS	C	T	0,3951	0,6049	0,1419	0,8969	0,4425	0,4561	0,498800
7	131057307	rs7803075	SELDIN	A	G	0,3473	0,6527	0,1325	0,8754	0,4288	0,4997	*0,001447*
7	139747578	rs10236187	SELDIN	A	C							
7	152176768	rs6464211	SELDIN	C	T	0,4163	0,5837	0,1125	0,8301	0,3977	0,4232	0,175400
8	4333271	rs10108270	SELDIN	C	A,T	0,3804	0,6196	0,1702	0,9607	0,4795	0,4850	0,785500
8	4980500	rs4875427	SANTOS	G	A	0,4351	0,5649	0,1390	0,8902	0,4384	0,4148	0,240300
8	5061156	rs17072176	SANTOS	T	G	0,5451	0,4549	0,0625	0,7117	0,2975	0,2902	0,649300
8	8912944	rs400404	SANTOS	T	G	0,3974	0,6026	0,1365	0,8845	0,4347	0,4514	0,433700
8	9415737	rs11776114	SANTOS	G	A	0,4228	0,5772	0,1104	0,8252	0,3941	0,4160	0,241900
8	11605506	rs17807624	SANTOS	C	T	0,3975	0,6025	0,1325	0,8754	0,4288	0,4495	0,325700
8	13501991	rs3943253	SELDIN	A	G	0,5318	0,4682	0,0644	0,7165	0,3021	0,3030	1,000000
8	29083788	rs1471939	SELDIN	C	T	0,4304	0,5696	0,1090	0,8221	0,3918	0,4080	0,386600
8	32039076	rs1462906	KIDD	T	C	0,5795	0,4205	0,0327	0,6333	0,2105	0,2765	**0,000001**
8	81092622	rs7018273	SANTOS	A	G	0,5718	0,4282	0,0441	0,6645	0,2476	0,2725	*0,049940*
8	85512387	rs12544346	SELDIN	G	A	0,3619	0,6381	0,1624	0,9430	0,4698	0,4994	0,185100
8	109590088	rs6990312	KIDD	G	T	0,4373	0,5627	0,1173	0,8410	0,4055	0,4038	1,000000
8	113789154	rs896680	SANTOS	T	C	0,4901	0,5099	0,0848	0,7657	0,3470	0,3435	0,898000
8	116169693	rs17742080	PIMA	C	T	0,6595	0,3405	0,0294	0,6241	0,1988	0,2010	0,823400
8	121112062	rs2196051	KIDD	A	G	0,3691	0,6309	0,1767	0,9753	0,4873	0,4998	0,596100
8	121896264	rs7844723	SELDIN	C	T	0,3706	0,6294	0,1639	0,9465	0,4717	0,4915	0,369500
8	125966119	rs4870958	SANTOS	A	G	0,5212	0,4788	0,0694	0,7287	0,3138	0,3127	1,000000
8	133861643	rs6993205	PIMA|SANTOS	A	G	0,3544	0,6456	0,1338	0,8784	0,4308	0,4932	*0,004197*
8	139228938	rs2001907	SELDIN	C	T	0,6625	0,3375	0,0269	0,6166	0,1891	0,2021	0,184300
8	144414297	rs1871534	KIDD	C	G	0,5412	0,4588	0,0344	0,6381	0,2164	0,3127	**1,17E-10**
9	12672320	rs1408801	SELDIN	A	G	0,4494	0,5506	0,0917	0,7820	0,3606	0,3850	0,168100

9	16795288	rs10962599	SANTOS	C	T	0,3574	0,6426	0,1549	0,9260	0,4600	0,5000	0,077210
9	25872207	rs10812316	SANTOS	T	C	0,4490	0,5510	0,0719	0,7349	0,3196	0,3846	*0,000206*
9	28628502	rs10511828	SELDIN	T	C	0,6629	0,3371	0,0294	0,6241	0,1988	0,1975	1,000000
9	69044364	rs3793451	SELDIN	C	T	0,7426	0,2574	0,0160	0,5830	0,1423	0,1454	0,544300
9	89256441	rs4877068	SANTOS	G	A	0,6343	0,3657	0,0351	0,6400	0,2188	0,2188	1,000000
9	90878917	rs2306040	SELDIN	T	C	0,7011	0,2989	0,0234	0,6064	0,1754	0,1696	0,604600
9	110246294	rs4135211	SANTOS	A	C	0,6788	0,3212	0,0254	0,6122	0,1832	0,1884	0,484100
9	113681989	rs4979274	PIMA|SANTOS	C	A	0,5719	0,4281	0,0478	0,6742	0,2584	0,2698	0,324900
9	117367927	rs10513300	SELDIN	T	C	0,7005	0,2995	0,0230	0,6050	0,1735	0,1712	1,000000
9	124505410	rs3814134	KIDD	A	G	0,5258	0,4742	0,0537	0,6895	0,2749	0,3127	*0,007293*
9	133057735	rs2073821	SELDIN	C	T	0,6606	0,3394	0,0300	0,6256	0,2008	0,1990	1,000000
9	135482018	rs1412512	SANTOS	T	C	0,4258	0,5742	0,1298	0,8695	0,4250	0,4202	0,834200
10	1488620	rs4880510	SANTOS	T	G	0,6138	0,3862	0,0364	0,6438	0,2233	0,2377	0,185100
10	1488689	rs4880511	SANTOS	C	A	0,6366	0,3634	0,0320	0,6313	0,2080	0,2202	0,220900
10	8272081	rs17143387	SANTOS	T	G	0,5798	0,4202	0,0441	0,6645	0,2476	0,2646	0,179300
10	10714775	rs7923994	SANTOS	G	A	0,4191	0,5809	0,1185	0,8438	0,4074	0,4225	0,405000
10	10903388	rs912069	SANTOS	T	C	0,4412	0,5588	0,1045	0,8117	0,3840	0,3960	0,503700
10	17022993	rs7897550	34PLEX	G	A	0,4829	0,5171	0,0877	0,7726	0,3528	0,3510	1,000000
10	20220010	rs6482103	SANTOS	C	A	0,6204	0,3797	0,0349	0,6397	0,2183	0,2328	0,181000
10	34466420	rs1978806	34PLEX	A	G	0,6989	0,3011	0,0193	0,5938	0,1579	0,1806	*0,010950*
10	49633658	rs3793791	SELDIN	C	T	0,5546	0,4454	0,0530	0,6877	0,2729	0,2842	0,352800
10	52771028	rs201381710/rs17287498	SANTOS|PIMA	C	T,A	0,2513	0,7487	0,3204	1,3290	0,4930	0,4531	0,104100
10	60376448	rs10821745	SANTOS	T	G	0,6843	0,3157	0,0249	0,6107	0,1813	0,1837	0,808400
10	60425736	rs4948418	SANTOS	C	T	0,7100	0,2900	0,0207	0,5979	0,1637	0,1664	0,601500
10	60519366	rs10994397	SANTOS	C	T	0,7011	0,2989	0,0211	0,5993	0,1657	0,1744	0,303800
10	62308584	rs3998448	SANTOS	C	T	0,4068	0,5932	0,1314	0,8731	0,4273	0,4398	0,526900
10	73541236	rs4746136	SELDIN|SANTOS	G	A	0,5999	0,4001	0,0428	0,6611	0,2437	0,2455	0,857200
10	81963701	rs7902158	SANTOS	A	G	0,3614	0,6386	0,1325	0,8754	0,4288	0,4856	*0,008382*
10	86368524	rs3814614	PIMA	A	G	0,3695	0,6305	0,1677	0,9551	0,4765	0,4946	0,420600

10	93161308	rs4918664	KIDD	A	G	0,5946	0,4054	0,0409	0,6560	0,2378	0,2524	0,218000
10	93215254	rs12220128	SANTOS	A	C	0,5407	0,4593	0,0545	0,6916	0,2770	0,2974	0,134800
10	111873049	rs7917775	SANTOS	A	C	0,6101	0,3899	0,0395	0,6522	0,2333	0,2383	0,580600
10	113436260	rs17130385	PIMA|SANTOS	G	T,A	0,6274	0,3726	0,0338	0,6365	0,2144	0,2271	0,239600
10	113479843	rs734241	SANTOS	G	A	0,6057	0,3943	0,0367	0,6445	0,2242	0,2455	0,069110
10	113515606	rs7919248	SANTOS	T	G	0,5655	0,4345	0,0474	0,6732	0,2573	0,2765	0,112200
10	113532698	rs6585227	SANTOS	T	C	0,5181	0,4819	0,0586	0,7019	0,2877	0,3184	*0,036150*
10	113557053	rs4918842	SELDIN|SANTOS	T	C	0,5273	0,4727	0,0628	0,7125	0,2982	0,3079	0,473900
10	117971770	rs1419138	SANTOS	C	T	0,4339	0,5661	0,0800	0,7544	0,3372	0,3995	*0,000571*
10	118061825	rs853577	SANTOS	A	G	0,4485	0,5515	0,0728	0,7371	0,3216	0,3850	*0,000337*
10	132836599	rs4880436	SELDIN	C	T	0,7571	0,2429	0,0144	0,5777	0,1345	0,1355	0,745300
11	7828769	rs10839880	SELDIN	C	T	0,3629	0,6371	0,1639	0,9465	0,4717	0,4992	0,216200
11	11316631	rs2403512	SANTOS	T	C	0,5738	0,4262	0,0423	0,6597	0,2421	0,2721	*0,020910*
11	15816591	rs1837606	SELDIN	T	C	0,3799	0,6201	0,1505	0,9161	0,4542	0,4753	0,308300
11	23988984	rs2946788	SELDIN	G	T	0,3976	0,6024	0,1462	0,9064	0,4483	0,4555	0,771100
11	25563634	rs1561937	SANTOS	G	A	0,4811	0,5189	0,0537	0,6895	0,2749	0,3574	**0,000001**
11	32402843	rs5030240	34PLEX	C	G,A	0,2731	0,7269	0,2389	1,1201	0,4315	0,4188	0,627400
11	45953579	rs3736508	SANTOS	C	T	0,7372	0,2628	0,0168	0,5856	0,1462	0,1487	0,562900
11	61781087	rs509360	PIMA	A	G	0,3784	0,6216	0,1718	0,9643	0,4815	0,4878	0,786100
11	61802358	rs174546	SANTOS	C	T	0,4364	0,5636	0,1106	0,8258	0,3945	0,4026	0,660800
11	61829740	rs174570	KIDD|SANTOS	C	T	0,6046	0,3954	0,0422	0,6594	0,2417	0,2413	1,000000
11	61830500	rs1535	SANTOS	A	G	0,4271	0,5729	0,1304	0,8707	0,4258	0,4191	0,752700
11	62048331	rs4963452	SANTOS	T	C	0,5265	0,4735	0,0737	0,7391	0,3235	0,3070	0,309800
11	67131021	rs11227699	SELDIN	G	A	0,5477	0,4523	0,0516	0,6840	0,2690	0,2918	0,094190
11	67195954	rs674499	SANTOS	A	G	0,4774	0,5226	0,0716	0,7342	0,3190	0,3562	*0,024690*
11	67356862	rs1638566	SANTOS	T	G	0,4700	0,5300	0,0728	0,7371	0,3216	0,3635	*0,010560*
11	67357752	rs1638567	SANTOS	T	C	0,4700	0,5300	0,0728	0,7371	0,3216	0,3635	*0,010560*
11	69614368	rs599190	SANTOS	A	G	0,4573	0,5427	0,1023	0,8066	0,3801	0,3793	1,000000
11	80266327	rs2619206	SANTOS	G	A	0,5811	0,4189	0,0474	0,6733	0,2574	0,2609	0,734600

11	94830565	rs2241667	SANTOS	T	G	0,4799	0,5201	0,0984	0,7975	0,3730	0,3558	0,320500
11	94831082	rs16921059	SANTOS	T	C	0,4685	0,5315	0,1060	0,8153	0,3867	0,3691	0,337700
11	94858124	rs4309121	SANTOS	T	C	0,5431	0,4569	0,0636	0,7145	0,3002	0,2918	0,649400
11	113425564	rs1079597	KIDD	C	T	0,5587	0,4413	0,0502	0,6804	0,2651	0,2817	0,207800
11	120773738	rs948028	SELDIN	A	C	0,4184	0,5816	0,1406	0,8937	0,4405	0,4321	0,759100
11	129348709	rs4937417	PIMA	A	G	0,3784	0,6216	0,1406	0,8937	0,4405	0,4722	0,134700
12	4290330	rs3217863	PIMA	C	T	0,7053	0,2947	0,0220	0,6021	0,1696	0,1682	1,000000
12	11548554	rs2416791	SELDIN	A	G	0,4225	0,5775	0,1012	0,8041	0,3782	0,4138	0,054670
12	17254858	rs1513056	SELDIN	G	A	0,4836	0,5164	0,0848	0,7657	0,3470	0,3500	0,802100
12	29216938	rs10843344	34PLEX	C	T	0,4360	0,5640	0,1012	0,8041	0,3782	0,4004	0,224700
12	47283167	rs214678	SELDIN	T	C	0,6240	0,3760	0,0355	0,6413	0,2203	0,2285	0,437200
12	47562248	rs2051827	PIMA|SANTOS	G	A	0,6669	0,3331	0,0279	0,6196	0,1930	0,1960	0,654600
12	51955304	rs11612312	SANTOS	T	C	0,4688	0,5312	0,0969	0,7941	0,3704	0,3666	0,904200
12	52077375	rs11521	SANTOS	C	T	0,4792	0,5208	0,0914	0,7813	0,3601	0,3552	0,804500
12	55769950	rs772262	SELDIN	G	A	0,4502	0,5498	0,0800	0,7544	0,3372	0,3831	*0,007862*
12	56210050	rs773658	34PLEX	G	C	0,5977	0,4023	0,0391	0,6510	0,2320	0,2510	0,110400
12	68457144	rs2216437	SANTOS	A	G	0,5600	0,4400	0,0552	0,6932	0,2788	0,2778	1,000000
12	68461952	rs10878834	SANTOS	T	C	0,5628	0,4372	0,0539	0,6900	0,2754	0,2756	1,000000
12	68483055	rs2052386	SANTOS	G	A	0,5736	0,4264	0,0516	0,6840	0,2690	0,2659	1,000000
12	71217099	rs3851621	SANTOS	T	C	0,6788	0,3212	0,0254	0,6122	0,1832	0,1884	0,484100
12	71225011	rs1397560	SANTOS	C	A	0,5109	0,4891	0,0731	0,7378	0,3223	0,3226	1,000000
12	108883944	rs2070586	SELDIN	G	A	0,4672	0,5328	0,0938	0,7868	0,3645	0,3676	0,811300
12	111774029	rs2238151	KIDD	T	C	0,3560	0,6440	0,1378	0,8875	0,4366	0,4934	*0,009472*
12	111803962	rs671	KIDD	G	A	0,9692	0,0308	0,0001	0,5049	0,0097	0,0212	**0,000006**
12	118451683	rs1503767	SELDIN	T	G,A	0,4068	0,5932	0,1056	0,8143	0,3860	0,4307	*0,018640*
13	20327585	rs1335873	34PLEX	T	A	0,3722	0,6278	0,1655	0,9500	0,4737	0,4907	0,471400
13	21800561	rs1886510	34PLEX	G	A	0,3742	0,6258	0,1534	0,9227	0,4581	0,4825	0,272000
13	23093195	rs17359176	SANTOS	G	A	0,6721	0,3279	0,0294	0,6241	0,1988	0,1884	0,343300
13	27050219	rs9319336	SELDIN	T	C	0,5927	0,4073	0,0416	0,6577	0,2398	0,2538	0,221700

13	33086648	rs566996	SANTOS	C	T	0,4912	0,5088	0,0946	0,7888	0,3661	0,3437	0,157200
13	34273600	rs7997709	SELDIN|KIDD	C	T	0,5665	0,4335	0,0502	0,6804	0,2651	0,2739	0,517700
13	34290103	rs2065982	34PLEX	T	C	0,6376	0,3624	0,0327	0,6333	0,2105	0,2184	0,417000
13	41141146	rs1572018	KIDD	T	C	0,3952	0,6048	0,1185	0,8438	0,4074	0,4464	*0,048260*
13	42005849	rs2166624	KIDD|SANTOS	G	A	0,3893	0,6107	0,1378	0,8875	0,4366	0,4601	0,250200
13	48496376	rs7326934	KIDD	G	C	0,5289	0,4711	0,0481	0,6750	0,2593	0,3127	*0,000208*
13	62784039	rs9563982	SANTOS	A	G	0,5361	0,4639	0,0677	0,7246	0,3099	0,2981	0,458600
13	66094458	rs1409842	PIMA	A	G	0,3593	0,6407	0,1564	0,9293	0,4620	0,4989	0,093260
13	72981494	rs1000973	SANTOS	G	A	0,4387	0,5613	0,0935	0,7862	0,3640	0,3960	0,073560
13	75419751	rs9530435	SELDIN	T	C	0,4162	0,5838	0,0945	0,7886	0,3659	0,4187	*0,005827*
13	111174820	rs9522149	SELDIN|KIDD	T	C	0,3662	0,6338	0,1505	0,9161	0,4542	0,4890	0,104900
14	20349972	rs1760921	SELDIN	T	C	0,5260	0,4740	0,0559	0,6951	0,2807	0,3115	*0,032400*
14	32951006	rs11844034	SANTOS	G	A	0,4140	0,5860	0,1285	0,8666	0,4230	0,4314	0,682300
14	35701401	rs10141763	34PLEX	T	A	0,5935	0,4065	0,0379	0,6477	0,2281	0,2565	*0,015400*
14	46449839	rs7158003	PIMA	T	C	0,3902	0,6098	0,1750	0,9716	0,4854	0,4778	0,781700
14	47262835	rs1958099	SANTOS	T	G	0,4857	0,5143	0,0911	0,7807	0,3595	0,3487	0,527300
14	47780999	rs17119138	SANTOS	A	G	0,4591	0,5409	0,1090	0,8221	0,3918	0,3793	0,487400
14	52141249	rs2357442	SELDIN	A	C	0,5233	0,4767	0,0516	0,6840	0,2690	0,3162	*0,001179*
14	53720410	rs17126825	SANTOS	G	A	0,5814	0,4186	0,0474	0,6732	0,2573	0,2606	0,735800
14	57771969	rs1950993	SELDIN	G	T	0,3624	0,6376	0,1639	0,9465	0,4717	0,4997	0,216100
14	61555161	rs810087	SANTOS	A	G	0,4071	0,5929	0,1369	0,8854	0,4353	0,4418	0,763700
14	67420064	rs8021730	SELDIN	G	T	0,3967	0,6033	0,1260	0,8607	0,4191	0,4476	0,166900
14	78783801	rs7142344	SANTOS	C	T	0,4304	0,5696	0,0800	0,7544	0,3372	0,4030	*0,000284*
14	83006524	rs946918	SELDIN	G	T	0,4435	0,5565	0,0948	0,7892	0,3665	0,3915	0,144000
14	98908984	rs200354	SELDIN|KIDD	G	T	0,4078	0,5922	0,1338	0,8784	0,4308	0,4398	0,687900
14	100676553	rs730570	34PLEX	G	A	0,3941	0,6059	0,1173	0,8410	0,4055	0,4470	*0,037950*
14	105212718	rs3784230	SELDIN	A	G	0,3663	0,6337	0,1718	0,9643	0,4815	0,5000	0,426700
15	26665997	rs11161328	SANTOS	A	G	0,4081	0,5919	0,1405	0,8937	0,4405	0,4425	0,915800
15	27951891	rs1800414	KIDD	T	C	0,9541	0,0459	0,0002	0,5079	0,0156	0,0307	**0,000001**

15	28038295	rs1448484	SANTOS	A	G	0,5967	0,4033	0,0289	0,6226	0,1969	0,2646	**1,17E-07**
15	28120472	rs12913832	KIDD|34Plex	A	G	0,3666	0,6334	0,0887	0,7749	0,3548	0,4674	**5,98E-08**
15	35927834	rs12439433	SELDIN|KIDD	A	G	0,8964	0,1036	0,0023	0,5267	0,0507	0,0568	0,063360
15	36955045	rs2122497	SANTOS	C	T	0,4682	0,5318	0,0851	0,7666	0,3477	0,3655	0,275600
15	42689608	rs8030587	SANTOS	G	A	0,3964	0,6036	0,1034	0,8091	0,3821	0,4405	*0,003537*
15	42729788	rs12594483	SANTOS	G	A	0,4255	0,5745	0,0838	0,7634	0,3450	0,4080	*0,000738*
15	42731284	rs12917189	SANTOS	T	C	0,3925	0,6075	0,1125	0,8301	0,3977	0,4470	*0,013460*
15	44860173	rs735480	KIDD	T	C	0,4155	0,5845	0,0782	0,7500	0,3333	0,4178	**0,000008**
15	48134287	rs1426654	KIDD|34Plex	A	G	0,4516	0,5484	0,0537	0,6895	0,2749	0,3869	**2,63E-10**
15	48340956	rs11637235	SANTOS	C	T,G	0,3524	0,6476	0,1357	0,8828	0,4336	0,4960	*0,004391*
15	55034973	rs12594750	SANTOS	T	C	0,5617	0,4384	0,0544	0,6914	0,2768	0,2765	1,000000
15	60563356	rs974828	SANTOS	C	T	0,4251	0,5749	0,0887	0,7749	0,3548	0,4089	*0,003425*
15	61410580	rs3884558	SANTOS	A	G	0,5410	0,4590	0,0582	0,7008	0,2865	0,2956	0,457800
15	63365538	rs11631047	SANTOS	C	T	0,4248	0,5752	0,0803	0,7552	0,3379	0,4085	*0,000137*
15	64741112	rs6494489	SANTOS	C	T	0,3832	0,6168	0,1402	0,8929	0,4400	0,4672	0,213100
15	67673356	rs16951105	SANTOS	A	G	0,4412	0,5588	0,1209	0,8493	0,4113	0,4012	0,659700
15	74442159	rs2899826	SELDIN	A	G	0,4796	0,5204	0,0867	0,7703	0,3509	0,3542	0,803900
15	91562478	rs8035124	SELDIN	A	C	0,3970	0,6030	0,1519	0,9194	0,4561	0,4589	0,923400
16	690466	rs4984913	SELDIN	A	G	0,3776	0,6224	0,1609	0,9396	0,4678	0,4829	0,522000
16	8417726	rs4141505	SANTOS	G	A	0,6173	0,3827	0,0391	0,6510	0,2320	0,2314	1,000000
16	10881454	rs4781011	SELDIN	T	G	0,3975	0,6025	0,1454	0,9046	0,4473	0,4554	0,698100
16	12768150	rs4780476	PIMA	A	C	0,4361	0,5639	0,0980	0,7966	0,3723	0,3995	0,122600
16	19261586	rs2269793	SELDIN	T	G	0,5054	0,4946	0,0661	0,7205	0,3060	0,3290	0,138500
16	30811726	rs4889490	SANTOS	G	T	0,3889	0,6111	0,1517	0,9188	0,4558	0,4669	0,634900
16	31068050	rs881929	34PLEX	G	T	0,3803	0,6197	0,1245	0,8574	0,4168	0,4635	*0,022310*
16	65372805	rs818386	SELDIN	T	C	0,5513	0,4487	0,0544	0,6914	0,2768	0,2868	0,440900
16	85150076	rs2966849	SELDIN	A	G	0,5212	0,4788	0,0694	0,7287	0,3138	0,3127	1,000000
16	89664419	rs459920	KIDD	T	C	0,3788	0,6212	0,1799	0,9828	0,4912	0,4920	1,000000
16	90038925	rs3785181	34PLEX	C	T	0,6740	0,3260	0,0264	0,6151	0,1871	0,1914	0,641800

17	1498319	rs1879488	SELDIN	A	C	0,6442	0,3558	0,0310	0,6287	0,2047	0,2140	0,303700
17	5855770	rs9916327	PIMA|SANTOS	G	A	0,5522	0,4478	0,0628	0,7125	0,2982	0,2829	0,274600
17	42506515	rs4411548	KIDD	C	T	0,4869	0,5131	0,0857	0,7680	0,3489	0,3468	1,000000
17	42904228	rs2593595	KIDD	A	G	0,3873	0,6127	0,1234	0,8550	0,4152	0,4561	*0,042390*
17	49404152	rs1049620	SANTOS	T	C	0,4680	0,5320	0,0949	0,7895	0,3667	0,3670	1,000000
17	50648771	rs17642714	KIDD	A	T	0,5452	0,4548	0,0516	0,6840	0,2690	0,2944	0,051990
17	55491523	rs4471745	KIDD	G	A	0,8726	0,1274	0,0039	0,5355	0,0663	0,0677	0,470700
17	55710919	rs2033111	SELDIN	A	G	0,4710	0,5290	0,0829	0,7611	0,3431	0,3625	0,224000
17	61084285	rs2052074	SANTOS	A	G	0,3748	0,6252	0,1161	0,8382	0,4035	0,4659	*0,002447*
17	64991033	rs11652805	SELDIN|KIDD	C	T	0,4002	0,5998	0,1197	0,8465	0,4094	0,4418	0,109400
17	71515958	rs10512572	SELDIN	G	A	0,4866	0,5134	0,0764	0,7456	0,3294	0,3468	0,253000
17	74402197	rs4789659	SANTOS	G	A	0,3921	0,6079	0,1090	0,8221	0,3918	0,4464	*0,005686*
17	74512286	rs1107704	SANTOS	A	G	0,4271	0,5729	0,1007	0,8028	0,3772	0,4091	0,082940
17	75786110	rs2125345	SELDIN	T	C	0,3622	0,6378	0,1594	0,9361	0,4659	0,4975	0,155800
17	77555585	rs2304925	34PLEX	A	C	0,4582	0,5418	0,0980	0,7966	0,3723	0,3774	0,727000
18	9420506	rs4798812	SELDIN	G	A	0,3999	0,6001	0,1433	0,9000	0,4444	0,4520	0,696800
18	22072021	rs4800105	SELDIN	C	T	0,5102	0,4898	0,0737	0,7392	0,3236	0,3233	1,000000
18	37697659	rs2042762	KIDD	T	C	0,7891	0,2109	0,0105	0,5637	0,1131	0,1170	0,436000
18	42908314	rs7226659	KIDD	G	T	0,7944	0,2056	0,0092	0,5588	0,1053	0,1170	*0,038680*
18	52255174	rs7238445	SELDIN	G	A	0,3853	0,6147	0,1197	0,8465	0,4094	0,4567	*0,020260*
18	61665875	rs881728	SELDIN	C	A	0,5896	0,4104	0,0474	0,6732	0,2573	0,2524	0,860700
18	69911695	rs3916235	KIDD	T	C	0,5312	0,4688	0,0502	0,6804	0,2651	0,3091	*0,002383*
18	70200427	rs4891825	SELDIN|KIDD	G	A	0,4621	0,5379	0,0857	0,7680	0,3489	0,3716	0,190300
18	77344328	rs874299	SELDIN	T	C	0,3671	0,6329	0,1534	0,9227	0,4581	0,4896	0,149300
18	77720430	rs1024116	34PLEX	C	T	0,3631	0,6369	0,1655	0,9500	0,4737	0,4998	0,250800
19	2396245	rs7252511	SANTOS	C	T	0,4617	0,5383	0,0907	0,7796	0,3587	0,3726	0,407000
19	4077098	rs7251928	KIDD	A	C	0,4091	0,5909	0,1549	0,9260	0,4600	0,4483	0,622400
19	18805179	rs2313048	SANTOS	A	G	0,5207	0,4793	0,0652	0,7185	0,3041	0,3139	0,482100
19	33161341	rs8113143	SELDIN	C	A	0,3695	0,6305	0,1734	0,9679	0,4834	0,4977	0,534700

19	39738239	rs2074928	SANTOS	C	T	0,6867	0,3133	0,0220	0,6021	0,1696	0,1868	0,053150
19	41889215	rs717225	SANTOS	A	G	0,5621	0,4379	0,0467	0,6715	0,2554	0,2804	0,056880
19	41894377	rs12104228	SANTOS	T	C	0,5755	0,4245	0,0428	0,6611	0,2437	0,2699	*0,032650*
19	41906179	rs2303798	34PLEX	G	A	0,6217	0,3783	0,0316	0,6302	0,2066	0,2357	*0,007991*
19	50916290	rs1701930	PIMA	A	G	0,3849	0,6151	0,1365	0,8845	0,4347	0,4638	0,154200
19	52398652	rs3745099	SELDIN	G	A	0,4242	0,5758	0,1012	0,8041	0,3782	0,4122	0,068240
19	55103555	rs2532060	SELDIN	T	C	0,3598	0,6402	0,1519	0,9194	0,4561	0,4962	0,074920
20	10214785	rs6104567	SELDIN	T	G	0,4497	0,5503	0,1023	0,8066	0,3801	0,3869	0,731900
20	32460579	rs1555133	SANTOS	G	A	0,4084	0,5916	0,1283	0,8660	0,4226	0,4369	0,475700
20	40221002	rs1321333	34PLEX	A	G	0,3973	0,6027	0,1313	0,8727	0,4271	0,4492	0,309200
20	55384376	rs3907047	SELDIN	T	C	0,6955	0,3046	0,0216	0,6007	0,1676	0,1791	0,140800
20	56362487	rs4811693	SANTOS	C	T	0,6356	0,3644	0,0355	0,6413	0,2203	0,2169	0,840800
20	58090617	rs6026036	SANTOS	A	G	0,4088	0,5912	0,1328	0,8763	0,4294	0,4384	0,686100
20	63528151	rs310644	KIDD	T	C	0,4633	0,5367	0,0605	0,7066	0,2924	0,3726	**0,000003**
21	15313279	rs722098	34PLEX	A	G	0,3999	0,6001	0,1351	0,8814	0,4327	0,4483	0,431600
21	16232918	rs1626109	SANTOS	T	C	0,5156	0,4844	0,0538	0,6897	0,2750	0,3228	*0,001412*
21	16338103	rs239031	34PLEX	G	A	0,5934	0,4066	0,0339	0,6368	0,2148	0,2610	*0,000251*
21	16505474	rs2823850	SANTOS	C	T	0,5689	0,4311	0,0373	0,6461	0,2261	0,2817	**0,000029**
21	24300147	rs2572307	34PLEX	A	G	0,6735	0,3265	0,0185	0,5910	0,1540	0,2081	**3,03E-07**
21	31156642	rs2248656	SANTOS	A	G	0,5080	0,4920	0,0848	0,7657	0,3470	0,3256	0,174100
21	36513327	rs2835370	SELDIN	C	T	0,5131	0,4869	0,0685	0,7266	0,3119	0,3209	0,494000
21	40011216	rs2837371	SANTOS	C	A	0,5033	0,4967	0,0767	0,7463	0,3300	0,3301	1,000000
21	41377262	rs379839	PIMA	A	G	0,3774	0,6226	0,1454	0,9047	0,4473	0,4755	0,205400
21	42755486	rs2269160	SANTOS	G	T	0,3666	0,6334	0,1354	0,8820	0,4331	0,4816	*0,025930*
21	44728959	rs2838665	SANTOS	G	A	0,4794	0,5206	0,0801	0,7545	0,3373	0,3540	0,314700
21	44739169	rs717177	SANTOS	T	C	0,4782	0,5218	0,0819	0,7589	0,3411	0,3553	0,384000
22	17593780	rs1296819	SELDIN	C	A	0,4096	0,5904	0,1209	0,8493	0,4113	0,4328	0,262500
22	25954137	rs5997008	34PLEX	C	A,T	0,6327	0,3673	0,0284	0,6211	0,1902	0,2239	*0,002200*
22	31970372	rs4821004	SELDIN	C	T	0,3539	0,6461	0,1476	0,9096	0,4503	0,5000	*0,027140*

22	40819668	rs13054099	SANTOS	T	C	0,4535	0,5465	0,1023	0,8066	0,3801	0,3831	0,908200
22	41301334	rs2024566	KIDD|SANTOS	A	G	0,4101	0,5899	0,1476	0,9096	0,4503	0,4438	0,766400
22	41326860	rs3927	SANTOS	T	C	0,4395	0,5605	0,1101	0,8247	0,3937	0,3993	0,739600
22	41333212	rs8137373	PIMA	G	A	0,4189	0,5811	0,1245	0,8574	0,4168	0,4249	0,677300
22	47440662	rs2040411	34PLEX	G	A	0,4247	0,5753	0,1209	0,8493	0,4113	0,4178	0,751200
22	47812123	rs5768007	SELDIN	C	T	0,5980	0,4020	0,0435	0,6628	0,2456	0,2469	0,858700

Tabela A2. Probabilidade de aderência ao equilíbrio de Hardy-Weinberg para todos os 396 AIMs utilizados no projeto na população de Ribeirão Preto, considerando apenas os 343 indivíduos que foram coletados de maneira randômica. Os valores estatisticamente significantes ao nível de 5% estão indicados em itálico, e aquele que permaneceu significante após a aplicação da correção de Bonferroni para múltiplos testes ($\alpha_{BONFERRONI}$ = 0.05/396 = 0.000126) se encontra em negrito.

CHR	SNP	O(HET)	E(HET)	P-*value*
1	rs11804831	0,41860	0,47020	0,05079
1	rs12135987	0,40670	0,44340	0,13550
1	rs12119556	0,21950	0,21680	1,00000
1	rs2986742	0,34120	0,35360	0,53930
1	rs6541030	0,21510	0,21880	0,80380
1	rs647325	0,44770	0,46420	0,56120
1	rs4908343	0,44870	0,46670	0,48690
1	rs10493066	0,32560	0,33770	0,52280
1	rs1573020	0,15120	0,15920	0,30860
1	rs1325502	0,37790	0,38360	0,77920
1	rs6429644	0,31690	0,31900	0,86650
1	rs3827730	0,43600	0,41240	0,35960
1	rs12130799	0,11920	0,12230	0,64620
1	rs3118378	0,51600	0,47000	0,08475
1	rs12402499	0,17060	0,18000	0,35760
1	rs3737576	0,14240	0,16640	*0,01492*
1	rs7554936	0,44480	0,49900	0,05149
1	rs2814778	0,27620	0,30470	0,10870
1	rs1040404	0,46800	0,49950	0,23700
1	rs10127540	0,45640	0,48990	0,22570
1	rs1407434	0,26740	0,27260	0,69360
1	rs2065160	0,32850	0,34590	0,35060
1	rs4844413	0,34400	0,32490	0,32150
1	rs4951629	0,23550	0,20770	*0,00771*
1	rs726570	0,39230	0,40730	0,50560
1	rs955613	0,37500	0,40870	0,14590
1	rs316873	0,20350	0,21440	0,31700
2	rs896788	0,33140	0,33770	0,74940
2	rs798443	0,45350	0,46880	0,56550
2	rs7421394	0,45350	0,48860	0,18620
2	rs11096686	0,40990	0,43400	0,32030
2	rs1876482	0,17150	0,19430	*0,04484*
2	rs1834619	0,23550	0,24220	0,65420
2	rs4666200	0,45290	0,46340	0,72540
2	rs4670767	0,28780	0,27450	0,43760
2	rs2711302	0,26020	0,29500	*0,04145*
2	rs10173381	0,47230	0,45660	0,55650
2	rs721390	0,38950	0,39720	0,68610

CHR	SNP	O(HET)	E(HET)	P-*value*
2	rs12714093	0,47900	0,47680	1,00000
2	rs2280355	0,34710	0,35360	0,75890
2	rs10202644	0,27410	0,29240	0,26600
2	rs11123869	0,38620	0,37200	0,55750
2	rs1466020	0,40120	0,38070	0,39530
2	rs2084713	0,38600	0,39330	0,78310
2	rs4851839	0,36800	0,41740	*0,03614*
2	rs3754675	0,19590	0,20880	0,29350
2	rs3827760	0,11050	0,12980	*0,01701*
2	rs260714	0,44310	0,45130	0,72090
2	rs260690	0,37210	0,37210	1,00000
2	rs4241135	0,30810	0,32070	0,50030
2	rs17044697	0,31870	0,33410	0,41760
2	rs10205952	0,33140	0,35070	0,35500
2	rs182549	0,30230	0,34420	*0,02779*
2	rs6754311	0,30520	0,34260	0,05736
2	rs10496971	0,18600	0,19650	0,27960
2	rs1863086	0,49850	0,50000	1,00000
2	rs10497191	0,34990	0,40560	*0,01153*
2	rs1991600	0,40640	0,42660	0,37660
2	rs12470024	0,36920	0,38220	0,57210
2	rs2627037	0,27620	0,30830	0,05510
2	rs10179812	0,25660	0,24070	0,26940
2	rs1569175	0,25870	0,23380	0,06164
2	rs4852013	0,31100	0,28990	0,26180
2	rs12471054	0,23840	0,23170	0,81540
2	rs2176046	0,23390	0,21980	0,32660
3	rs10510228	0,22090	0,21440	0,80090
3	rs7432238	0,19760	0,19680	1,00000
3	rs17008327	0,18900	0,18510	1,00000
3	rs4955316	0,23550	0,24640	0,38380
3	rs9809104	0,38950	0,45600	*0,00904*
3	rs3870336	0,25950	0,26730	0,54730
3	rs4955903	0,43310	0,44790	0,54800
3	rs9871910	0,37060	0,37790	0,77380
3	rs6548616	0,48840	0,49760	0,74550
3	rs4855697	0,55520	0,47490	*0,00213*
3	rs12629908	0,21220	0,22090	0,46200
3	rs12498138	0,15990	0,15680	1,00000
3	rs702032	0,35170	0,41840	*0,00434*
3	rs9845457	0,43310	0,49880	*0,01721*
3	rs10935323	0,37500	0,39590	0,34060
3	rs868767	0,21220	0,22950	0,15910
3	rs734873	0,19770	0,21000	0,29890
3	rs1498444	0,43310	0,47090	0,13760
3	rs2030763	0,27330	0,28420	0,45030
3	rs1513181	0,31980	0,34100	0,26740

CHR	SNP	O(HET)	E(HET)	P-*value*
3	rs710493	0,26450	0,30830	*0,01313*
4	rs9291090	0,10470	0,10950	0,31610
4	rs4698702	0,31100	0,30470	0,85940
4	rs10012227	0,31100	0,30470	0,85940
4	rs2874414	0,39070	0,39260	0,89140
4	rs4540055	0,46150	0,48680	0,42270
4	rs4833103	0,36340	0,39850	0,10490
4	rs1876495	0,34590	0,35540	0,64880
4	rs10007810	0,45930	0,46110	1,00000
4	rs1369093	0,26450	0,27060	0,68960
4	rs7660290	0,43440	0,44280	0,71560
4	rs385194	0,46510	0,49940	0,19710
4	rs1229984	0,15990	0,16640	0,50830
4	rs3811801	0,01450	0,02010	*0,03034*
4	rs4532240	0,38950	0,37210	0,46870
4	rs7657799	0,24710	0,25860	0,40400
4	rs2026721	0,25870	0,25460	1,00000
4	rs4431170	0,17730	0,19430	0,15620
4	rs2063393	0,41570	0,45160	0,15180
4	rs2702414	0,26160	0,28040	0,24570
4	rs4647693	0,40350	0,41620	0,60320
5	rs316598	0,49710	0,48810	0,82510
5	rs870347	0,18600	0,21000	*0,04006*
5	rs10473594	0,34880	0,38640	0,07086
5	rs13357733	0,28490	0,32420	*0,02970*
5	rs12522914	0,24000	0,23860	1,00000
5	rs16891982	0,41690	0,49100	*0,00588*
5	rs28777	0,39830	0,45160	*0,03144*
5	rs37369	0,34880	0,35690	0,65200
5	rs6451722	0,42440	0,43290	0,70960
5	rs12657828	0,29940	0,33600	0,05281
5	rs1895218	0,35330	0,36120	0,65210
5	rs10079352	0,45640	0,48990	0,22570
5	rs6880706	0,47370	0,50000	0,33160
5	rs252155	0,52200	0,49280	0,32250
5	rs4145160	0,23550	0,24640	0,38380
5	rs1366220	0,36920	0,42970	*0,01172*
5	rs17053011	0,26160	0,26070	1,00000
5	rs6556352	0,52030	0,49740	0,44800
5	rs1500127	0,27030	0,28230	0,44400
5	rs10454965	0,23170	0,23130	1,00000
5	rs7728435	0,42260	0,44480	0,37340
5	rs2042314	0,29650	0,33100	0,07126
5	rs7722456	0,40990	0,43400	0,32030
5	rs4868237	0,28200	0,27840	1,00000
5	rs6422347	0,34300	0,37210	0,14830
5	rs4701170	0,30230	0,29180	0,58300

CHR	SNP	O(HET)	E(HET)	P-*value*
6	rs1040045	0,47230	0,45310	0,47600
6	rs2504853	0,47970	0,48720	0,82480
6	rs7745461	0,34010	0,35220	0,54030
6	rs2397060	0,35470	0,39720	0,05651
6	rs192655	0,29650	0,32760	0,09780
6	rs11153123	0,42440	0,40990	0,59890
6	rs4896046	0,38530	0,39610	0,58660
6	rs3823159	0,23550	0,29360	*0,00068*
6	rs7753890	0,21800	0,25860	*0,00588*
6	rs4463276	0,45060	0,48720	0,18380
6	rs6570644	0,45350	0,49670	0,10470
6	rs4143683	0,40410	0,41120	0,79300
6	rs1744173	0,26740	0,26860	0,84330
6	rs350289	0,26450	0,26270	1,00000
6	rs10945764	0,37790	0,41950	0,07171
6	rs16888746	0,22120	0,24520	0,07624
6	rs10484813	0,27330	0,29180	0,26530
6	rs4458655	0,32270	0,32240	1,00000
6	rs1871428	0,45770	0,49560	0,15770
7	rs917118	0,46220	0,47490	0,64980
7	rs1550392	0,37790	0,36000	0,45340
7	rs17166119	0,43600	0,41240	0,35960
7	rs731257	0,25580	0,25660	1,00000
7	rs917115	0,45930	0,46730	0,73070
7	rs32314	0,45350	0,43080	0,38140
7	rs12701660	0,45880	0,46180	0,90680
7	rs2330442	0,47090	0,49450	0,38390
7	rs2471552	0,33430	0,33600	0,87360
7	rs4717865	0,15990	0,16160	0,73940
7	rs705308	0,44190	0,48480	0,09682
7	rs978874	0,43020	0,46270	0,20030
7	rs2040502	0,40120	0,45070	*0,04233*
7	rs7803075	0,43900	0,49950	*0,02385*
7	rs6464211	0,39830	0,42070	0,30820
8	rs10108270	0,49130	0,48100	0,73740
8	rs4875427	0,41810	0,41260	0,89580
8	rs17072176	0,29740	0,28480	0,56840
8	rs400404	0,42730	0,44790	0,40030
8	rs11776114	0,40350	0,42090	0,44220
8	rs17807624	0,40990	0,45160	0,09441
8	rs3943253	0,30230	0,29550	0,85490
8	rs1471939	0,41280	0,40740	0,89490
8	rs1462906	0,23260	0,26470	*0,03850*
8	rs7018273	0,25870	0,27060	0,42440
8	rs12544346	0,47970	0,50000	0,45160
8	rs6990312	0,43020	0,40490	0,28760
8	rs896680	0,36340	0,36460	1,00000

CHR	SNP	O(HET)	E(HET)	P-*value*
8	rs17742080	0,22490	0,22210	1,00000
8	rs2196051	0,50870	0,49950	0,82900
8	rs7844723	0,49420	0,49620	1,00000
8	rs4870958	0,30810	0,31010	0,86290
8	rs6993205	0,43900	0,49540	*0,03847*
8	rs2001907	0,17730	0,19430	0,15620
8	rs1871534	0,22970	0,28990	*0,00029*
9	rs1408801	0,37210	0,38640	0,48680
9	rs10962599	0,47670	0,49980	0,38960
9	rs10812316	0,33240	0,38430	*0,01627*
9	rs2073821	0,19770	0,18740	0,55980
9	rs3793451	0,14530	0,13970	0,70720
9	rs10511828	0,19480	0,19880	0,59560
9	rs2306040	0,18020	0,17810	1,00000
9	rs4135211	0,19770	0,19200	0,78110
9	rs4979274	0,29200	0,28170	0,69870
9	rs4877068	0,22740	0,22360	1,00000
9	rs3814134	0,27910	0,29180	0,45810
9	rs10513300	0,16570	0,16160	1,00000
9	rs1412512	0,39530	0,40240	0,78860
10	rs4880510	0,24480	0,24970	0,66390
10	rs4880511	0,23370	0,23730	0,81770
10	rs17143387	0,27030	0,27840	0,56300
10	rs7923994	0,44480	0,43810	0,90200
10	rs912069	0,36340	0,37930	0,47660
10	rs7897550	0,36050	0,36910	0,66190
10	rs6482103	0,24420	0,26470	0,15370
10	rs1978806	0,15120	0,16870	0,09775
10	rs3793791	0,26160	0,26860	0,68680
10	rs201381710	0,48760	0,46500	0,49180
10	rs10821745	0,17730	0,17110	0,75310
10	rs4948418	0,14830	0,14710	1,00000
10	rs10994397	0,15410	0,15680	0,72660
10	rs3998448	0,42900	0,44480	0,52490
10	rs4746136	0,23840	0,24840	0,51230
10	rs7902158	0,43020	0,48160	0,05661
10	rs3814614	0,47080	0,49810	0,32870
10	rs4918664	0,22380	0,23380	0,48550
10	rs12220128	0,25000	0,25710	0,67140
10	rs7917775	0,24050	0,24610	0,65960
10	rs17130385	0,20640	0,22090	0,22100
10	rs734241	0,21800	0,23800	0,11540
10	rs7919248	0,25870	0,26670	0,54630
10	rs6585227	0,28570	0,29990	0,36960
10	rs4918842	0,31400	0,30650	0,72800
10	rs1419138	0,35170	0,39850	*0,03069*
10	rs853577	0,33720	0,37790	*0,04648*

CHR	SNP	O(HET)	E(HET)	P-*value*
10	rs4880436	0,13370	0,13970	0,42600
11	rs10839880	0,49130	0,49930	0,74730
11	rs2403512	0,25070	0,28940	*0,02233*
11	rs1837606	0,47380	0,47490	1,00000
11	rs2946788	0,47090	0,45770	0,63870
11	rs5030240	0,41800	0,40600	0,75820
11	rs3736508	0,13950	0,14470	0,45100
11	rs509360	0,49130	0,48100	0,73740
11	rs174546	0,42860	0,41670	0,69740
11	rs599190	0,36630	0,37210	0,77240
11	rs1535	0,44900	0,42930	0,45110
11	rs4963452	0,30970	0,30080	0,71820
11	rs11227699	0,28490	0,29180	0,71080
11	rs674499	0,34690	0,35920	0,54750
11	rs1638566	0,34590	0,36460	0,37410
11	rs1638567	0,34590	0,36460	0,37410
11	rs174570	0,24710	0,24640	1,00000
11	rs2619206	0,25290	0,25500	0,83250
11	rs2241667	0,38780	0,36230	0,23370
11	rs16921059	0,40230	0,37570	0,24930
11	rs4309121	0,31400	0,29180	0,19670
11	rs1079597	0,25580	0,26070	0,68110
11	rs948028	0,43600	0,43290	1,00000
11	rs4937417	0,45350	0,47020	0,49380
11	rs1561937	0,29360	0,35220	*0,00326*
12	rs2216437	0,29940	0,28230	0,33970
12	rs2416791	0,38950	0,39180	0,89110
12	rs1513056	0,34880	0,36000	0,55070
12	rs10843344	0,40120	0,40990	0,69380
12	rs214678	0,22970	0,22950	1,00000
12	rs2051827	0,22090	0,21880	1,00000
12	rs11612312	0,39240	0,38770	0,89000
12	rs11521	0,37210	0,36610	0,88300
12	rs772262	0,34880	0,38070	0,12030
12	rs773658	0,22970	0,23380	0,65020
12	rs2052386	0,29070	0,27650	0,43630
12	rs10878834	0,29450	0,27910	0,43710
12	rs3217863	0,17700	0,18050	0,76020
12	rs3851621	0,18900	0,20330	0,18460
12	rs1397560	0,34110	0,33670	0,87400
12	rs2070586	0,35170	0,37060	0,38180
12	rs2238151	0,44770	0,48580	0,14960
12	rs671	0,00580	0,01160	*0,00873*
12	rs1503767	0,38950	0,42860	0,10150
13	rs1335873	0,49710	0,49140	0,91260
13	rs1886510	0,48550	0,48320	1,00000
13	rs17359176	0,20060	0,19430	0,78130

CHR	SNP	O(HET)	E(HET)	P-*value*
13	rs9319336	0,25580	0,26070	0,68110
13	rs566996	0,38940	0,35430	0,09049
13	rs7997709	0,28200	0,28610	0,85000
13	rs2065982	0,22380	0,22950	0,63750
13	rs1572018	0,41860	0,44510	0,27630
13	rs2166624	0,43020	0,45770	0,28880
13	rs7326934	0,28490	0,31370	0,08741
13	rs9563982	0,31100	0,29740	0,47100
13	rs1409842	0,44190	0,49970	*0,03137*
13	rs1000973	0,36550	0,38640	0,32750
13	rs9530435	0,39070	0,41790	0,24460
13	rs9522149	0,47970	0,48530	0,82470
14	rs1760921	0,26160	0,29180	0,06330
14	rs11844034	0,43600	0,44510	0,71670
14	rs10141763	0,21800	0,24220	0,07278
14	rs7158003	0,52330	0,48270	0,14620
14	rs1958099	0,38120	0,35600	0,22500
14	rs17119138	0,40410	0,38500	0,40290
14	rs2357442	0,29940	0,33600	0,05281
14	rs17126825	0,26450	0,26670	0,84050
14	rs1950993	0,48260	0,49980	0,51880
14	rs810087	0,43150	0,42930	1,00000
14	rs8021730	0,42440	0,45250	0,28300
14	rs7142344	0,34880	0,40240	*0,01561*
14	rs946918	0,37210	0,39720	0,27660
14	rs200354	0,44190	0,43500	0,80550
14	rs730570	0,38080	0,43190	*0,03334*
14	rs3784230	0,51160	0,50000	0,74630
15	rs11161328	0,44370	0,43610	0,89490
15	rs1800414	0,01450	0,02010	*0,03034*
15	rs1448484	0,23260	0,26070	0,05941
15	rs12913832	0,39530	0,41950	0,30370
15	rs12439433	0,06110	0,06460	0,31630
15	rs2122497	0,34120	0,35990	0,36510
15	rs8030587	0,40120	0,43710	0,13850
15	rs12594483	0,36920	0,40110	0,14030
15	rs12917189	0,41570	0,44410	0,22710
15	rs735480	0,35470	0,40740	*0,01729*
15	rs1426654	0,31400	0,37790	*0,00250*
15	rs11637235	0,45480	0,49420	0,15520
15	rs12594750	0,28490	0,28800	0,85130
15	rs974828	0,36920	0,40110	0,14030
15	rs3884558	0,28780	0,28610	1,00000
15	rs11631047	0,32650	0,41060	*0,00021*
15	rs6494489	0,47430	0,45790	0,55030
15	rs16951105	0,42440	0,40490	0,42540
15	rs2899826	0,36630	0,35070	0,53770

CHR	SNP	O(HET)	E(HET)	P-*value*
15	rs8035124	0,47380	0,46020	0,64000
16	rs4984913	0,47670	0,48050	0,91070
16	rs4141505	0,25580	0,24010	0,36520
16	rs4781011	0,44900	0,44180	0,80790
16	rs4780476	0,36630	0,39980	0,13710
16	rs2269793	0,32560	0,34100	0,42840
16	rs4889490	0,44710	0,46800	0,41800
16	rs881929	0,42980	0,46460	0,16400
16	rs818386	0,27910	0,29550	0,27830
16	rs2966849	0,30230	0,30290	1,00000
16	rs459920	0,51450	0,49540	0,51460
16	rs3785181	0,19190	0,20100	0,41590
17	rs1879488	0,20060	0,19880	1,00000
17	rs9916327	0,32270	0,30470	0,37490
17	rs4411548	0,34010	0,34590	0,75570
17	rs2593595	0,42730	0,46020	0,19750
17	rs1049620	0,38190	0,36530	0,46160
17	rs17642714	0,26450	0,28990	0,13370
17	rs4471745	0,06980	0,07270	0,38780
17	rs2033111	0,33720	0,36000	0,23300
17	rs2052074	0,42730	0,46950	0,10750
17	rs11652805	0,38950	0,44120	*0,03719*
17	rs10512572	0,33140	0,34100	0,63480
17	rs4789659	0,40700	0,45250	0,07333
17	rs1107704	0,37540	0,40450	0,18200
17	rs2125345	0,47670	0,49860	0,44900
17	rs2304925	0,36340	0,38220	0,39660
18	rs4798812	0,42730	0,44790	0,40030
18	rs4800105	0,32560	0,32070	0,86750
18	rs2042762	0,13080	0,12730	1,00000
18	rs7226659	0,09880	0,11460	*0,02875*
18	rs7238445	0,43600	0,46270	0,29480
18	rs881728	0,23550	0,24220	0,65420
18	rs3916235	0,27620	0,29740	0,20270
18	rs4891825	0,34300	0,35690	0,45240
18	rs874299	0,42150	0,48720	*0,01470*
18	rs1024116	0,49130	0,49850	0,82870
19	rs7252511	0,34300	0,36310	0,30030
19	rs7251928	0,46220	0,44600	0,54690
19	rs2313048	0,31980	0,32070	1,00000
19	rs8113143	0,51740	0,49800	0,51640
19	rs2074928	0,16860	0,17350	0,53500
19	rs717225	0,26450	0,27450	0,55390
19	rs12104228	0,24710	0,26270	0,30160
19	rs2303798	0,20350	0,22740	0,05700
19	rs1701930	0,43600	0,45600	0,40970
19	rs3745099	0,37210	0,40240	0,17960

CHR	SNP	O(HET)	E(HET)	P-*value*
19	rs2532060	0,47380	0,49740	0,38670
20	rs6104567	0,35760	0,38500	0,20670
20	rs1555133	0,43030	0,44480	0,54180
20	rs1321333	0,44300	0,45580	0,62240
20	rs3907047	0,16570	0,18510	0,07136
20	rs4811693	0,21220	0,20770	1,00000
20	rs6026036	0,41520	0,42550	0,70280
20	rs310644	0,31100	0,36160	*0,01108*
21	rs722098	0,44480	0,43610	0,80470
21	rs1626109	0,31090	0,32280	0,50220
21	rs239031	0,21280	0,25520	*0,00464*
21	rs2823850	0,25580	0,26860	0,41950
21	rs2572307	0,15410	0,21220	**0,00001**
21	rs2248656	0,33720	0,32420	0,51030
21	rs2835370	0,31100	0,31540	0,73540
21	rs2837371	0,33430	0,33700	0,87210
21	rs379839	0,48700	0,47700	0,81110
21	rs2269160	0,44050	0,48300	0,11380
21	rs2838665	0,34120	0,36600	0,23480
21	rs717177	0,34590	0,36760	0,30360
22	rs1296819	0,40990	0,41840	0,70000
22	rs5997008	0,17780	0,20830	*0,01518*
22	rs4821004	0,45930	0,49980	0,13210
22	rs13054099	0,38370	0,39450	0,58700
22	rs2024566	0,44480	0,44220	1,00000
22	rs3927	0,38120	0,39130	0,67770
22	rs8137373	0,40820	0,41790	0,69820
22	rs2040411	0,39830	0,41120	0,59960
22	rs5768007	0,25870	0,24640	0,50970

Apêndice B

Tabela B. Populações e número de indivíduos analisados pelos projetos 1KGEN e HGDP.

População	Banco de Dados	Subpopulação	*n*
AFR	HGDP	BantuKenya	11
		BantuSouthAfrica	8
		Biaka	22
		Mandenka	22
		Mbuti	13
		San	6
		Yoruba	22
		Total	104
	1KGEN	ESN - Esan in Nigeria	99
		GWD - Gambian in Western Division	113
		LWK - Luhya in Webuye, Kenya	99
		MSL - Mende in Sierra Leone	85
		YRI - Yoruba in Ibadan, Nigeria	108
		Total	504
AMR	HGDP	Colombian	7
		Karitiana	12
		Maya	21
		Pima	13
		Surui	8
		Total	61
EAS	HGDP	Cambodian	9
		Dai	9
		Daur	9
		Han	33
		Hezhen	9
		Japanese	27
		Lahu	8
		Miao	10
		Mongolian	9
		Naxi	8
		NorthernHan	10
		Oroqen	9
		She	10
		Tu	10
		Tujia	9

População	Banco de Dados	Subpopulação	*n*
		Xibo	9
		Yakut	25
		Yi	10
		Total	223
	1KGEN	CDX - Chinese Dai	93
		CHB - Han Chinese, Beijing	103
		CHS - Han Chinese South	105
		JPT - Japanese in Tokyo	104
		KHV - Vietnam	99
		Total	504
EUR	HGDP	Adygei	16
		Basque	23
		BergamoItalian	12
		French	28
		Orcadian	15
		Russian	25
		Sardinian	28
		Tuscan	8
		Total	155
	1KGEN	CEU - Northern Europeans from Utah	99
		FIN - Finnish	99
		GBR - British from England and Scotland	91
		IBS - Iberian Populations in Spain	107
		TSI - Toscani in Italy	107
		Total	503
MISCIGENADOS	1KGEN	ACB - African Caribbean in Barbados	96
		ASW - African Ancestry is SW USA	61
		CLM - Colombian in Medellín	94
		MXL - Mexican Ancestry in Los Angeles	64
		PEL - Peruvian in Lima	85
		PUR - Puerto Rican	104
		Total	504

Apêndice C

Tabela C. Frequências alélicas dos 396 *loci* estudados para cada grupo populacional de referência, populações miscigenadas do 1KGEN e para a população de Ribeirão Preto. Entre parênteses encontram-se os tamanhos amostrais de cada população.

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs11804831	T	0,040	0,033	0,014	0,827	0,151	0,189	0,559	0,406	0,200	0,630	0,612
	C	0,960	0,967	0,986	0,173	0,849	0,811	0,441	0,594	0,800	0,370	0,388
rs12135987	G	0,875	0,049	0,333	0,708	0,849	0,754	0,580	0,422	0,200	0,611	0,672
	A	0,125	0,951	0,667	0,292	0,151	0,246	0,420	0,578	0,800	0,389	0,328
rs12119556	G	0,995	0,262	0,530	0,868	0,969	0,926	0,718	0,617	0,353	0,784	0,875
	A	0,005	0,738	0,470	0,132	0,031	0,074	0,282	0,383	0,647	0,216	0,125
rs2986742	T	0,172	0,943	0,750	0,877	0,250	0,385	0,793	0,836	0,918	0,793	0,765
	C	0,828	0,057	0,250	0,123	0,750	0,615	0,207	0,164	0,082	0,207	0,235
rs6541030	G	0,743	0,566	0,684	0,934	0,781	0,754	0,739	0,742	0,541	0,870	0,877
	A	0,257	0,434	0,316	0,066	0,219	0,246	0,261	0,258	0,459	0,130	0,123
rs647325	A	0,340	0,057	0,509	0,779	0,391	0,443	0,537	0,367	0,247	0,630	0,642
	G	0,660	0,943	0,491	0,221	0,609	0,557	0,463	0,633	0,753	0,370	0,358
rs4908343	G	0,938	0,098	0,386	0,222	0,922	0,795	0,191	0,195	0,124	0,284	0,379
	A	0,062	0,902	0,614	0,778	0,078	0,205	0,809	0,805	0,876	0,716	0,621
rs10493066	C	0,359	0,238	0,335	0,972	0,328	0,492	0,707	0,594	0,300	0,755	0,778
	T	0,641	0,762	0,665	0,028	0,672	0,508	0,293	0,406	0,700	0,245	0,222
rs1573020	G	0,511	0,000	0,044	0,001	0,464	0,402	0,053	0,031	0,035	0,082	0,091
	A	0,489	1,000	0,956	0,999	0,536	0,598	0,947	0,969	0,965	0,918	0,909
rs1325502	G	0,246	0,932	0,588	0,841	0,276	0,303	0,878	0,867	0,894	0,822	0,737
	A	0,754	0,068	0,412	0,159	0,724	0,697	0,122	0,133	0,106	0,178	0,263
rs6429644	T	0,117	0,967	0,903	0,919	0,208	0,295	0,888	0,844	0,918	0,813	0,788
	C	0,883	0,033	0,097	0,081	0,792	0,705	0,112	0,156	0,082	0,188	0,212
rs3827730	T	0,933	0,033	0,215	0,697	0,911	0,893	0,527	0,430	0,212	0,673	0,720
	C	0,067	0,967	0,785	0,303	0,089	0,107	0,473	0,570	0,788	0,327	0,280
rs12130799	A	1,000	0,598	0,998	0,932	0,995	0,975	0,809	0,773	0,541	0,885	0,928
	G	0,000	0,402	0,002	0,068	0,005	0,025	0,191	0,227	0,459	0,115	0,072
rs3118378	A	0,545	0,074	0,479	0,693	0,589	0,549	0,516	0,336	0,182	0,500	0,615
	G	0,455	0,926	0,521	0,307	0,411	0,451	0,484	0,664	0,818	0,500	0,385
rs12402499	G	1,000	0,270	1,000	0,929	0,984	0,943	0,761	0,711	0,388	0,865	0,911
	A	0,000	0,730	0,000	0,071	0,016	0,057	0,239	0,289	0,612	0,135	0,089
rs3737576	T	0,988	0,270	0,916	0,956	0,979	0,967	0,787	0,633	0,306	0,899	0,917
	C	0,012	0,730	0,084	0,044	0,021	0,033	0,213	0,367	0,694	0,101	0,083
rs7554936	C	0,985	0,156	0,139	0,333	0,922	0,910	0,394	0,211	0,259	0,370	0,471
	T	0,015	0,844	0,861	0,667	0,078	0,090	0,606	0,789	0,741	0,630	0,529
rs2814778	T	0,012	0,984	1,000	0,993	0,115	0,205	0,926	0,969	0,959	0,861	0,807
	C	0,988	0,016	0,000	0,007	0,885	0,795	0,074	0,031	0,041	0,139	0,193
rs1040404	G	0,129	0,180	0,577	0,697	0,156	0,213	0,410	0,383	0,218	0,447	0,503

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	A	0,871	0,820	0,423	0,303	0,844	0,787	0,590	0,617	0,782	0,553	0,497
rs10127540	T	0,016	1,000	0,867	0,669	0,047	0,189	0,676	0,781	0,859	0,630	0,576
	C	0,984	0,000	0,133	0,331	0,953	0,811	0,324	0,219	0,141	0,370	0,424
rs1407434	G	0,764	0,167	0,493	0,921	0,745	0,811	0,734	0,531	0,376	0,827	0,828
	A	0,236	0,833	0,507	0,079	0,255	0,189	0,266	0,469	0,624	0,173	0,172
rs2065160	A	0,527	0,090	0,229	0,917	0,552	0,615	0,654	0,508	0,271	0,740	0,772
	G	0,473	0,910	0,771	0,083	0,448	0,385	0,346	0,492	0,729	0,260	0,228
rs4844413	C	0,854	0,041	0,302	0,869	0,854	0,836	0,601	0,508	0,224	0,750	0,790
	T	0,146	0,959	0,698	0,131	0,146	0,164	0,399	0,492	0,776	0,250	0,210
rs4951629	T	0,839	0,434	0,765	0,963	0,854	0,828	0,782	0,727	0,547	0,832	0,878
	C	0,161	0,566	0,235	0,037	0,146	0,172	0,218	0,273	0,453	0,168	0,122
rs726570	T	0,080	0,959	0,610	0,249	0,156	0,148	0,468	0,594	0,835	0,394	0,305
	C	0,920	0,041	0,390	0,751	0,844	0,852	0,532	0,406	0,165	0,606	0,695
rs955613	T	0,067	0,959	0,691	0,251	0,135	0,148	0,463	0,602	0,824	0,413	0,309
	C	0,933	0,041	0,309	0,749	0,865	0,852	0,537	0,398	0,176	0,587	0,691
rs316873	C	0,909	0,393	0,561	0,911	0,880	0,852	0,750	0,656	0,582	0,846	0,877
	T	0,091	0,607	0,439	0,089	0,120	0,148	0,250	0,344	0,418	0,154	0,123
rs896788	C	0,674	0,418	0,367	0,829	0,667	0,680	0,771	0,625	0,482	0,712	0,777
	T	0,326	0,582	0,633	0,171	0,333	0,320	0,229	0,375	0,518	0,288	0,223
rs798443	G	0,949	0,213	0,375	0,204	0,922	0,820	0,303	0,281	0,282	0,404	0,390
	A	0,051	0,787	0,625	0,796	0,078	0,180	0,697	0,719	0,718	0,596	0,610
rs7421394	A	0,033	0,418	0,522	0,716	0,068	0,164	0,574	0,531	0,394	0,553	0,575
	G	0,967	0,582	0,478	0,284	0,932	0,836	0,426	0,469	0,606	0,447	0,425
rs11096686	T	0,229	0,139	0,376	0,899	0,229	0,336	0,686	0,602	0,459	0,620	0,691
	C	0,771	0,861	0,624	0,101	0,771	0,664	0,314	0,398	0,541	0,380	0,309
rs1876482	G	0,999	0,533	0,225	0,944	0,990	1,000	0,835	0,797	0,547	0,875	0,896
	A	0,001	0,467	0,775	0,056	0,010	0,000	0,165	0,203	0,453	0,125	0,104
rs1834619	G	1,000	0,049	0,291	0,952	0,984	0,959	0,729	0,516	0,276	0,803	0,870
	A	0,000	0,951	0,709	0,048	0,016	0,041	0,271	0,484	0,724	0,197	0,130
rs4666200	G	0,909	0,090	0,207	0,281	0,807	0,754	0,293	0,227	0,118	0,370	0,371
	A	0,091	0,910	0,793	0,719	0,193	0,246	0,707	0,773	0,882	0,630	0,629
rs4670767	G	0,969	0,567	0,424	0,896	0,958	0,934	0,681	0,586	0,518	0,841	0,834
	T	0,031	0,433	0,576	0,104	0,042	0,066	0,319	0,414	0,482	0,159	0,166
rs2711302	A	0,981	0,225	0,455	0,865	0,969	0,918	0,771	0,469	0,376	0,832	0,825
	G	0,019	0,775	0,545	0,135	0,031	0,082	0,229	0,531	0,624	0,168	0,175
rs10173381	A	0,017	0,992	0,990	0,796	0,099	0,180	0,766	0,828	0,906	0,750	0,652
	G	0,983	0,008	0,010	0,204	0,901	0,820	0,234	0,172	0,094	0,250	0,348
rs721390	G	0,922	0,098	0,413	0,746	0,917	0,885	0,617	0,430	0,259	0,654	0,727
	T	0,078	0,902	0,587	0,254	0,083	0,115	0,383	0,570	0,741	0,346	0,273
rs12714093	G	0,974	0,008	0,275	0,605	0,911	0,902	0,479	0,305	0,129	0,514	0,605
	A	0,026	0,992	0,725	0,395	0,089	0,098	0,521	0,695	0,871	0,486	0,395
rs2280355	A	0,021	0,897	0,506	0,215	0,031	0,041	0,388	0,492	0,618	0,332	0,231
	G	0,979	0,103	0,494	0,785	0,969	0,959	0,612	0,508	0,382	0,668	0,769

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs10202644	T	0,078	0,885	0,569	0,154	0,052	0,090	0,335	0,469	0,647	0,255	0,183
	C	0,922	0,115	0,431	0,846	0,948	0,910	0,665	0,531	0,353	0,745	0,817
rs11123869	A	0,031	0,967	0,600	0,243	0,042	0,057	0,426	0,578	0,712	0,250	0,237
	G	0,969	0,033	0,400	0,757	0,958	0,943	0,574	0,422	0,288	0,750	0,763
rs1466020	G	0,030	0,959	0,677	0,235	0,036	0,049	0,356	0,563	0,653	0,245	0,251
	A	0,970	0,041	0,323	0,765	0,964	0,951	0,644	0,438	0,347	0,755	0,749
rs2084713	G	0,992	0,058	0,485	0,697	0,958	0,951	0,596	0,383	0,318	0,688	0,725
	A	0,008	0,942	0,515	0,303	0,042	0,049	0,404	0,617	0,682	0,313	0,275
rs4851839	G	0,940	0,033	0,381	0,676	0,917	0,934	0,543	0,367	0,200	0,663	0,703
	A	0,057	0,967	0,619	0,324	0,083	0,066	0,457	0,633	0,800	0,337	0,297
	T	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
rs3754675	T	0,970	0,082	0,653	0,926	0,964	0,918	0,777	0,516	0,229	0,793	0,888
	C	0,030	0,918	0,347	0,074	0,036	0,082	0,223	0,484	0,771	0,207	0,112
rs3827760	A	1,000	0,107	0,137	0,999	1,000	0,967	0,761	0,516	0,241	0,827	0,927
	G	0,000	0,893	0,863	0,001	0,000	0,033	0,239	0,484	0,759	0,173	0,073
rs260714	C	0,192	0,125	0,054	0,866	0,245	0,287	0,601	0,422	0,182	0,625	0,663
	T	0,808	0,875	0,946	0,134	0,755	0,713	0,399	0,578	0,818	0,375	0,337
rs260690	C	0,618	0,877	0,931	0,057	0,625	0,631	0,314	0,516	0,771	0,317	0,241
	A	0,382	0,123	0,069	0,943	0,375	0,369	0,686	0,484	0,229	0,683	0,759
rs4241135	A	0,913	0,148	0,622	0,795	0,880	0,844	0,734	0,539	0,306	0,683	0,788
	G	0,087	0,852	0,378	0,205	0,120	0,156	0,266	0,461	0,694	0,317	0,212
rs17044697	T	0,907	0,156	0,613	0,793	0,880	0,844	0,734	0,547	0,282	0,702	0,777
	C	0,093	0,844	0,387	0,207	0,120	0,156	0,266	0,453	0,718	0,298	0,223
rs10205952	T	0,919	0,164	0,623	0,788	0,891	0,869	0,734	0,547	0,300	0,692	0,767
	G	0,081	0,836	0,377	0,212	0,109	0,131	0,266	0,453	0,700	0,308	0,233
rs182549	C	0,998	0,992	0,998	0,541	0,932	0,828	0,691	0,750	0,888	0,793	0,783
	T	0,002	0,008	0,002	0,459	0,068	0,172	0,309	0,250	0,112	0,207	0,217
rs6754311	T	0,000	0,008	0,002	0,459	0,068	0,172	0,309	0,250	0,112	0,207	0,215
	C	1,000	0,992	0,998	0,541	0,932	0,828	0,691	0,750	0,888	0,793	0,785
rs10496971	T	0,930	0,443	0,215	0,937	0,932	0,943	0,771	0,680	0,600	0,851	0,879
	G	0,070	0,557	0,785	0,063	0,068	0,057	0,229	0,320	0,400	0,149	0,121
rs1863086	G	0,324	0,934	0,043	0,510	0,391	0,525	0,553	0,656	0,759	0,577	0,484
	T	0,676	0,066	0,957	0,490	0,609	0,475	0,447	0,344	0,241	0,423	0,516
rs10497191	T	0,953	0,074	0,090	0,128	0,844	0,803	0,154	0,156	0,129	0,279	0,285
	C	0,047	0,926	0,910	0,872	0,156	0,197	0,846	0,844	0,871	0,721	0,715
rs1991600	C	0,892	0,016	0,335	0,752	0,880	0,861	0,516	0,422	0,212	0,673	0,701
	T	0,108	0,984	0,665	0,248	0,120	0,139	0,484	0,578	0,788	0,327	0,299
rs12470024	G	0,984	0,016	0,330	0,781	0,969	0,910	0,564	0,414	0,224	0,712	0,756
	A	0,016	0,984	0,670	0,219	0,031	0,090	0,436	0,586	0,776	0,288	0,244
rs2627037	G	0,511	0,369	0,788	0,924	0,542	0,557	0,718	0,609	0,353	0,822	0,801
	A	0,489	0,631	0,212	0,076	0,458	0,443	0,282	0,391	0,647	0,178	0,199
rs10179812	T	0,953	0,230	0,803	0,913	0,938	0,926	0,707	0,555	0,306	0,832	0,875
	G	0,047	0,770	0,197	0,087	0,063	0,074	0,293	0,445	0,694	0,168	0,125

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs1569175	T	0,217	0,730	0,155	0,048	0,177	0,180	0,282	0,297	0,494	0,192	0,135
	C	0,783	0,270	0,845	0,952	0,823	0,820	0,718	0,703	0,506	0,808	0,866
rs4852013	G	0,961	0,230	0,692	0,809	0,953	0,910	0,681	0,555	0,365	0,813	0,831
	A	0,039	0,770	0,308	0,191	0,047	0,090	0,319	0,445	0,635	0,188	0,169
rs12471054	C	0,960	0,115	0,341	0,922	0,964	0,918	0,739	0,602	0,259	0,817	0,862
	T	0,040	0,885	0,659	0,078	0,036	0,082	0,261	0,398	0,741	0,183	0,138
rs2176046	G	0,985	0,115	0,376	0,923	0,964	0,926	0,734	0,602	0,282	0,822	0,870
	A	0,015	0,885	0,624	0,077	0,036	0,074	0,266	0,398	0,718	0,178	0,130
rs10510228	A	0,889	0,443	0,684	0,934	0,880	0,852	0,713	0,633	0,553	0,832	0,878
	G	0,111	0,557	0,316	0,066	0,120	0,148	0,287	0,367	0,447	0,168	0,122
rs7432238	T	0,978	0,164	0,469	0,933	0,974	0,943	0,713	0,500	0,265	0,798	0,881
	G	0,022	0,836	0,531	0,067	0,026	0,057	0,287	0,500	0,735	0,202	0,119
rs17008327	G	0,940	0,117	0,736	0,950	0,964	0,918	0,697	0,477	0,265	0,808	0,894
	A	0,060	0,883	0,264	0,050	0,036	0,082	0,303	0,523	0,735	0,192	0,106
rs4955316	T	0,813	0,320	0,764	0,922	0,813	0,787	0,750	0,711	0,482	0,822	0,865
	G	0,188	0,680	0,236	0,078	0,188	0,213	0,250	0,289	0,518	0,178	0,135
rs9809104	T	0,093	0,689	0,693	0,766	0,115	0,254	0,798	0,695	0,647	0,654	0,656
	C	0,907	0,311	0,307	0,234	0,885	0,746	0,202	0,305	0,353	0,346	0,344
rs3870336	G	0,908	0,117	0,240	0,917	0,880	0,852	0,686	0,508	0,365	0,803	0,851
	A	0,092	0,883	0,760	0,083	0,120	0,148	0,314	0,492	0,635	0,197	0,149
rs4955903	T	0,963	0,049	0,307	0,663	0,964	0,885	0,500	0,344	0,159	0,635	0,663
	C	0,037	0,951	0,693	0,337	0,036	0,115	0,500	0,656	0,841	0,365	0,337
rs9871910	G	0,096	0,958	0,585	0,206	0,125	0,139	0,404	0,516	0,682	0,356	0,261
	A	0,904	0,042	0,415	0,794	0,875	0,861	0,596	0,484	0,318	0,644	0,739
	T	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
rs6548616	T	0,058	0,951	0,831	0,659	0,089	0,213	0,761	0,813	0,871	0,572	0,549
	C	0,942	0,049	0,169	0,341	0,911	0,787	0,239	0,188	0,129	0,428	0,451
rs4855697	A	0,900	0,025	0,152	0,593	0,828	0,820	0,436	0,367	0,124	0,587	0,619
	G	0,100	0,975	0,848	0,407	0,172	0,180	0,564	0,633	0,876	0,413	0,381
rs12629908	G	0,905	0,180	0,605	0,936	0,885	0,861	0,745	0,492	0,388	0,808	0,874
	A	0,095	0,820	0,395	0,064	0,115	0,139	0,255	0,508	0,612	0,192	0,126
rs12498138	G	0,998	0,090	0,913	0,936	0,990	0,959	0,761	0,500	0,353	0,846	0,909
	A	0,002	0,910	0,087	0,064	0,010	0,041	0,239	0,500	0,647	0,154	0,091
rs702032	C	0,142	0,262	0,381	0,922	0,177	0,262	0,761	0,578	0,476	0,726	0,705
	T	0,858	0,738	0,619	0,078	0,823	0,738	0,239	0,422	0,524	0,274	0,296
rs9845457	G	0,972	0,100	0,203	0,398	0,927	0,795	0,282	0,344	0,176	0,452	0,481
	A	0,028	0,900	0,797	0,602	0,073	0,205	0,718	0,656	0,824	0,548	0,519
rs10935323	T	0,932	0,049	0,139	0,772	0,917	0,861	0,553	0,516	0,265	0,625	0,748
	G	0,068	0,951	0,861	0,228	0,083	0,139	0,447	0,484	0,735	0,375	0,252
rs868767	A	0,975	0,167	0,263	0,899	0,953	0,926	0,739	0,648	0,341	0,808	0,845
	G	0,025	0,833	0,737	0,101	0,047	0,074	0,261	0,352	0,659	0,192	0,155
rs734873	G	0,987	0,344	0,574	0,892	0,984	0,975	0,739	0,719	0,518	0,851	0,878
	A	0,013	0,656	0,426	0,108	0,016	0,025	0,261	0,281	0,482	0,149	0,122

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs1498444	T	0,935	0,746	0,754	0,532	0,922	0,836	0,660	0,586	0,753	0,678	0,622
	G	0,065	0,254	0,246	0,468	0,078	0,164	0,340	0,414	0,247	0,322	0,378
rs2030763	G	0,979	0,270	0,494	0,837	0,974	0,934	0,670	0,563	0,382	0,779	0,829
	A	0,021	0,730	0,507	0,163	0,026	0,066	0,330	0,438	0,618	0,221	0,171
rs1513181	G	0,831	0,098	0,188	0,859	0,839	0,787	0,633	0,547	0,247	0,712	0,773
	A	0,169	0,902	0,812	0,141	0,161	0,213	0,367	0,453	0,753	0,288	0,227
rs710493	G	0,877	0,016	0,188	0,021	0,786	0,656	0,138	0,063	0,065	0,154	0,187
	A	0,123	0,984	0,812	0,979	0,214	0,344	0,862	0,938	0,935	0,846	0,813
rs9291090	A	0,961	0,582	0,900	0,962	0,953	0,951	0,851	0,766	0,553	0,870	0,940
	C	0,039	0,418	0,100	0,038	0,047	0,049	0,149	0,234	0,447	0,130	0,060
rs4698702	C	0,902	0,041	0,575	0,853	0,823	0,861	0,633	0,391	0,312	0,736	0,812
	A	0,098	0,959	0,425	0,147	0,177	0,139	0,367	0,609	0,688	0,264	0,188
rs10012227	G	0,904	0,041	0,576	0,854	0,823	0,861	0,633	0,391	0,312	0,736	0,812
	A	0,096	0,959	0,424	0,146	0,177	0,139	0,367	0,609	0,688	0,264	0,188
rs2874414	A	0,788	0,041	0,467	0,762	0,729	0,762	0,527	0,359	0,282	0,635	0,730
	G	0,212	0,959	0,533	0,238	0,271	0,238	0,473	0,641	0,718	0,365	0,270
rs4540055	A	0,055	0,574	0,377	0,675	0,125	0,197	0,564	0,492	0,494	0,529	0,499
	T	0,387	0,426	0,533	0,282	0,370	0,336	0,351	0,445	0,459	0,337	0,330
	C	0,558	0,000	0,090	0,043	0,505	0,467	0,085	0,063	0,047	0,135	0,171
rs4833103	A	0,000	0,016	0,001	0,460	0,042	0,131	0,266	0,164	0,094	0,274	0,298
	C	1,000	0,984	0,999	0,540	0,958	0,869	0,734	0,836	0,906	0,726	0,702
rs1876495	C	0,979	0,139	0,390	0,805	0,953	0,934	0,718	0,484	0,424	0,745	0,777
	T	0,021	0,861	0,610	0,195	0,047	0,066	0,282	0,516	0,576	0,255	0,223
rs10007810	G	0,013	0,730	0,893	0,787	0,109	0,189	0,766	0,734	0,759	0,716	0,635
	A	0,987	0,270	0,108	0,213	0,891	0,811	0,234	0,266	0,241	0,284	0,365
rs1369093	T	0,592	0,237	0,739	0,974	0,677	0,623	0,750	0,602	0,406	0,856	0,848
	C	0,408	0,763	0,261	0,026	0,323	0,377	0,250	0,398	0,594	0,144	0,152
rs7660290	G	0,903	0,893	0,610	0,160	0,807	0,803	0,457	0,586	0,659	0,346	0,330
	T	0,097	0,107	0,390	0,840	0,193	0,197	0,543	0,414	0,341	0,654	0,670
rs385194	A	0,983	0,828	0,578	0,329	0,896	0,893	0,559	0,625	0,700	0,529	0,498
	G	0,017	0,172	0,422	0,671	0,104	0,107	0,441	0,375	0,300	0,471	0,502
rs1229984	T	0,000	0,016	0,665	0,045	0,010	0,000	0,074	0,086	0,012	0,063	0,097
	C	1,000	0,984	0,335	0,955	0,990	1,000	0,926	0,914	0,988	0,938	0,903
rs3811801	G	1,000	1,000	0,512	0,997	1,000	1,000	1,000	1,000	1,000	1,000	0,982
	A	0,000	0,000	0,488	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,018
rs4532240	G	0,942	0,156	0,484	0,788	0,885	0,918	0,638	0,500	0,347	0,712	0,756
	T	0,058	0,844	0,516	0,212	0,115	0,082	0,362	0,500	0,653	0,288	0,244
rs7657799	T	0,170	0,984	0,832	0,978	0,354	0,352	0,947	0,938	0,947	0,942	0,847
	G	0,830	0,016	0,168	0,022	0,646	0,648	0,053	0,063	0,053	0,058	0,153
rs2026721	T	0,580	0,008	0,005	0,057	0,526	0,393	0,059	0,117	0,024	0,130	0,149
	C	0,420	0,992	0,995	0,943	0,474	0,607	0,941	0,883	0,976	0,870	0,851
rs4431170	A	0,943	0,279	0,731	0,957	0,953	0,918	0,734	0,625	0,459	0,870	0,898
	G	0,057	0,721	0,269	0,043	0,047	0,082	0,266	0,375	0,541	0,130	0,102

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs2063393	T	0,924	0,082	0,541	0,676	0,854	0,869	0,564	0,344	0,159	0,644	0,666
	C	0,076	0,918	0,459	0,324	0,146	0,131	0,436	0,656	0,841	0,356	0,334
rs2702414	G	0,961	0,270	0,606	0,907	0,901	0,910	0,713	0,570	0,341	0,870	0,833
	A	0,039	0,730	0,394	0,093	0,099	0,090	0,287	0,430	0,659	0,130	0,167
rs4647693	T	0,800	0,041	0,183	0,743	0,849	0,795	0,537	0,406	0,218	0,731	0,707
	C	0,200	0,959	0,817	0,257	0,151	0,205	0,463	0,594	0,782	0,269	0,293
rs316598	T	0,011	0,680	0,418	0,724	0,083	0,172	0,644	0,523	0,524	0,620	0,569
	C	0,989	0,320	0,582	0,276	0,917	0,828	0,356	0,477	0,476	0,380	0,431
rs870347	A	0,937	0,230	0,534	0,942	0,943	0,902	0,798	0,609	0,259	0,856	0,877
	C	0,063	0,770	0,466	0,058	0,057	0,098	0,202	0,391	0,741	0,144	0,123
rs10473594	T	0,954	0,148	0,708	0,797	0,953	0,934	0,633	0,492	0,200	0,654	0,747
	C	0,046	0,852	0,292	0,203	0,047	0,066	0,367	0,508	0,800	0,346	0,253
rs13357733	G	0,927	0,148	0,633	0,875	0,927	0,926	0,718	0,578	0,218	0,707	0,810
	A	0,073	0,852	0,367	0,125	0,073	0,074	0,282	0,422	0,782	0,293	0,190
rs12522914	C	0,947	0,180	0,549	0,932	0,932	0,893	0,771	0,586	0,329	0,875	0,873
	T	0,053	0,820	0,451	0,068	0,068	0,107	0,229	0,414	0,671	0,125	0,127
rs16891982	C	0,998	0,951	0,992	0,085	0,911	0,787	0,362	0,586	0,841	0,413	0,415
	G	0,002	0,049	0,008	0,915	0,089	0,213	0,638	0,414	0,159	0,587	0,585
rs28777	C	0,845	0,943	0,884	0,065	0,708	0,656	0,346	0,578	0,818	0,356	0,337
	A	0,155	0,057	0,116	0,935	0,292	0,344	0,654	0,422	0,182	0,644	0,663
rs37369	C	0,363	0,352	0,423	0,912	0,401	0,557	0,771	0,688	0,435	0,779	0,773
	T	0,637	0,648	0,577	0,088	0,599	0,443	0,229	0,313	0,565	0,221	0,227
rs6451722	G	0,093	0,984	0,782	0,768	0,151	0,311	0,766	0,875	0,929	0,644	0,670
	A	0,907	0,016	0,218	0,232	0,849	0,689	0,234	0,125	0,071	0,356	0,330
rs12657828	A	0,705	0,254	0,810	0,900	0,797	0,689	0,697	0,594	0,412	0,764	0,790
	G	0,295	0,746	0,190	0,100	0,203	0,311	0,303	0,406	0,588	0,236	0,210
rs1895218	C	0,996	0,115	0,410	0,811	0,974	0,910	0,622	0,484	0,265	0,788	0,776
	A	0,001	0,885	0,590	0,189	0,026	0,090	0,378	0,516	0,735	0,212	0,224
	T	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
rs10079352	G	0,020	0,984	0,976	0,431	0,068	0,205	0,553	0,656	0,859	0,466	0,409
	A	0,980	0,016	0,024	0,569	0,932	0,795	0,447	0,344	0,141	0,534	0,591
rs6880706	A	0,011	0,992	0,965	0,557	0,073	0,221	0,660	0,727	0,888	0,505	0,497
	G	0,989	0,008	0,035	0,443	0,927	0,779	0,340	0,273	0,112	0,495	0,503
rs252155	T	0,012	0,967	0,296	0,479	0,068	0,082	0,585	0,664	0,788	0,423	0,430
	C	0,988	0,033	0,704	0,521	0,932	0,918	0,415	0,336	0,212	0,577	0,570
rs4145160	G	0,892	0,082	0,403	0,933	0,938	0,885	0,723	0,508	0,324	0,889	0,867
	A	0,108	0,918	0,597	0,067	0,063	0,115	0,277	0,492	0,676	0,111	0,133
rs1366220	G	0,949	0,057	0,100	0,701	0,927	0,910	0,585	0,422	0,194	0,567	0,677
	A	0,051	0,943	0,900	0,299	0,073	0,090	0,415	0,578	0,806	0,433	0,323
rs17053011	G	0,995	0,098	0,776	0,863	0,984	0,959	0,681	0,492	0,282	0,793	0,850
	T	0,005	0,902	0,224	0,137	0,016	0,041	0,319	0,508	0,718	0,207	0,150
rs6556352	C	0,016	0,361	0,428	0,663	0,083	0,082	0,521	0,602	0,553	0,563	0,557
	T	0,984	0,639	0,572	0,337	0,917	0,918	0,479	0,398	0,447	0,438	0,443

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs1500127	C	0,502	0,467	0,662	0,939	0,505	0,508	0,702	0,758	0,512	0,750	0,820
	T	0,498	0,533	0,338	0,061	0,495	0,492	0,298	0,242	0,488	0,250	0,180
rs10454965	A	0,965	0,082	0,419	0,905	0,979	0,951	0,670	0,586	0,329	0,837	0,867
	G	0,035	0,918	0,581	0,095	0,021	0,049	0,330	0,414	0,671	0,163	0,133
rs7728435	T	0,928	0,033	0,449	0,654	0,891	0,852	0,452	0,453	0,247	0,649	0,660
	C	0,072	0,967	0,551	0,346	0,109	0,148	0,548	0,547	0,753	0,351	0,340
rs2042314	C	0,874	0,033	0,267	0,873	0,901	0,803	0,569	0,430	0,259	0,692	0,792
	T	0,126	0,967	0,733	0,127	0,099	0,197	0,431	0,570	0,741	0,308	0,208
rs7722456	C	0,343	0,008	0,028	0,221	0,313	0,311	0,160	0,117	0,065	0,212	0,320
	T	0,657	0,992	0,972	0,779	0,688	0,689	0,840	0,883	0,935	0,788	0,680
rs4868237	C	0,982	0,164	0,281	0,887	0,984	0,967	0,633	0,539	0,282	0,793	0,834
	T	0,018	0,836	0,719	0,113	0,016	0,033	0,367	0,461	0,718	0,207	0,166
rs6422347	T	0,112	0,967	0,624	0,908	0,203	0,287	0,798	0,859	0,865	0,784	0,751
	C	0,888	0,033	0,376	0,092	0,797	0,713	0,202	0,141	0,135	0,216	0,249
rs4701170	G	0,948	0,049	0,516	0,845	0,938	0,910	0,654	0,445	0,235	0,702	0,819
	A	0,052	0,951	0,484	0,155	0,063	0,090	0,346	0,555	0,765	0,298	0,181
rs1040045	G	0,881	0,033	0,068	0,280	0,818	0,672	0,261	0,172	0,076	0,274	0,358
	A	0,119	0,967	0,932	0,720	0,182	0,328	0,739	0,828	0,924	0,726	0,642
rs2504853	T	0,040	0,500	0,724	0,700	0,099	0,197	0,537	0,594	0,412	0,615	0,585
	C	0,960	0,500	0,276	0,300	0,901	0,803	0,463	0,406	0,588	0,385	0,415
rs7745461	A	0,362	0,557	0,756	0,131	0,313	0,328	0,282	0,469	0,512	0,207	0,227
	G	0,638	0,443	0,244	0,869	0,688	0,672	0,718	0,531	0,488	0,793	0,773
rs2397060	T	0,189	0,926	0,812	0,851	0,276	0,361	0,846	0,867	0,871	0,697	0,726
	C	0,811	0,074	0,188	0,149	0,724	0,639	0,154	0,133	0,129	0,303	0,274
rs192655	G	0,391	0,697	0,660	0,090	0,380	0,361	0,282	0,344	0,641	0,303	0,202
	A	0,609	0,303	0,340	0,910	0,620	0,639	0,718	0,656	0,359	0,697	0,798
rs11153123	A	0,919	0,090	0,684	0,730	0,911	0,828	0,532	0,406	0,294	0,567	0,726
	G	0,081	0,910	0,316	0,270	0,089	0,172	0,468	0,594	0,706	0,433	0,274
rs4896046	C	0,934	0,049	0,387	0,788	0,917	0,852	0,548	0,461	0,253	0,668	0,750
	T	0,066	0,951	0,613	0,212	0,083	0,148	0,452	0,539	0,747	0,332	0,250
rs3823159	A	0,109	0,566	0,885	0,992	0,245	0,270	0,787	0,727	0,641	0,774	0,801
	G	0,891	0,434	0,115	0,008	0,755	0,730	0,213	0,273	0,359	0,226	0,199
rs7753890	T	0,085	0,984	0,847	0,980	0,208	0,311	0,888	0,906	0,947	0,803	0,828
	C	0,915	0,016	0,153	0,020	0,792	0,689	0,112	0,094	0,053	0,197	0,172
rs4463276	G	0,066	0,025	0,129	0,765	0,156	0,189	0,468	0,328	0,200	0,466	0,552
	A	0,934	0,975	0,871	0,235	0,844	0,811	0,532	0,672	0,800	0,534	0,448
rs6570644	A	0,020	0,025	0,130	0,738	0,083	0,156	0,468	0,328	0,182	0,466	0,514
	G	0,980	0,975	0,870	0,262	0,917	0,844	0,532	0,672	0,818	0,534	0,486
rs4143683	T	0,877	0,057	0,499	0,761	0,875	0,795	0,559	0,438	0,212	0,659	0,732
	C	0,123	0,943	0,501	0,239	0,125	0,205	0,441	0,563	0,788	0,341	0,268
rs1744173	A	0,014	0,918	0,408	0,124	0,026	0,033	0,298	0,469	0,712	0,240	0,161
	G	0,986	0,082	0,592	0,876	0,974	0,967	0,702	0,531	0,288	0,760	0,839
rs350289	A	0,017	0,918	0,397	0,110	0,021	0,049	0,293	0,469	0,706	0,236	0,155

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	G	0,983	0,082	0,603	0,890	0,979	0,951	0,707	0,531	0,294	0,764	0,845
rs10945764	A	0,007	0,959	0,608	0,318	0,089	0,148	0,457	0,578	0,806	0,370	0,300
	G	0,993	0,041	0,392	0,682	0,911	0,852	0,543	0,422	0,194	0,630	0,700
rs16888746	G	0,995	0,164	0,795	0,882	0,969	0,893	0,734	0,516	0,394	0,798	0,864
	A	0,005	0,836	0,204	0,118	0,031	0,107	0,266	0,484	0,606	0,202	0,136
	C	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
rs10484813	A	0,951	0,156	0,778	0,844	0,938	0,861	0,670	0,508	0,388	0,784	0,829
	G	0,049	0,844	0,222	0,156	0,063	0,139	0,330	0,492	0,612	0,216	0,171
rs4458655	T	0,600	0,213	0,862	0,876	0,625	0,680	0,734	0,648	0,382	0,683	0,792
	C	0,400	0,787	0,138	0,124	0,375	0,320	0,266	0,352	0,618	0,317	0,208
rs1871428	G	0,011	0,221	0,198	0,610	0,078	0,156	0,383	0,344	0,229	0,442	0,440
	A	0,989	0,779	0,802	0,390	0,922	0,844	0,617	0,656	0,771	0,558	0,560
rs917118	C	0,277	0,491	0,711	0,720	0,375	0,377	0,606	0,609	0,506	0,702	0,622
	T	0,723	0,509	0,289	0,280	0,625	0,623	0,394	0,391	0,494	0,298	0,378
rs1550392	C	0,916	0,115	0,514	0,781	0,875	0,852	0,585	0,461	0,282	0,764	0,769
	T	0,084	0,885	0,486	0,219	0,125	0,148	0,415	0,539	0,718	0,236	0,231
rs17166119	A	0,893	0,107	0,512	0,730	0,870	0,836	0,548	0,422	0,265	0,721	0,714
	G	0,107	0,893	0,488	0,270	0,130	0,164	0,452	0,578	0,735	0,279	0,286
rs731257	G	0,986	0,123	0,652	0,865	0,990	0,951	0,691	0,500	0,347	0,808	0,843
	A	0,014	0,877	0,348	0,135	0,010	0,049	0,309	0,500	0,653	0,192	0,157
rs917115	T	0,210	0,156	0,039	0,758	0,250	0,279	0,574	0,453	0,312	0,596	0,618
	C	0,790	0,844	0,961	0,242	0,750	0,721	0,426	0,547	0,688	0,404	0,382
rs32314	C	0,297	0,967	0,595	0,257	0,234	0,246	0,372	0,531	0,800	0,375	0,310
	T	0,703	0,033	0,405	0,743	0,766	0,754	0,628	0,469	0,200	0,625	0,690
rs12701660	T	0,178	0,984	0,449	0,303	0,234	0,287	0,447	0,633	0,776	0,447	0,355
	C	0,822	0,016	0,551	0,697	0,766	0,713	0,553	0,367	0,224	0,553	0,645
rs2330442	A	0,016	0,197	0,713	0,604	0,094	0,107	0,606	0,438	0,400	0,563	0,438
	G	0,984	0,803	0,287	0,396	0,906	0,893	0,394	0,563	0,600	0,438	0,562
rs2471552	C	0,067	0,943	0,202	0,185	0,094	0,148	0,309	0,438	0,682	0,308	0,208
	T	0,933	0,057	0,798	0,815	0,906	0,852	0,691	0,563	0,318	0,692	0,792
rs4717865	G	0,999	0,525	0,615	0,919	0,995	0,959	0,777	0,656	0,524	0,880	0,908
	A	0,001	0,475	0,385	0,081	0,005	0,041	0,223	0,344	0,476	0,120	0,092
rs705308	C	0,022	0,033	0,162	0,546	0,042	0,123	0,356	0,195	0,153	0,471	0,401
	A	0,978	0,967	0,838	0,454	0,958	0,877	0,644	0,805	0,847	0,529	0,599
rs978874	C	0,142	0,975	0,703	0,290	0,182	0,139	0,484	0,617	0,812	0,462	0,360
	T	0,858	0,025	0,297	0,710	0,818	0,861	0,516	0,383	0,188	0,538	0,640
rs2040502	C	0,143	0,975	0,686	0,297	0,130	0,156	0,537	0,609	0,818	0,399	0,352
	T	0,857	0,025	0,314	0,703	0,870	0,844	0,463	0,391	0,182	0,601	0,648
rs7803075	A	0,989	0,844	0,876	0,290	0,896	0,820	0,441	0,617	0,788	0,490	0,487
	G	0,011	0,156	0,124	0,710	0,104	0,180	0,559	0,383	0,212	0,510	0,513
rs6464211	C	0,199	0,680	0,703	0,861	0,302	0,426	0,755	0,766	0,794	0,736	0,696
	T	0,801	0,320	0,297	0,139	0,698	0,574	0,245	0,234	0,206	0,264	0,304
rs10108270	C	0,067	0,967	0,663	0,713	0,161	0,213	0,718	0,867	0,800	0,668	0,587

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	A	0,933	0,033	0,337	0,287	0,839	0,787	0,282	0,133	0,200	0,332	0,413
	T	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
rs4875427	G	0,108	0,861	0,562	0,295	0,141	0,172	0,436	0,641	0,765	0,375	0,294
	A	0,892	0,139	0,438	0,705	0,859	0,828	0,564	0,359	0,235	0,625	0,706
rs17072176	T	0,987	0,164	0,466	0,866	0,964	0,951	0,649	0,477	0,276	0,779	0,824
	G	0,013	0,836	0,534	0,134	0,036	0,049	0,351	0,523	0,724	0,221	0,176
rs400404	T	0,854	0,016	0,152	0,738	0,797	0,770	0,553	0,328	0,182	0,736	0,656
	G	0,146	0,984	0,848	0,262	0,203	0,230	0,447	0,672	0,818	0,264	0,344
rs11776114	G	0,984	0,033	0,236	0,685	0,943	0,893	0,628	0,313	0,229	0,668	0,705
	A	0,016	0,967	0,764	0,315	0,057	0,107	0,372	0,688	0,771	0,332	0,295
rs17807624	C	0,903	0,033	0,079	0,686	0,891	0,754	0,622	0,398	0,182	0,587	0,659
	T	0,097	0,967	0,921	0,314	0,109	0,246	0,378	0,602	0,818	0,413	0,341
rs3943253	A	0,674	0,189	0,406	0,918	0,708	0,672	0,697	0,563	0,288	0,803	0,814
	G	0,326	0,811	0,594	0,082	0,292	0,328	0,303	0,438	0,712	0,197	0,186
rs1471939	C	0,317	0,779	0,726	0,199	0,292	0,311	0,362	0,570	0,735	0,284	0,286
	T	0,683	0,221	0,274	0,801	0,708	0,689	0,638	0,430	0,265	0,716	0,714
rs1462906	T	0,755	0,000	0,006	0,030	0,682	0,615	0,085	0,008	0,012	0,111	0,166
	C	0,245	1,000	0,994	0,970	0,318	0,385	0,915	0,992	0,988	0,889	0,834
rs7018273	A	0,169	1,000	0,999	0,979	0,250	0,270	0,894	0,945	0,965	0,817	0,837
	G	0,831	0,000	0,001	0,021	0,750	0,730	0,106	0,055	0,035	0,183	0,163
rs12544346	G	0,979	0,508	0,543	0,366	0,901	0,861	0,495	0,516	0,571	0,519	0,482
	A	0,021	0,492	0,457	0,634	0,099	0,139	0,505	0,484	0,429	0,481	0,518
rs6990312	G	0,276	0,648	0,984	0,827	0,318	0,434	0,766	0,695	0,682	0,740	0,719
	T	0,724	0,352	0,016	0,173	0,682	0,566	0,234	0,305	0,318	0,260	0,281
rs896680	T	0,169	0,852	0,553	0,159	0,167	0,246	0,404	0,539	0,788	0,264	0,220
	C	0,831	0,148	0,447	0,841	0,833	0,754	0,596	0,461	0,212	0,736	0,780
rs17742080	C	0,998	0,123	0,571	0,922	0,990	0,943	0,665	0,578	0,382	0,846	0,887
	T	0,002	0,877	0,429	0,078	0,010	0,057	0,335	0,422	0,618	0,154	0,113
rs2196051	A	0,013	0,033	0,004	0,707	0,057	0,148	0,441	0,320	0,135	0,538	0,489
	G	0,987	0,967	0,996	0,293	0,943	0,852	0,559	0,680	0,865	0,462	0,511
rs7844723	C	0,992	0,811	0,885	0,415	0,974	0,877	0,574	0,633	0,718	0,620	0,565
	T	0,008	0,189	0,115	0,585	0,026	0,123	0,426	0,367	0,282	0,380	0,435
rs4870958	A	0,905	0,107	0,441	0,841	0,917	0,885	0,633	0,484	0,282	0,764	0,806
	G	0,095	0,893	0,559	0,159	0,083	0,115	0,367	0,516	0,718	0,236	0,194
rs6993205	A	0,018	1,000	0,949	0,513	0,052	0,172	0,606	0,805	0,847	0,481	0,442
	G	0,982	0,000	0,051	0,487	0,948	0,828	0,394	0,195	0,153	0,519	0,558
rs2001907	C	0,973	0,418	0,786	0,933	0,932	0,910	0,851	0,578	0,406	0,856	0,886
	T	0,027	0,582	0,214	0,067	0,068	0,090	0,149	0,422	0,594	0,144	0,114
rs1871534	C	0,925	0,025	0,000	0,007	0,880	0,762	0,085	0,047	0,047	0,101	0,806
	G	0,075	0,975	1,000	0,993	0,120	0,238	0,915	0,953	0,953	0,899	0,194
rs1408801	A	0,461	0,197	0,229	0,873	0,458	0,459	0,585	0,523	0,241	0,769	0,740
	G	0,539	0,803	0,771	0,127	0,542	0,541	0,415	0,477	0,759	0,231	0,260
rs10962599	C	0,998	0,959	0,997	0,287	0,911	0,910	0,543	0,742	0,876	0,510	0,501

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	T	0,002	0,041	0,003	0,713	0,089	0,090	0,457	0,258	0,124	0,490	0,499
rs10812316	T	0,221	0,270	0,442	0,929	0,344	0,402	0,691	0,523	0,276	0,764	0,740
	C	0,779	0,730	0,558	0,071	0,656	0,598	0,309	0,477	0,724	0,236	0,260
rs10511828	T	0,891	0,295	0,854	0,939	0,901	0,852	0,798	0,633	0,506	0,875	0,889
	C	0,109	0,705	0,146	0,061	0,099	0,148	0,202	0,367	0,494	0,125	0,111
rs3793451	C	0,954	0,680	0,748	0,937	0,958	0,943	0,851	0,680	0,665	0,899	0,921
	T	0,046	0,320	0,252	0,063	0,042	0,057	0,149	0,320	0,335	0,101	0,079
rs4877068	G	0,955	0,131	0,540	0,929	0,896	0,869	0,686	0,445	0,271	0,870	0,875
	A	0,045	0,869	0,460	0,071	0,104	0,131	0,314	0,555	0,729	0,130	0,125
rs2306040	T	0,942	0,393	0,770	0,950	0,932	0,934	0,777	0,641	0,394	0,889	0,906
	C	0,058	0,607	0,230	0,050	0,068	0,066	0,223	0,359	0,606	0,111	0,094
rs4135211	A	0,994	0,197	0,444	0,945	0,990	0,959	0,777	0,617	0,529	0,875	0,895
	C	0,006	0,803	0,556	0,055	0,010	0,041	0,223	0,383	0,471	0,125	0,105
rs4979274	C	0,995	0,090	0,582	0,912	0,990	0,926	0,654	0,516	0,382	0,822	0,839
	A	0,005	0,910	0,418	0,088	0,010	0,074	0,346	0,484	0,618	0,178	0,161
rs10513300	T	0,999	0,311	0,742	0,945	0,984	1,000	0,766	0,703	0,471	0,813	0,905
	C	0,001	0,689	0,258	0,055	0,016	0,000	0,234	0,297	0,529	0,188	0,095
rs3814134	A	0,121	0,975	0,687	0,978	0,240	0,402	0,904	0,922	0,941	0,885	0,806
	G	0,879	0,025	0,313	0,022	0,760	0,598	0,096	0,078	0,059	0,115	0,194
rs2073821	C	1,000	0,410	0,701	0,919	0,995	0,951	0,846	0,695	0,635	0,856	0,888
	T	0,000	0,590	0,299	0,081	0,005	0,049	0,154	0,305	0,365	0,144	0,112
rs1412512	T	0,691	0,074	0,333	0,778	0,714	0,795	0,580	0,414	0,288	0,688	0,700
	C	0,309	0,926	0,667	0,222	0,286	0,205	0,420	0,586	0,712	0,313	0,300
rs4880510	T	0,910	0,156	0,641	0,915	0,875	0,885	0,638	0,578	0,318	0,793	0,862
	G	0,090	0,844	0,359	0,085	0,125	0,115	0,362	0,422	0,682	0,207	0,138
rs4880511	C	0,972	0,156	0,665	0,916	0,938	0,959	0,676	0,578	0,318	0,798	0,874
	A	0,028	0,844	0,335	0,084	0,063	0,041	0,324	0,422	0,682	0,202	0,126
rs17143387	T	0,950	0,369	0,535	0,854	0,932	0,910	0,686	0,484	0,312	0,813	0,843
	G	0,050	0,631	0,465	0,146	0,068	0,090	0,314	0,516	0,688	0,188	0,157
rs7923994	G	0,382	0,984	0,424	0,184	0,307	0,402	0,426	0,625	0,788	0,322	0,303
	A	0,618	0,016	0,576	0,816	0,693	0,598	0,574	0,375	0,212	0,678	0,697
rs912069	T	0,756	0,180	0,368	0,778	0,734	0,672	0,590	0,492	0,400	0,740	0,728
	C	0,244	0,820	0,632	0,222	0,266	0,328	0,410	0,508	0,600	0,260	0,272
rs7897550	G	0,945	0,492	0,602	0,750	0,917	0,836	0,670	0,516	0,476	0,788	0,773
	A	0,055	0,508	0,398	0,250	0,083	0,164	0,330	0,484	0,524	0,212	0,227
rs6482103	C	0,094	0,742	0,439	0,096	0,068	0,148	0,239	0,477	0,635	0,226	0,135
	A	0,906	0,258	0,561	0,904	0,932	0,852	0,761	0,523	0,365	0,774	0,866
rs1978806	A	0,488	1,000	1,000	0,995	0,531	0,549	0,963	0,938	0,976	0,904	0,900
	G	0,512	0,000	0,000	0,005	0,469	0,451	0,037	0,063	0,024	0,096	0,100
rs3793791	C	0,264	0,689	0,231	0,067	0,229	0,189	0,223	0,328	0,629	0,159	0,172
	T	0,736	0,311	0,769	0,933	0,771	0,811	0,777	0,672	0,371	0,841	0,828
rs201381710	C	0,403	0,984	0,712	0,566	0,396	0,443	0,644	0,711	0,918	0,596	0,557
	T	0,054	0,008	0,162	0,226	0,073	0,107	0,138	0,125	0,035	0,173	0,161

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	A	0,543	0,008	0,126	0,209	0,531	0,451	0,218	0,164	0,047	0,231	0,283
rs10821745	T	0,924	0,189	0,726	0,952	0,901	0,885	0,747	0,672	0,559	0,885	0,898
	G	0,076	0,811	0,274	0,048	0,099	0,115	0,253	0,328	0,441	0,115	0,102
rs4948418	C	0,995	0,213	0,775	0,946	0,979	0,967	0,739	0,680	0,559	0,885	0,908
	T	0,005	0,787	0,225	0,054	0,021	0,033	0,261	0,320	0,441	0,115	0,092
rs10994397	C	0,937	0,254	0,702	0,949	0,896	0,861	0,755	0,680	0,565	0,870	0,904
	T	0,063	0,746	0,298	0,051	0,104	0,139	0,245	0,320	0,435	0,130	0,096
rs3998448	C	0,996	0,197	0,319	0,648	0,943	0,902	0,527	0,453	0,265	0,620	0,674
	T	0,004	0,803	0,681	0,352	0,057	0,098	0,473	0,547	0,735	0,380	0,326
rs4746136	G	0,999	0,197	0,478	0,876	0,974	0,910	0,686	0,555	0,294	0,803	0,857
	A	0,001	0,803	0,522	0,124	0,026	0,090	0,314	0,445	0,706	0,197	0,143
rs7902158	A	0,067	0,057	0,234	0,761	0,167	0,205	0,479	0,344	0,159	0,524	0,585
	G	0,933	0,943	0,766	0,239	0,833	0,795	0,521	0,656	0,841	0,476	0,415
rs3814614	A	0,122	0,992	0,252	0,480	0,161	0,230	0,548	0,688	0,829	0,563	0,448
	G	0,878	0,008	0,748	0,520	0,839	0,770	0,452	0,313	0,171	0,438	0,552
rs4918664	A	0,993	0,148	0,131	0,883	0,953	0,934	0,665	0,492	0,341	0,832	0,852
	G	0,007	0,852	0,869	0,117	0,047	0,066	0,335	0,508	0,659	0,168	0,148
rs12220128	A	0,945	0,131	0,133	0,861	0,917	0,844	0,665	0,461	0,318	0,808	0,818
	C	0,055	0,869	0,867	0,139	0,083	0,156	0,335	0,539	0,682	0,192	0,182
rs7917775	A	0,921	0,328	0,440	0,912	0,906	0,926	0,809	0,586	0,506	0,827	0,862
	C	0,079	0,672	0,560	0,088	0,094	0,074	0,191	0,414	0,494	0,173	0,138
	G	0,956	0,083	0,746	0,945	0,958	0,918	0,771	0,570	0,241	0,861	0,869
rs17130385	T	0,043	0,917	0,254	0,055	0,042	0,082	0,229	0,430	0,759	0,139	0,131
	A	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
rs734241	G	0,939	0,049	0,613	0,939	0,958	0,918	0,739	0,531	0,224	0,837	0,857
	A	0,061	0,951	0,387	0,061	0,042	0,082	0,261	0,469	0,776	0,163	0,143
rs7919248	T	0,903	0,049	0,488	0,931	0,885	0,861	0,713	0,539	0,218	0,798	0,834
	G	0,097	0,951	0,512	0,069	0,115	0,139	0,287	0,461	0,782	0,202	0,166
rs6585227	T	0,253	0,959	0,507	0,078	0,250	0,246	0,293	0,484	0,782	0,207	0,199
	C	0,747	0,041	0,493	0,922	0,750	0,754	0,707	0,516	0,218	0,793	0,801
rs4918842	T	0,926	0,074	0,530	0,873	0,901	0,910	0,734	0,469	0,247	0,837	0,810
	C	0,074	0,926	0,470	0,127	0,099	0,090	0,266	0,531	0,753	0,163	0,190
rs1419138	C	0,860	0,811	0,623	0,068	0,750	0,672	0,356	0,469	0,629	0,308	0,276
	T	0,140	0,189	0,377	0,932	0,250	0,328	0,644	0,531	0,371	0,692	0,724
rs853577	A	0,733	0,811	0,784	0,064	0,615	0,607	0,303	0,422	0,618	0,288	0,260
	G	0,267	0,189	0,216	0,936	0,385	0,393	0,697	0,578	0,382	0,712	0,740
rs4880436	C	0,996	0,574	0,998	0,940	0,990	0,992	0,846	0,828	0,700	0,942	0,927
	T	0,004	0,426	0,002	0,060	0,010	0,008	0,154	0,172	0,300	0,058	0,073
rs10839880	C	0,956	0,287	0,514	0,375	0,922	0,820	0,356	0,359	0,388	0,500	0,480
	T	0,044	0,713	0,486	0,625	0,078	0,180	0,644	0,641	0,612	0,500	0,520
rs2403512	T	0,975	0,172	0,753	0,861	0,948	0,910	0,686	0,555	0,271	0,817	0,838
	C	0,025	0,828	0,247	0,139	0,052	0,090	0,314	0,445	0,729	0,183	0,162
rs1837606	T	0,290	0,230	0,741	0,742	0,323	0,508	0,564	0,477	0,247	0,500	0,611

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	C	0,710	0,770	0,259	0,258	0,677	0,492	0,436	0,523	0,753	0,500	0,389
rs2946788	G	0,881	0,213	0,642	0,261	0,802	0,754	0,271	0,250	0,176	0,317	0,351
	T	0,119	0,787	0,358	0,739	0,198	0,246	0,729	0,750	0,824	0,683	0,649
rs1561937	G	0,896	0,000	0,033	0,096	0,818	0,705	0,149	0,055	0,100	0,178	0,233
	A	0,104	1,000	0,967	0,904	0,182	0,295	0,851	0,945	0,900	0,822	0,767
	C	0,296	0,344	0,279	0,684	0,339	0,418	0,644	0,578	0,494	0,601	0,604
rs5030240	G	0,308	0,656	0,676	0,228	0,302	0,311	0,282	0,398	0,471	0,293	0,264
	A	0,396	0,000	0,045	0,089	0,359	0,270	0,074	0,023	0,035	0,106	0,132
rs3736508	C	1,000	0,164	0,418	0,978	0,990	0,984	0,835	0,672	0,565	0,808	0,919
	T	0,000	0,836	0,582	0,022	0,010	0,016	0,165	0,328	0,435	0,192	0,081
rs509360	A	0,942	0,016	0,456	0,304	0,865	0,820	0,239	0,141	0,124	0,284	0,422
	G	0,058	0,984	0,544	0,696	0,135	0,180	0,761	0,859	0,876	0,716	0,578
rs174546	C	0,991	0,033	0,467	0,698	0,969	0,893	0,574	0,313	0,188	0,510	0,721
	T	0,009	0,967	0,533	0,302	0,031	0,107	0,426	0,688	0,812	0,490	0,279
rs174570	C	0,997	0,049	0,475	0,870	0,995	0,943	0,660	0,406	0,224	0,683	0,860
	T	0,003	0,951	0,525	0,130	0,005	0,057	0,340	0,594	0,776	0,317	0,140
rs1535	A	0,901	0,033	0,467	0,694	0,865	0,844	0,569	0,305	0,194	0,495	0,701
	G	0,099	0,967	0,533	0,306	0,135	0,156	0,431	0,695	0,806	0,505	0,299
rs4963452	T	0,847	0,057	0,576	0,854	0,880	0,803	0,638	0,516	0,276	0,668	0,811
	C	0,153	0,943	0,424	0,146	0,120	0,197	0,362	0,484	0,724	0,332	0,189
rs11227699	G	0,347	0,475	0,861	0,968	0,432	0,475	0,750	0,664	0,359	0,793	0,823
	A	0,653	0,525	0,139	0,032	0,568	0,525	0,250	0,336	0,641	0,207	0,177
rs674499	A	0,846	0,557	0,245	0,061	0,724	0,746	0,293	0,398	0,682	0,269	0,232
	G	0,154	0,443	0,755	0,939	0,276	0,254	0,707	0,602	0,318	0,731	0,768
rs1638566	T	0,151	0,443	0,753	0,939	0,276	0,254	0,702	0,602	0,318	0,731	0,761
	G	0,849	0,557	0,247	0,061	0,724	0,746	0,298	0,398	0,682	0,269	0,239
rs1638567	T	0,151	0,443	0,753	0,939	0,276	0,254	0,702	0,602	0,318	0,731	0,761
	C	0,849	0,557	0,247	0,061	0,724	0,746	0,298	0,398	0,682	0,269	0,239
rs599190	A	0,035	0,917	0,766	0,221	0,063	0,082	0,298	0,625	0,618	0,356	0,254
	G	0,965	0,083	0,234	0,779	0,938	0,918	0,702	0,375	0,382	0,644	0,746
rs2619206	G	0,921	0,230	0,439	0,874	0,906	0,918	0,723	0,523	0,547	0,755	0,846
	A	0,079	0,770	0,561	0,126	0,094	0,082	0,277	0,477	0,453	0,245	0,154
rs2241667	T	0,934	0,131	0,590	0,794	0,922	0,910	0,638	0,469	0,259	0,712	0,769
	G	0,066	0,869	0,410	0,206	0,078	0,090	0,362	0,531	0,741	0,288	0,231
rs16921059	T	0,874	0,131	0,590	0,794	0,870	0,861	0,633	0,453	0,259	0,707	0,756
	C	0,126	0,869	0,410	0,206	0,130	0,139	0,367	0,547	0,741	0,293	0,244
rs4309121	T	0,982	0,123	0,717	0,839	0,964	0,926	0,676	0,492	0,276	0,774	0,823
	C	0,018	0,877	0,283	0,161	0,036	0,074	0,324	0,508	0,724	0,226	0,177
rs1079597	C	0,838	0,385	0,580	0,858	0,875	0,787	0,851	0,609	0,582	0,813	0,830
	T	0,162	0,615	0,420	0,142	0,125	0,213	0,149	0,391	0,418	0,188	0,170
rs948028	A	0,231	0,754	0,837	0,828	0,313	0,336	0,723	0,633	0,612	0,716	0,684
	C	0,769	0,246	0,163	0,172	0,688	0,664	0,277	0,367	0,388	0,284	0,316
rs4937417	A	0,979	0,049	0,596	0,558	0,943	0,828	0,548	0,352	0,235	0,635	0,618

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	G	0,021	0,951	0,404	0,442	0,057	0,172	0,452	0,648	0,765	0,365	0,382
rs3217863	C	1,000	0,205	0,672	0,931	0,984	0,951	0,777	0,648	0,347	0,846	0,907
	T	0,000	0,795	0,328	0,069	0,016	0,049	0,223	0,352	0,653	0,154	0,093
rs2416791	A	0,960	0,508	0,270	0,099	0,875	0,762	0,314	0,344	0,382	0,293	0,292
	G	0,040	0,492	0,730	0,901	0,125	0,238	0,686	0,656	0,618	0,707	0,708
rs1513056	G	0,668	0,150	0,486	0,860	0,646	0,738	0,617	0,500	0,476	0,716	0,774
	A	0,332	0,850	0,514	0,140	0,354	0,262	0,383	0,500	0,524	0,284	0,226
rs10843344	C	0,996	0,283	0,885	0,668	0,953	0,893	0,580	0,508	0,529	0,702	0,723
	T	0,004	0,717	0,115	0,332	0,047	0,107	0,420	0,492	0,471	0,298	0,277
rs214678	T	0,927	0,367	0,834	0,905	0,896	0,893	0,734	0,570	0,441	0,788	0,868
	C	0,073	0,633	0,166	0,095	0,104	0,107	0,266	0,430	0,559	0,212	0,132
rs2051827	G	0,948	0,167	0,926	0,950	0,964	0,934	0,766	0,563	0,388	0,832	0,890
	A	0,052	0,833	0,074	0,050	0,036	0,066	0,234	0,438	0,612	0,168	0,110
rs11612312	T	0,974	0,051	0,711	0,788	0,964	0,885	0,548	0,391	0,253	0,673	0,758
	C	0,026	0,949	0,289	0,212	0,036	0,115	0,452	0,609	0,747	0,327	0,242
rs11521	C	0,946	0,107	0,561	0,786	0,911	0,885	0,638	0,430	0,324	0,697	0,769
	T	0,054	0,893	0,439	0,214	0,089	0,115	0,362	0,570	0,676	0,303	0,231
rs772262	G	0,192	0,342	0,836	0,917	0,240	0,336	0,739	0,633	0,594	0,774	0,742
	A	0,808	0,658	0,164	0,083	0,760	0,664	0,261	0,367	0,406	0,226	0,258
rs773658	G	0,680	0,000	0,023	0,004	0,531	0,418	0,064	0,039	0,024	0,115	0,147
	C	0,320	1,000	0,977	0,996	0,469	0,582	0,936	0,961	0,976	0,885	0,853
rs2216437	A	0,801	0,074	0,465	0,924	0,802	0,779	0,702	0,484	0,335	0,827	0,833
	G	0,199	0,926	0,535	0,076	0,198	0,221	0,298	0,516	0,665	0,173	0,167
rs10878834	T	0,805	0,075	0,466	0,928	0,802	0,779	0,702	0,484	0,335	0,827	0,835
	C	0,195	0,925	0,534	0,072	0,198	0,221	0,298	0,516	0,665	0,173	0,165
rs2052386	G	0,911	0,074	0,463	0,913	0,885	0,869	0,697	0,492	0,341	0,817	0,842
	A	0,089	0,926	0,537	0,087	0,115	0,131	0,303	0,508	0,659	0,183	0,158
rs3851621	T	0,973	0,328	0,647	0,953	0,958	0,943	0,846	0,711	0,459	0,875	0,895
	C	0,027	0,672	0,353	0,047	0,042	0,057	0,154	0,289	0,541	0,125	0,105
rs1397560	C	0,799	0,082	0,394	0,862	0,781	0,787	0,670	0,477	0,335	0,779	0,798
	A	0,201	0,918	0,606	0,138	0,219	0,213	0,330	0,523	0,665	0,221	0,202
rs2070586	G	0,600	0,300	0,684	0,826	0,536	0,689	0,612	0,531	0,335	0,678	0,757
	A	0,400	0,700	0,316	0,174	0,464	0,311	0,388	0,469	0,665	0,322	0,243
rs2238151	T	0,002	0,467	0,069	0,707	0,089	0,205	0,537	0,484	0,700	0,577	0,558
	C	0,998	0,533	0,931	0,293	0,911	0,795	0,463	0,516	0,300	0,423	0,443
rs671	G	1,000	1,000	0,825	1,000	0,995	1,000	1,000	0,992	0,994	1,000	0,989
	A	0,000	0,000	0,175	0,000	0,005	0,000	0,000	0,008	0,006	0,000	0,011
rs1503767	T	0,228	0,500	0,461	0,845	0,292	0,410	0,676	0,695	0,671	0,668	0,686
	G	0,768	0,500	0,539	0,155	0,698	0,590	0,324	0,305	0,329	0,332	0,314
	A	0,004	0,000	0,000	0,000	0,010	0,000	0,000	0,000	0,000	0,000	0,000
rs1335873	T	0,944	0,262	0,338	0,295	0,901	0,787	0,410	0,281	0,259	0,457	0,432
	A	0,056	0,738	0,662	0,705	0,099	0,213	0,590	0,719	0,741	0,543	0,568
rs1886510	G	0,925	0,648	0,855	0,462	0,818	0,787	0,596	0,477	0,488	0,582	0,594

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	A	0,075	0,352	0,145	0,538	0,182	0,213	0,404	0,523	0,512	0,418	0,406
rs17359176	G	0,990	0,148	0,828	0,934	0,984	0,967	0,745	0,602	0,347	0,837	0,895
	A	0,010	0,852	0,172	0,066	0,016	0,033	0,255	0,398	0,653	0,163	0,105
rs9319336	T	0,840	0,262	0,328	0,949	0,849	0,910	0,729	0,563	0,324	0,827	0,851
	C	0,160	0,738	0,672	0,051	0,151	0,090	0,271	0,438	0,676	0,173	0,149
rs566996	C	0,952	0,107	0,813	0,772	0,969	0,893	0,617	0,398	0,229	0,697	0,780
	T	0,048	0,893	0,187	0,228	0,031	0,107	0,383	0,602	0,771	0,303	0,220
rs7997709	C	0,185	0,893	0,683	0,047	0,260	0,279	0,255	0,563	0,706	0,221	0,164
	T	0,815	0,107	0,317	0,953	0,740	0,721	0,745	0,438	0,294	0,779	0,836
rs2065982	T	0,936	0,107	0,303	0,954	0,885	0,869	0,750	0,438	0,294	0,793	0,875
	C	0,064	0,893	0,697	0,046	0,115	0,131	0,250	0,563	0,706	0,207	0,125
rs1572018	T	0,977	0,074	0,652	0,144	0,839	0,795	0,176	0,148	0,147	0,264	0,336
	C	0,023	0,926	0,348	0,856	0,161	0,205	0,824	0,852	0,853	0,736	0,664
rs2166624	G	0,998	0,025	0,596	0,635	0,938	0,918	0,484	0,406	0,188	0,572	0,641
	A	0,002	0,975	0,404	0,365	0,063	0,082	0,516	0,594	0,812	0,428	0,359
rs7326934	G	0,158	0,910	0,988	0,963	0,297	0,336	0,910	0,906	0,965	0,841	0,806
	C	0,842	0,090	0,012	0,037	0,703	0,664	0,090	0,094	0,035	0,159	0,194
rs9563982	A	0,992	0,148	0,213	0,847	0,990	0,893	0,660	0,578	0,588	0,813	0,818
	G	0,008	0,852	0,787	0,153	0,010	0,107	0,340	0,422	0,412	0,188	0,182
rs1409842	A	0,749	0,008	0,561	0,514	0,760	0,582	0,394	0,281	0,159	0,481	0,523
	G	0,251	0,992	0,439	0,486	0,240	0,418	0,606	0,719	0,841	0,519	0,477
rs1000973	G	0,910	0,066	0,547	0,739	0,891	0,820	0,521	0,484	0,318	0,712	0,728
	A	0,090	0,934	0,453	0,261	0,109	0,180	0,479	0,516	0,682	0,288	0,272
rs9530435	T	0,868	0,049	0,022	0,191	0,818	0,721	0,218	0,180	0,094	0,308	0,298
	C	0,132	0,951	0,978	0,809	0,182	0,279	0,782	0,820	0,906	0,692	0,702
rs9522149	T	0,969	0,942	0,983	0,224	0,885	0,803	0,431	0,641	0,794	0,457	0,426
	C	0,031	0,058	0,017	0,776	0,115	0,197	0,569	0,359	0,206	0,543	0,574
rs1760921	T	0,279	0,910	0,810	0,952	0,339	0,475	0,910	0,805	0,759	0,880	0,807
	C	0,721	0,090	0,190	0,048	0,661	0,525	0,090	0,195	0,241	0,120	0,193
rs11844034	G	0,164	0,984	0,869	0,307	0,198	0,254	0,484	0,617	0,818	0,370	0,315
	A	0,836	0,016	0,132	0,693	0,802	0,746	0,516	0,383	0,182	0,630	0,685
rs10141763	T	0,524	0,918	0,539	0,949	0,552	0,639	0,862	0,883	0,971	0,861	0,849
	A	0,476	0,082	0,461	0,051	0,448	0,361	0,138	0,117	0,029	0,139	0,151
rs7158003	T	0,295	0,967	0,598	0,631	0,349	0,434	0,745	0,828	0,894	0,591	0,605
	C	0,705	0,033	0,402	0,369	0,651	0,566	0,255	0,172	0,106	0,409	0,395
rs1958099	T	0,962	0,082	0,965	0,816	0,938	0,902	0,617	0,531	0,259	0,721	0,775
	G	0,038	0,918	0,035	0,184	0,063	0,098	0,383	0,469	0,741	0,279	0,225
rs17119138	A	0,951	0,049	0,627	0,765	0,896	0,869	0,495	0,453	0,194	0,745	0,746
	G	0,049	0,951	0,373	0,235	0,104	0,131	0,505	0,547	0,806	0,255	0,254
rs2357442	A	0,835	0,262	0,318	0,873	0,875	0,828	0,574	0,531	0,341	0,750	0,803
	C	0,165	0,738	0,682	0,127	0,125	0,172	0,426	0,469	0,659	0,250	0,197
rs17126825	G	0,977	0,131	0,549	0,875	0,974	0,967	0,697	0,531	0,312	0,784	0,846
	A	0,023	0,869	0,451	0,125	0,026	0,033	0,303	0,469	0,688	0,216	0,154

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs1950993	G	0,043	0,303	0,162	0,633	0,115	0,172	0,410	0,406	0,335	0,442	0,487
	T	0,957	0,697	0,838	0,367	0,885	0,828	0,590	0,594	0,665	0,558	0,513
rs810087	A	0,975	0,667	0,547	0,142	0,901	0,852	0,553	0,414	0,682	0,351	0,329
	G	0,025	0,333	0,453	0,858	0,099	0,148	0,447	0,586	0,318	0,649	0,671
rs8021730	G	0,188	0,549	0,516	0,829	0,271	0,369	0,644	0,594	0,476	0,649	0,662
	T	0,812	0,451	0,484	0,171	0,729	0,631	0,356	0,406	0,524	0,351	0,338
rs7142344	C	0,040	0,984	0,616	0,892	0,156	0,213	0,851	0,828	0,924	0,817	0,720
	T	0,960	0,016	0,384	0,108	0,844	0,787	0,149	0,172	0,076	0,183	0,280
rs946918	G	0,801	0,221	0,618	0,809	0,792	0,787	0,739	0,508	0,318	0,731	0,733
	T	0,199	0,779	0,382	0,191	0,208	0,213	0,261	0,492	0,682	0,269	0,267
rs200354	G	0,315	0,083	0,254	0,837	0,365	0,352	0,569	0,328	0,271	0,649	0,673
	T	0,685	0,917	0,746	0,163	0,635	0,648	0,431	0,672	0,729	0,351	0,327
rs730570	G	0,773	0,934	0,771	0,138	0,703	0,713	0,378	0,648	0,724	0,356	0,337
	A	0,227	0,066	0,229	0,862	0,297	0,287	0,622	0,352	0,276	0,644	0,663
rs3784230	A	0,024	0,607	0,929	0,630	0,109	0,107	0,569	0,578	0,371	0,519	0,498
	G	0,976	0,393	0,071	0,370	0,891	0,893	0,431	0,422	0,629	0,481	0,502
rs11161328	A	0,814	0,025	0,112	0,715	0,766	0,803	0,580	0,422	0,206	0,663	0,670
	G	0,186	0,975	0,888	0,285	0,234	0,197	0,420	0,578	0,794	0,337	0,330
rs1800414	T	1,000	1,000	0,423	0,999	0,995	1,000	1,000	1,000	1,000	1,000	0,984
	C	0,000	0,000	0,577	0,001	0,005	0,000	0,000	0,000	0,000	0,000	0,016
rs1448484	A	0,168	1,000	1,000	0,991	0,286	0,328	0,947	0,945	0,947	0,909	0,843
	G	0,832	0,000	0,000	0,009	0,714	0,672	0,053	0,055	0,053	0,091	0,157
rs12913832	A	0,996	0,893	0,999	0,458	0,922	0,861	0,734	0,820	0,888	0,769	0,628
	G	0,004	0,107	0,001	0,542	0,078	0,139	0,266	0,180	0,112	0,231	0,372
rs12439433	A	1,000	0,475	0,995	0,968	0,984	0,984	0,883	0,703	0,671	0,894	0,971
	G	0,000	0,525	0,005	0,032	0,016	0,016	0,117	0,297	0,329	0,106	0,029
rs2122497	C	0,913	0,049	0,237	0,788	0,870	0,893	0,638	0,352	0,212	0,678	0,759
	T	0,087	0,951	0,763	0,212	0,130	0,107	0,362	0,648	0,788	0,322	0,241
rs8030587	G	0,033	0,984	0,774	0,810	0,089	0,156	0,771	0,875	0,888	0,779	0,673
	A	0,967	0,016	0,226	0,190	0,911	0,844	0,229	0,125	0,112	0,221	0,327
rs12594483	G	0,024	0,869	0,723	0,893	0,109	0,156	0,803	0,867	0,800	0,784	0,714
	A	0,976	0,131	0,277	0,107	0,891	0,844	0,197	0,133	0,200	0,216	0,286
rs12917189	T	0,025	0,984	0,750	0,798	0,089	0,156	0,787	0,875	0,894	0,779	0,663
	C	0,975	0,016	0,250	0,202	0,911	0,844	0,213	0,125	0,106	0,221	0,337
rs735480	T	0,019	0,467	0,591	0,907	0,125	0,189	0,729	0,586	0,418	0,750	0,703
	C	0,981	0,533	0,409	0,093	0,875	0,811	0,271	0,414	0,582	0,250	0,297
rs1426654	A	0,052	0,033	0,017	0,998	0,104	0,189	0,718	0,516	0,282	0,769	0,738
	G	0,948	0,967	0,983	0,002	0,896	0,811	0,282	0,484	0,718	0,231	0,262
rs11637235	C	0,964	0,992	0,991	0,235	0,917	0,861	0,489	0,625	0,800	0,433	0,455
	T	0,036	0,008	0,009	0,765	0,083	0,139	0,511	0,375	0,200	0,567	0,545
rs12594750	T	0,886	0,189	0,594	0,871	0,906	0,893	0,676	0,594	0,353	0,788	0,834
	C	0,115	0,811	0,406	0,129	0,094	0,107	0,324	0,406	0,647	0,212	0,166
rs974828	C	0,076	0,975	0,840	0,909	0,135	0,172	0,872	0,891	0,924	0,779	0,713

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	T	0,924	0,025	0,160	0,091	0,865	0,828	0,128	0,109	0,076	0,221	0,287
rs3884558	A	0,185	0,926	0,615	0,096	0,188	0,205	0,319	0,453	0,753	0,178	0,180
	G	0,815	0,074	0,385	0,904	0,813	0,795	0,681	0,547	0,247	0,822	0,820
rs11631047	C	0,925	0,115	0,446	0,764	0,854	0,836	0,628	0,453	0,235	0,615	0,714
	T	0,075	0,885	0,554	0,236	0,146	0,164	0,372	0,547	0,765	0,385	0,286
rs6494489	C	0,869	0,025	0,102	0,648	0,807	0,779	0,505	0,281	0,153	0,587	0,628
	T	0,131	0,975	0,898	0,352	0,193	0,221	0,495	0,719	0,847	0,413	0,372
rs16951105	A	0,944	0,082	0,480	0,753	0,932	0,910	0,649	0,430	0,324	0,726	0,722
	G	0,056	0,918	0,520	0,247	0,068	0,090	0,351	0,570	0,676	0,274	0,278
rs2899826	A	0,605	0,270	0,521	0,917	0,615	0,697	0,660	0,602	0,235	0,784	0,770
	G	0,395	0,730	0,479	0,083	0,385	0,303	0,340	0,398	0,765	0,216	0,230
rs8035124	A	0,248	0,131	0,222	0,848	0,302	0,393	0,633	0,469	0,259	0,654	0,643
	C	0,752	0,869	0,779	0,152	0,698	0,607	0,367	0,531	0,741	0,346	0,357
rs4984913	A	0,109	0,250	0,260	0,747	0,167	0,270	0,622	0,469	0,376	0,644	0,593
	G	0,891	0,750	0,740	0,253	0,833	0,730	0,378	0,531	0,624	0,356	0,407
rs4141505	G	0,876	0,115	0,446	0,943	0,875	0,869	0,761	0,648	0,412	0,870	0,866
	A	0,124	0,885	0,554	0,057	0,125	0,131	0,239	0,352	0,588	0,130	0,134
rs4781011	T	0,570	0,852	0,883	0,246	0,589	0,557	0,404	0,625	0,835	0,322	0,351
	G	0,430	0,148	0,117	0,754	0,411	0,443	0,596	0,375	0,165	0,678	0,649
rs4780476	A	0,192	0,943	0,253	0,291	0,219	0,311	0,378	0,602	0,682	0,385	0,276
	C	0,808	0,057	0,747	0,709	0,781	0,689	0,622	0,398	0,318	0,615	0,724
rs2269793	T	0,395	0,328	0,624	0,914	0,380	0,459	0,745	0,672	0,482	0,788	0,792
	G	0,605	0,672	0,376	0,086	0,620	0,541	0,255	0,328	0,518	0,212	0,208
rs4889490	G	0,996	0,098	0,116	0,581	0,958	0,893	0,505	0,359	0,259	0,649	0,629
	T	0,004	0,902	0,884	0,419	0,042	0,107	0,495	0,641	0,741	0,351	0,371
rs881929	G	0,953	0,598	0,113	0,570	0,932	0,852	0,585	0,523	0,612	0,611	0,635
	T	0,047	0,402	0,887	0,430	0,068	0,148	0,415	0,477	0,388	0,389	0,365
rs818386	T	0,000	0,582	0,388	0,173	0,031	0,057	0,250	0,430	0,535	0,188	0,173
	C	1,000	0,418	0,612	0,827	0,969	0,943	0,750	0,570	0,465	0,813	0,827
rs2966849	A	0,352	0,787	0,434	0,089	0,307	0,262	0,223	0,328	0,594	0,159	0,194
	G	0,648	0,213	0,566	0,911	0,693	0,738	0,777	0,672	0,406	0,841	0,806
rs459920	T	0,842	0,893	0,977	0,495	0,786	0,795	0,606	0,656	0,871	0,563	0,563
	C	0,158	0,107	0,023	0,505	0,214	0,205	0,394	0,344	0,129	0,438	0,437
rs3785181	C	0,900	0,500	0,391	0,942	0,922	0,910	0,840	0,648	0,459	0,904	0,893
	T	0,100	0,500	0,609	0,058	0,078	0,090	0,160	0,352	0,541	0,096	0,107
rs1879488	A	0,214	0,689	0,493	0,056	0,182	0,221	0,165	0,406	0,606	0,178	0,122
	C	0,786	0,311	0,507	0,944	0,818	0,779	0,835	0,594	0,394	0,822	0,878
rs9916327	G	0,982	0,057	0,414	0,848	0,979	0,869	0,649	0,484	0,300	0,731	0,829
	A	0,018	0,943	0,586	0,152	0,021	0,131	0,351	0,516	0,700	0,269	0,171
rs4411548	C	0,828	0,525	0,978	0,759	0,792	0,779	0,766	0,672	0,594	0,822	0,777
	T	0,172	0,475	0,022	0,241	0,208	0,221	0,234	0,328	0,406	0,178	0,223
rs2593595	A	0,014	0,615	0,837	0,791	0,141	0,221	0,660	0,695	0,694	0,721	0,648
	G	0,986	0,385	0,163	0,209	0,859	0,779	0,340	0,305	0,306	0,279	0,352

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs1049620	T	0,176	0,910	0,522	0,180	0,156	0,246	0,372	0,523	0,771	0,380	0,242
	C	0,824	0,090	0,478	0,820	0,844	0,754	0,628	0,477	0,229	0,620	0,758
rs17642714	A	0,997	0,877	0,996	0,735	0,979	0,959	0,787	0,789	0,824	0,813	0,821
	T	0,003	0,123	0,004	0,265	0,021	0,041	0,213	0,211	0,176	0,188	0,179
rs4471745	G	0,931	0,983	0,868	0,948	0,953	0,975	0,957	0,945	0,959	0,976	0,965
	A	0,069	0,017	0,132	0,052	0,047	0,025	0,043	0,055	0,041	0,024	0,035
rs2033111	A	0,488	0,443	0,629	0,908	0,536	0,500	0,707	0,648	0,471	0,760	0,762
	G	0,512	0,557	0,371	0,092	0,464	0,500	0,293	0,352	0,529	0,240	0,238
rs2052074	A	0,024	0,959	0,837	0,816	0,141	0,238	0,670	0,859	0,829	0,736	0,631
	G	0,976	0,041	0,163	0,184	0,859	0,762	0,330	0,141	0,171	0,264	0,369
rs11652805	C	0,974	0,123	0,255	0,197	0,865	0,820	0,324	0,203	0,165	0,260	0,329
	T	0,026	0,877	0,745	0,803	0,135	0,180	0,676	0,797	0,835	0,740	0,671
rs10512572	G	0,804	0,283	0,425	0,867	0,729	0,779	0,766	0,484	0,365	0,793	0,777
	A	0,196	0,717	0,575	0,133	0,271	0,221	0,234	0,516	0,635	0,207	0,223
rs4789659	G	0,957	0,213	0,206	0,169	0,839	0,836	0,282	0,188	0,206	0,298	0,336
	A	0,043	0,787	0,794	0,831	0,161	0,164	0,718	0,813	0,794	0,702	0,664
rs1107704	A	0,070	0,926	0,648	0,890	0,193	0,221	0,761	0,859	0,847	0,755	0,713
	G	0,930	0,074	0,352	0,110	0,807	0,779	0,239	0,141	0,153	0,245	0,287
rs2125345	T	0,022	0,853	0,499	0,660	0,120	0,131	0,622	0,609	0,824	0,611	0,535
	C	0,978	0,147	0,501	0,340	0,880	0,869	0,378	0,391	0,176	0,389	0,465
rs2304925	A	0,436	0,852	0,443	0,802	0,510	0,549	0,798	0,844	0,935	0,731	0,748
	C	0,564	0,148	0,557	0,198	0,490	0,451	0,202	0,156	0,065	0,269	0,252
rs4798812	G	0,766	0,262	0,613	0,685	0,792	0,721	0,649	0,438	0,294	0,668	0,655
	A	0,234	0,738	0,387	0,315	0,208	0,279	0,351	0,563	0,706	0,332	0,345
rs4800105	C	0,752	0,262	0,730	0,894	0,755	0,697	0,734	0,578	0,576	0,788	0,797
	T	0,248	0,738	0,270	0,106	0,245	0,303	0,266	0,422	0,424	0,212	0,203
rs2042762	T	1,000	0,443	0,999	0,963	1,000	0,984	0,862	0,664	0,594	0,894	0,938
	C	0,000	0,557	0,001	0,037	0,000	0,016	0,138	0,336	0,406	0,106	0,062
rs7226659	G	0,977	0,557	0,403	0,976	0,974	0,934	0,835	0,820	0,647	0,923	0,938
	T	0,023	0,443	0,597	0,024	0,026	0,066	0,165	0,180	0,353	0,077	0,062
rs7238445	G	0,099	0,656	0,620	0,767	0,151	0,270	0,697	0,695	0,647	0,697	0,647
	A	0,901	0,344	0,380	0,233	0,849	0,730	0,303	0,305	0,353	0,303	0,353
rs881728	C	0,893	0,402	0,744	0,891	0,917	0,926	0,846	0,688	0,594	0,793	0,852
	A	0,107	0,598	0,256	0,109	0,083	0,074	0,154	0,313	0,406	0,207	0,148
rs3916235	T	0,876	0,025	0,004	0,048	0,786	0,738	0,069	0,055	0,047	0,168	0,191
	C	0,124	0,975	0,996	0,952	0,214	0,262	0,931	0,945	0,953	0,832	0,809
rs4891825	G	0,902	0,164	0,006	0,104	0,818	0,787	0,144	0,133	0,129	0,231	0,247
	A	0,098	0,836	0,994	0,896	0,182	0,213	0,856	0,867	0,871	0,769	0,753
rs874299	T	0,908	0,877	0,645	0,288	0,854	0,877	0,394	0,445	0,829	0,447	0,428
	C	0,092	0,123	0,355	0,712	0,146	0,123	0,606	0,555	0,171	0,553	0,572
rs1024116	C	0,735	0,607	0,896	0,452	0,792	0,713	0,383	0,578	0,435	0,452	0,490
	T	0,265	0,393	0,104	0,548	0,208	0,287	0,617	0,422	0,565	0,548	0,510
rs7252511	C	0,954	0,041	0,453	0,803	0,927	0,877	0,601	0,492	0,135	0,673	0,752

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
	T	0,046	0,959	0,547	0,197	0,073	0,123	0,399	0,508	0,865	0,327	0,248
rs7251928	A	0,082	1,000	0,983	0,759	0,151	0,180	0,793	0,836	0,906	0,736	0,661
	C	0,918	0,000	0,017	0,241	0,849	0,820	0,207	0,164	0,094	0,264	0,339
rs2313048	A	0,977	0,098	0,411	0,834	0,938	0,926	0,691	0,445	0,335	0,764	0,805
	G	0,023	0,902	0,589	0,166	0,063	0,074	0,309	0,555	0,665	0,236	0,195
rs8113143	C	0,002	0,205	0,272	0,634	0,094	0,156	0,479	0,383	0,229	0,471	0,466
	A	0,998	0,795	0,728	0,366	0,906	0,844	0,521	0,617	0,771	0,529	0,534
rs2074928	C	1,000	0,230	0,613	0,948	0,990	0,943	0,734	0,688	0,612	0,837	0,896
	T	0,000	0,770	0,387	0,052	0,010	0,057	0,266	0,313	0,388	0,163	0,104
rs717225	A	0,159	1,000	0,790	0,988	0,229	0,344	0,936	0,953	0,965	0,865	0,831
	G	0,841	0,000	0,210	0,012	0,771	0,656	0,064	0,047	0,035	0,135	0,169
rs12104228	T	0,210	1,000	0,789	0,989	0,255	0,369	0,947	0,953	0,965	0,865	0,839
	C	0,790	0,000	0,211	0,011	0,745	0,631	0,053	0,047	0,035	0,135	0,161
rs2303798	G	0,305	1,000	0,791	0,991	0,344	0,418	0,952	0,961	0,965	0,875	0,864
	A	0,695	0,000	0,209	0,009	0,656	0,582	0,048	0,039	0,035	0,125	0,136
rs1701930	A	0,306	0,025	0,898	0,414	0,286	0,303	0,372	0,211	0,124	0,346	0,366
	G	0,694	0,975	0,102	0,586	0,714	0,697	0,628	0,789	0,876	0,654	0,635
rs3745099	G	0,853	0,303	0,335	0,132	0,750	0,705	0,266	0,313	0,371	0,288	0,290
	A	0,147	0,697	0,665	0,868	0,250	0,295	0,734	0,688	0,629	0,712	0,710
rs2532060	T	0,070	0,385	0,239	0,607	0,109	0,197	0,559	0,484	0,512	0,423	0,456
	C	0,930	0,615	0,761	0,393	0,891	0,803	0,441	0,516	0,488	0,577	0,544
rs6104567	T	0,891	0,107	0,475	0,740	0,865	0,820	0,564	0,352	0,294	0,712	0,738
	G	0,109	0,893	0,525	0,260	0,135	0,180	0,436	0,648	0,706	0,288	0,262
rs1555133	G	0,867	0,041	0,184	0,739	0,844	0,803	0,596	0,414	0,159	0,721	0,678
	A	0,133	0,959	0,816	0,261	0,156	0,197	0,404	0,586	0,841	0,279	0,322
rs1321333	A	0,032	0,270	0,048	0,481	0,120	0,074	0,426	0,367	0,194	0,361	0,341
	G	0,968	0,730	0,952	0,519	0,880	0,926	0,574	0,633	0,806	0,639	0,659
rs3907047	T	0,999	0,336	0,462	0,962	0,984	0,959	0,809	0,609	0,488	0,938	0,901
	C	0,001	0,664	0,538	0,038	0,016	0,041	0,191	0,391	0,512	0,063	0,099
rs4811693	C	0,910	0,352	0,396	0,947	0,885	0,902	0,824	0,609	0,406	0,904	0,876
	T	0,090	0,648	0,604	0,053	0,115	0,098	0,176	0,391	0,594	0,096	0,124
rs6026036	A	0,062	0,967	0,815	0,338	0,063	0,090	0,516	0,648	0,812	0,337	0,325
	G	0,938	0,033	0,185	0,662	0,938	0,910	0,484	0,352	0,188	0,663	0,675
rs310644	T	0,056	0,951	0,965	0,924	0,135	0,320	0,878	0,930	0,894	0,822	0,752
	C	0,944	0,049	0,035	0,076	0,865	0,680	0,122	0,070	0,106	0,178	0,248
rs722098	A	0,141	0,246	0,456	0,827	0,214	0,287	0,543	0,578	0,394	0,611	0,661
	G	0,859	0,754	0,544	0,173	0,786	0,713	0,457	0,422	0,606	0,389	0,339
rs1626109	T	0,874	0,008	0,107	0,069	0,807	0,721	0,122	0,078	0,024	0,178	0,202
	C	0,126	0,992	0,893	0,931	0,193	0,279	0,878	0,922	0,976	0,822	0,798
rs239031	G	0,708	0,205	0,044	0,011	0,625	0,516	0,154	0,203	0,124	0,130	0,154
	A	0,292	0,795	0,956	0,989	0,375	0,484	0,846	0,797	0,876	0,870	0,846
rs2823850	C	0,118	0,992	0,929	0,993	0,172	0,328	0,915	0,953	0,941	0,817	0,830
	T	0,882	0,008	0,071	0,007	0,828	0,672	0,085	0,047	0,059	0,183	0,170

AIM	Alelo	AFR (608)	AMR (61)	EAS (693)	EUR (534)	ACB (96)	ASW (61)	CLM (94)	MXL (64)	PEL (85)	PUR (104)	RPT (513)
rs2572307	A	0,521	0,000	0,000	0,007	0,599	0,475	0,059	0,023	0,006	0,106	0,118
	G	0,479	1,000	1,000	0,993	0,401	0,525	0,941	0,977	0,994	0,894	0,882
rs2248656	A	0,969	0,115	0,738	0,818	0,953	0,934	0,622	0,539	0,459	0,769	0,795
	G	0,031	0,885	0,262	0,182	0,047	0,066	0,378	0,461	0,541	0,231	0,205
rs2835370	C	0,493	0,623	0,424	0,070	0,495	0,369	0,293	0,438	0,571	0,159	0,201
	T	0,507	0,377	0,576	0,930	0,505	0,631	0,707	0,563	0,429	0,841	0,799
rs2837371	C	0,894	0,090	0,617	0,875	0,922	0,836	0,633	0,406	0,241	0,779	0,792
	A	0,106	0,910	0,383	0,125	0,078	0,164	0,367	0,594	0,759	0,221	0,209
rs379839	A	0,041	0,975	0,595	0,408	0,089	0,180	0,516	0,719	0,835	0,428	0,389
	G	0,959	0,025	0,405	0,592	0,911	0,820	0,484	0,281	0,165	0,572	0,611
rs2269160	G	0,914	0,049	0,590	0,537	0,849	0,770	0,441	0,297	0,153	0,572	0,596
	T	0,086	0,951	0,410	0,463	0,151	0,230	0,559	0,703	0,847	0,428	0,404
rs2838665	G	0,900	0,115	0,666	0,752	0,865	0,861	0,676	0,391	0,200	0,788	0,770
	A	0,100	0,885	0,334	0,248	0,135	0,139	0,324	0,609	0,800	0,212	0,230
rs717177	T	0,896	0,107	0,669	0,763	0,865	0,852	0,670	0,391	0,182	0,788	0,769
	C	0,104	0,893	0,331	0,237	0,135	0,148	0,330	0,609	0,818	0,212	0,231
rs1296819	C	0,214	0,762	0,826	0,764	0,276	0,369	0,745	0,797	0,771	0,712	0,683
	A	0,786	0,238	0,174	0,236	0,724	0,631	0,255	0,203	0,229	0,288	0,317
rs5997008	C	0,447	1,000	0,983	0,985	0,500	0,639	0,941	0,961	0,971	0,923	0,867
	A	0,553	0,000	0,017	0,009	0,500	0,361	0,053	0,039	0,024	0,072	0,130
	T	0,000	0,000	0,000	0,006	0,000	0,000	0,005	0,000	0,006	0,005	0,003
rs4821004	C	0,096	0,639	0,851	0,597	0,172	0,213	0,585	0,609	0,547	0,558	0,496
	T	0,904	0,361	0,149	0,403	0,828	0,787	0,415	0,391	0,453	0,442	0,504
rs13054099	T	0,994	0,041	0,940	0,744	0,953	0,885	0,537	0,336	0,224	0,644	0,742
	C	0,006	0,959	0,060	0,256	0,047	0,115	0,463	0,664	0,776	0,356	0,258
rs2024566	A	0,910	0,025	0,926	0,681	0,823	0,820	0,495	0,328	0,212	0,615	0,668
	G	0,090	0,975	0,074	0,319	0,177	0,180	0,505	0,672	0,788	0,385	0,332
rs3927	T	0,947	0,025	0,922	0,736	0,885	0,852	0,543	0,359	0,224	0,688	0,724
	C	0,053	0,975	0,078	0,264	0,115	0,148	0,457	0,641	0,776	0,313	0,276
rs8137373	G	0,843	0,025	0,923	0,732	0,760	0,795	0,537	0,344	0,212	0,644	0,694
	A	0,157	0,975	0,077	0,268	0,240	0,205	0,463	0,656	0,788	0,356	0,306
rs2040411	G	0,151	0,475	0,709	0,314	0,188	0,205	0,319	0,391	0,312	0,375	0,297
	A	0,849	0,525	0,291	0,686	0,813	0,795	0,681	0,609	0,688	0,625	0,703
rs5768007	C	0,999	0,172	0,491	0,886	0,984	0,951	0,718	0,531	0,388	0,750	0,856
	T	0,001	0,828	0,509	0,114	0,016	0,049	0,282	0,469	0,612	0,250	0,144

Anexos

Anexo A – Termo de consentimento livre e esclarecido

Nome da pesquisa: Sequenciamento de nova geração de genes envolvidos na biossíntese de melanina em amostra da população brasileira

Pesquisador responsável: Prof. Dr. Celso Teixeira Mendes Junior

Endereço para contato: Av. Bandeirantes, 3900; CEP: 14040-901, Ribeirão Preto - SP Departamento de Química, FFCLRP-USP Telefone: (16) 3602-0417

Prezado (a) doador (a),

A pigmentação humana é determinada pela presença de um pigmento denominado melanina. Proteínas produzidas por mais de 120 genes estão envolvidas na produção de melanina em células específicas e na distribuição desta melanina para outras células de nosso organismo. Variações normais nos genes envolvidos nestes processos levam aos diferentes tons de cor de cabelos, olhos e pele que vemos nas pessoas.

Nesse estudo, faremos a análise da sequência de algumas regiões do DNA que podem estar relacionadas à determinação das características físicas dos brasileiros, como a cor de olhos, pele e cabelos, tipo de cabelos e presença de sardas. Com isso buscamos identificar as variações (polimorfismos do DNA) e os mecanismos responsáveis pela determinação da cor das pessoas. Embora o conhecimento a ser produzido neste estudo não lhe trará benefícios diretos, os resultados desta pesquisa podem trazer grande contribuição para o desenvolvimento de terapias para tratar problemas de pele.

Para a realização desta pesquisa não haverá a necessidade de procedimentos que possam comprometer sua saúde. Caso concorde em participar da pesquisa, serão coletados cerca de 10 mL de seu sangue (aproximadamente uma colher de sopa). A coleta será feita por um profissional habilitado e altamente capacitado nessa atividade. Serão ainda obtidas imagens fotográficas de regiões pequenas e bem delimitadas de seu corpo (olho, raiz dos cabelos e regiões do braço não-expostas ao sol), as quais não permitirão sua identificação, mantendo, assim, o caráter confidencial da informação. Adicionalmente, um colorímetro portátil, equipamento frequentemente usado em pesquisas dermatológicas, será usado para definir a cor de sua pele e dos seus cabelos com maior precisão, em um procedimento rápido indolor e livre de qualquer desconforto. Em todo este procedimento não existem riscos previsíveis além de um pequeno desconforto durante a coleta de sangue.

Os seguintes aspectos lhe são esclarecidos para que você decida sobre sua participação voluntária:

a) Esta pesquisa não lhe trará qualquer custo. Visando lhe proporcionar maior conforto e comodidade, a coleta será realizada uma única vez em data e local a combinar.
b) Não haverá nenhuma forma de reembolso em dinheiro, uma vez que sua participação na pesquisa não proporcionará nenhum gasto.
c) Você terá a garantia de receber a resposta a qualquer pergunta ou esclarecimento a qualquer dúvida acerca dos procedimentos, riscos, benefícios e outros relacionados com a pesquisa. Isso poderá ser feito a qualquer momento. Nós temos o compromisso de proporcionar informações atualizadas durante o estudo, ainda que esta possa influenciar a sua vontade de continuar autorizando a sua participação.
d) Você terá a liberdade de retirar o seu consentimento a qualquer momento sem que isso lhe traga prejuízo algum ou qualquer tipo de constrangimento.
e) Você terá garantida a segurança de que não será identificado e que será mantida a sua privacidade e o caráter confidencial das informações concedidas.
f) A amostra biológica será coletada com total responsabilidade e custo dos pesquisadores. A coleta de seu sangue será única e feita com material descartável, sendo necessário apenas um pequeno furo com agulha em veia do seu braço, o que pode, de fato, ser um procedimento desconfortável, mas apenas pela "picada" com a agulha.
g) O sangue doado por você não será empregado na realização de pesquisas de outra natureza sem sua autorização.
h) Será aplicado um questionário visando uma melhor compreensão de sua origem e características físicas. Toda informação obtida será mantida em sigilo.
i) Os resultados da pesquisa serão apresentados em reuniões científicas (simpósios e congressos) e serão publicados em revistas científicas especializadas.

_____

Assinatura do pesquisador responsável

Eu, _____, RG nº_____, residente na _____ nº _____, cidade de _____, telefone _____ abaixo assinado, tendo recebido as informações acima e ciente dos meus direitos, concordo em participar da pesquisa por livre e espontânea vontade. Declaro ainda que recebi uma cópia deste termo de consentimento assinada pelo pesquisador responsável.

Tendo ciência do exposto acima, assino abaixo.

Ribeirão Preto,____de_____de 20___.

_____

Assinatura do doador

Anexo B – Termo de consentimento livre e esclarecido para armazenamento de material biológico

Nome da pesquisa: Sequenciamento de nova geração de genes envolvidos na biossíntese de melanina em amostra da população brasileira.

Pesquisador responsável: Prof. Dr. Celso Teixeira Mendes Junior

Endereço para contato: Av. Bandeirantes, 3900; CEP: 14040-901, Ribeirão Preto - SP Departamento de Química, FFCLRP-USP Telefone: (16) 3602-0417

Prezado (a) doador (a),

Como você concordou em participar da pesquisa "Sequenciamento de nova geração das regiões regulatórias de dez genes envolvidos na biossíntese de melanina em amostra da população brasileira", solicito agora uma nova autorização para que possamos armazenar parte do material biológico processado, para uso em pesquisas futuras desta mesma natureza, isto é, envolvendo a análise de outros genes e marcadores genéticos (polimorfismos do DNA).

Caso concorde, seu material biológico será armazenado sob minha responsabilidade em um freezer localizado em meu laboratório, compondo Biorrepositório devidamente registrado junto ao Comitê de Ética em Pesquisa desta Instituição. Este Biorrepositório será futuramente transformado em Biobanco, conforme determinação do Conselho Nacional em Saúde. Os seguintes aspectos lhe serão esclarecidos para que você decida sobre esta autorização voluntária:

a) Você tem a liberdade de decidir se seu material será armazenado.
b) Você terá a liberdade de retirar o seu consentimento a qualquer momento sem que isso lhe traga prejuízo algum ou qualquer tipo de constrangimento.
c) O armazenamento não lhe trará qualquer custo. A amostra biológica será armazenada com total responsabilidade e custo dos pesquisadores.
d) Toda nova pesquisa a ser realizada com o material armazenado será submetida para aprovação do Comitê de Ética em Pesquisa (CEP) institucional e, quando for o caso, da Comissão Nacional de Ética em Pesquisa (CONEP).
e) Caso deseje, você tem o direito de ser consultado sempre que houver intenção de uso de seu material biológico em pesquisas futuras, para que eu autorize ou não autorize sua nova utilização.

f) Você terá a garantia de receber a resposta a qualquer pergunta ou esclarecimento a qualquer dúvida acerca do seu material biológico, condições de armazenamento e sua utilização em outros projetos de pesquisa. Isso poderá ser feito a qualquer momento. Nós temos o compromisso de proporcionar informações atualizadas durante o estudo, ainda que esta possa influenciar a sua vontade de continuar autorizando a sua participação.

g) Você terá garantida a segurança de que não será identificado e que será mantida a sua privacidade e o caráter confidencial das informações concedidas. Toda informação obtida será mantida em sigilo.

h) Você tem o direito de ser informado sobre eventual perda ou destruição de suas amostras biológicas, bem como sobre eventual encerramento do Biorrepositório.

i) Os resultados de novas pesquisas serão apresentados em reuniões científicas (simpósios e congressos) e serão publicados em revistas científicas especializadas.

_____

Assinatura do pesquisador responsável

Eu, _____, RG no _____, residente na _____ nº _____, cidade de _____, telefone _____ abaixo assinado, tendo recebido as informações acima e ciente dos meus direitos, concordo com o armazenamento de meu material biológico por livre e espontânea vontade.

Declaro ainda que:

( ) desejo exercer o direito de ser consultado sempre que houver intenção de uso de meu material biológico em pesquisas futuras.

( ) dispenso o direito de ser consultado quando houver intenção de uso de meu material biológico em pesquisas futuras, autorizando automaticamente seu uso em estudos desta mesma natureza, isto é, envolvendo a análise de outros genes e marcadores genéticos (polimorfismos do DNA).

Atesto que recebi uma cópia deste termo de consentimento assinada pelo pesquisador responsável. Tendo ciência do exposto acima, assino abaixo.

Ribeirão Preto,____de_____de 20___.

_____

Assinatura do doador

Anexo C – Protocolo de análise laboratorial (HaloPlex) para sequenciamento de nova geração

Laboratório de Pesquisas Forenses e Genômicas
Departamento de Química – FFCLRP/USP

Protocolo de análise laboratorial (HaloPlex) para sequenciamento de nova geração

Oliveira MLG, Marcorin L, Pereira ALE, Debortoli G,

Fracasso NCA, Silva GV, Mendes-Junior CT

Versão 1 – Outubro de 2016

Adaptado de: Oliveira MLG. Sequenciamento de nova geração do gene *IRF4*: identificação de variações associadas a fenótipos de pigmentação na população brasileira. Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo. 2016. 151p.

1. EXTRAÇÃO DE DNA GENÔMICO 193

2. QUANTIFICAÇÃO E VERIFICAÇÃO DA QUALIDADE DO DNA 193

3. PREPARO DOS REAGENTES PARA ENRIQUECIMENTO DE ALVO 194

4. ENRIQUECIMENTO DO ALVO E PREPARO DA BIBLIOTECA 194

4.1. Digestão e hibridação do DNA 195

4.2. Captura, ligação, eluição e PCR 196

4.3. Purificação da biblioteca 197

5. CONTROLE DE QUALIDADE E QUANTIFICAÇÃO DA BIBLIOTECA 197

6. SEQUENCIAMENTO 198

7. REFERÊNCIAS 199

1. EXTRAÇÃO DE DNA GENÔMICO

Após a coleta, a extração do DNA deve ser realizada a partir das amostras de sangue total utilizando o protocolo de *salting-out* (Miller, Dykes, & Polesky, 1988) com modificações.

O sangue coletado e mantido em tubo *Vacutainer*® com EDTA é transferido para um tubo *falcon* graduado de 50mL, ao qual é adicionado tampão de Lise I gelado (Tris-HCl 0,01M pH 7,5; Sacarose 0,3M; $MgCl_2$ 0,005M; Triton 1%; H_2O qsp) até completar o volume de 45mL. De forma a lisar as hemácias, o conteúdo é delicadamente misturado por inversão e centrifugado durante 15 minutos a 5000rpm em centrífuga 5430 R (Eppendorf AG). A extração prossegue com o botão de células obtido, o qual deve ser ressuspendido em 4,5mL de tampão de Lise II (NaCl 0,075M, EDTA 0,024M, H2O qsp, pH 8), 125µL de SDS 10% e 1,1mL de perclorato de sódio (5M) e vortexado a temperatura ambiente por 10 segundos para rompimento dos glóbulos brancos. Para lisar as proteínas, são adicionados 2mL de NaCl 6M (saturado) ao tubo *falcon*, que é vortexado a temperatura ambiente por 15 segundos. A seguir, o conteúdo é centrifugado por 15 minutos a 5000rpm em centrífuga 5430 R (Eppendorf AG). O sobrenadante é transferido para um novo tubo *falcon* de 50mL, ao qual é adicionado 5mL de isopropanol absoluto a temperatura ambiente (Merck Millipore, Darmstadt, Alemanha) e homogeneizado por inversão até que o DNA se separe e fique suspenso na solução.

Finalmente, o DNA é retirado com o auxílio de uma pipeta estéril e transferido para um microtubo de 1,5mL contendo 1mL de etanol (Merck Millipore, Darmstadt, Alemanha) 70% gelado. O microtubo é então centrifugado por 5 minutos a 5000rpm em centrífuga MiniSpin Plus (Eppendorf AG), e o sobrenadante descartado a seguir. O microtubo é deixado em centrífuga a vácuo (Concentrator plus, Eppendorf AG) durante 5min para remoção completa do etanol. Por fim, o DNA aderido ao microtubo é dissolvido em 200µL de água MilliQ, mantido na geladeira por 1 semana e posteriormente armazenado a -20ºC no banco de amostras do laboratório.

2. QUANTIFICAÇÃO E VERIFICAÇÃO DA QUALIDADE DO DNA

O *NanoDrop*® ND-1000 (*Thermo Fisher Scientific Inc.*) é utilizado para avaliar o grau de contaminação do material genômico por proteínas e para estimativa inicial da concentração de DNA. É importante que as amostras possuam uma razão OD 260/280 variando de 1,8 a 2,0. Além disso, é realizada a eletroforese em gel de agarose, a fim de verificar o grau de

degradação do material, considerando como indicativo de degradação qualquer banda abaixo de 2,5kb (*Agilent Technologies*).

Na sequência, as amostras são quantificadas através de fluorescência por meio do *Qubit™ dsDNA BR Assay* (*Life Technologies*). Para isso, o DNA genômico é diluído em água MilliQ em uma proporção de 1:10. Inicialmente, é preciso fornecer ao instrumento uma curva padrão. Para tanto, o corante (fluoróforo) para dsDNA é diluído no *buffer* do Qubit (1:200) e 10µL dos padrões (1 e 2) são diluídos em 200µL da solução anterior. Realizada a leitura dos padrões, uma curva padrão é estabelecida, a partir da qual é calculada a concentração da amostra original, com base no volume adicionado para leitura.

Conhecendo agora a concentração das amostras, essas são diluídas à concentração de 5ng/µL, no intuito de normalizar as amostras a uma concentração compatível para a montagem das bibliotecas de fragmentos.

3. PREPARO DOS REAGENTES PARA ENRIQUECIMENTO DE ALVO

Regiões alvo, incluindo a sequência promotora, CDS (*Coding DNA Sequence*), 5'UTR e 3'UTR de genes humanos, além de outras regiões de interesse, foram submetidas ao ensaio *in silico* de desenho de sondas para captura de DNA alvo utilizando o aplicativo *SureDesign* da *Agilent Technologies* (https://earray.chem.agilent.com/suredesign/home.htm). Para desenho das sondas, o *SureDesign* tomou como referência as coordenadas dos genes em questão junto ao *release* hg19/GRCh37 do genoma humano. Visando a cobertura de 284.367pb, um total de 12.131 sondas foi desenhado, abrangendo 99,63% de tal extensão. Devido à distribuição dos sítios de restrição utilizados posteriormente no preparo das bibliotecas, as sondas desenhadas capturam uma região de 488.658pb, que garante o sequenciamento parcial dos introns flanqueadores dos exons alvo.

4. ENRIQUECIMENTO DO ALVO E PREPARO DA BIBLIOTECA

Após a quantificação, cada amostra de DNA é processada individualmente, gerando uma biblioteca contendo todos os amplicons. A captura das regiões-alvo utiliza reagentes do kit de enriquecimento de alvo customizado *HaloPlex* (*Agilent Technologies*) e segue o protocolo sugerido pelo fabricante. Resumidamente, o protocolo consiste dos seguintes passos: 1) digestão do DNA genômico com enzimas de restrição; 2) hibridação dos

fragmentos de DNA a sondas biotiniladas cujos finais são complementares a esses fragmentos-alvo (durante esse passo os fragmentos são circularizados e índices, que irão identificar cada indivíduo durante o sequenciamento, são adicionados); 3) captura dos fragmentos-alvo utilizando *beads* magnéticas recobertas por estreptavidina; 4) ligação dos fragmentos circularizados; 5) amplificação por PCR dos fragmentos-alvo capturados para montagem da biblioteca.

4.1. Digestão e hibridação do DNA

A digestão das amostras é feita conforme indicações do fabricante, com algumas modificações (*HaloPlex Target Enrichment System For Illumina Sequencing, Agilent Technologies*, 2013). Um total de 225ng (45µL de uma solução 5ng/µL) de DNA genômico de cada amostra e do *Enrichment Control DNA* (ECD), no volume total de 45µL, é separado e reservado em gelo.

Digestão: Para o mix de restrição, preparado para lotes de 12 amostras (11 amostras de DNA genômico e outra de ECD), são misturados 476µL de *RE buffer* e 11,9µL de BSA e esse conteúdo é distribuído em 8 tubos. Em cada um dos 8 tubos é adicionado um par de enzimas de restrição (7µL de cada). Cada amostra é digerida pelos 8 pares de enzima separadamente. Para isso, cada amostra é distribuída em 8 tubos (5µL por tubo) onde são adicionados 5µL dos mixes contendo as enzimas de restrição resultando, para um total de 12 amostras, em 96 tubos de digestão que são incubados no termociclador (*SureCycler 8800 Thermal Cycler*, *Agilent Technologies*) a 37ºC por 30 minutos.

Hibridação: Para o mix de hibridação são misturados 650µL da *Hybridization Solution* e 260µL das sondas *Haloplex* em um tubo e 70µL dessa solução é distribuída por amostra. Adicionalmente, são diluídos nessa solução 10µL de um dos índices disponíveis em cada um dos doze tubos (índices diferentes são utilizados para cada amostra, permitindo sua identificação após a mistura das amostras para o sequenciamento).

O DNA digerido de cada amostra (80µL somando todos os 8 tubos de digestão) é então transferido para os microtubos contendo o mix de hibridação (70µL) e seus índices correspondentes (10µL). Essas soluções são incubadas a 54ºC durante 16 horas. O tempo de duração da hibridação é determinado de acordo com a densidade de sondas do ensaio. Segundo instruções do fabricante (*Haloplex*, *Agilent Technologies*), ensaios que atingem cobertura de entre 500kb - 5Mb exigem 16 horas de hibridação, enquanto que ensaios com

cobertura de entre 1-500kb exigem apenas 3 horas de hibridação. Apesar de o presente ensaio estar voltado para a cobertura de apenas 488,7kb, os ensaios iniciais demonstraram a necessidade de um tempo de hibridação de 16 horas.

4.2. Captura, ligação, eluição e PCR

Captura: A captura dos fragmentos hibridados é feita com o uso de 520µL de *beads* magnéticas recobertas por estreptavidina. Com o auxílio de uma placa magnética (*DynaMag™-2 Magnet, Life Technologies*), o sobrenadante da solução de *beads* é removido e substituído por 520µL de *Capture Solution*. Deste conteúdo, 40µL é adicionado a cada solução de DNA hibridizado (160µL) e após 15min de incubação em temperatura ambiente, os tubos contendo as soluções são colocados junto a uma placa magnética de 96 poços (*Agencourt SPRIPlate 96R, Beckman Coulter*) para a remoção do sobrenadante. A seguir, adiciona-se 100µL de *Wash Solution* por amostra, e essa solução é incubada no termociclador a 46ºC por 10min. Após esse período, o sobrenadante é novamente removido.

Ligação: Procedendo com a ligação e circularização dos fragmentos alvo, adiciona-se 617,5µL de *Ligation Solution* e 32,5µL de DNA ligase e distribuí-se 50µL dessa solução a cada amostra hibridada e capturada. A solução é então incubada a 55ºC por 10min.

Eluição: A seguir, lavamos o conteúdo de DNA ligado com 100µL de *SSC buffer*, retirando o sobrenadante ao final da etapa com o auxílio da placa magnética. Para eluir o DNA das *beads* adicionamos 25µL de NaOH (50mM) e encubamos em temperatura ambiente por 1min..

PCR: Com o auxílio da placa magnética, coletamos o DNA eluído num total de 20µL por amostra e o transferimos para tubos contendo mix de PCR. Tal mix para a reação de PCR é preparado com 209,3µL de água MilliQ, 130µL de 5X *Herculase II Reaction Buffer*, 5,2µL de dNTPs (100mM, 25mM para cada dNTP), 13µL de cada um dos *primers forward* e *reverse* (25µM), 6,5µL de ácido acético (2M) e 13µL da polimerase *Herculase II Fusion*, somando um total de 30µL por amostra. O volume final da reação é de 50µL por amostra. As condições de ciclagem da PCR seguem uma etapa inicial de desnaturação por 10 minutos a 98ºC, seguida por 20 ciclos de 98ºC por 30 segundos, 60ºC por 30 segundos e 72ºC por 1 minuto, e um passo final a 72ºC por 10 minutos. Após o término da reação de PCR o produto é mantido a 8ºC.

4.3. Purificação da biblioteca

Os produtos de PCR são então purificados utilizando *AMPure XP beads* (*Beckman Coulter*) de acordo com as instruções da *Agilent Technologies*. Um total de 40µL do produto de PCR é misturado a 100µL de *AMPure XP beads* mais 40µL de água MilliQ por amostra. A seguir, essa solução é colocada junto à placa magnética para remoção do sobrenadante e é lavada duas vezes com álcool 70%, que é completamente removido ao final. Então, adicionamos 40µL de Tris-HCl (pH = 8,0) para eluir o DNA capturado pelas *AMPure XP beads*. O produto final de 40µL é armazenado a -20 ºC.

5. CONTROLE DE QUALIDADE E QUANTIFICAÇÃO DA BIBLIOTECA

A qualidade das bibliotecas é um dos principais determinantes para o sucesso da corrida de sequenciamento. Bibliotecas de fragmentos que não correspondem à distribuição de tamanho recomendada podem levar a uma baixa cobertura ou à falha de cobertura em determinadas porções das sequências alvo (Cher, 2011).

O controle de qualidade das bibliotecas é feito através do sistema de eletroforese capilar *2100 Bioanalyzer* (*Agilent Technologies*) utilizando um chip de alta-sensibilidade. A eletroforese é capaz de separar os fragmentos com base no tamanho, permitindo dessa forma avaliar a qualidade das bibliotecas geradas. É esperado para este ensaio que as bibliotecas apresentem um perfil de picos variando de 175 a 625pb, conforme previsto por um eletroferograma padrão (Figura 1). A partir da seleção manual da faixa de fragmentos gerados (175 a 625pb) é possível obter a concentração (em pg/µL e pmol/L) referente a este intervalo, sendo a molaridade utilizada para a quantificação de cada uma das bibliotecas para composição adequada do *pool* de bibliotecas para o sequenciamento. Adicionalmente, as bibliotecas individuais podem ser quantificadas por fluorescência (*Qubit™ dsDNA BR Assay*, *Life Technologies*) para composição adequada do *pool* de bibliotecas para o sequenciamento.

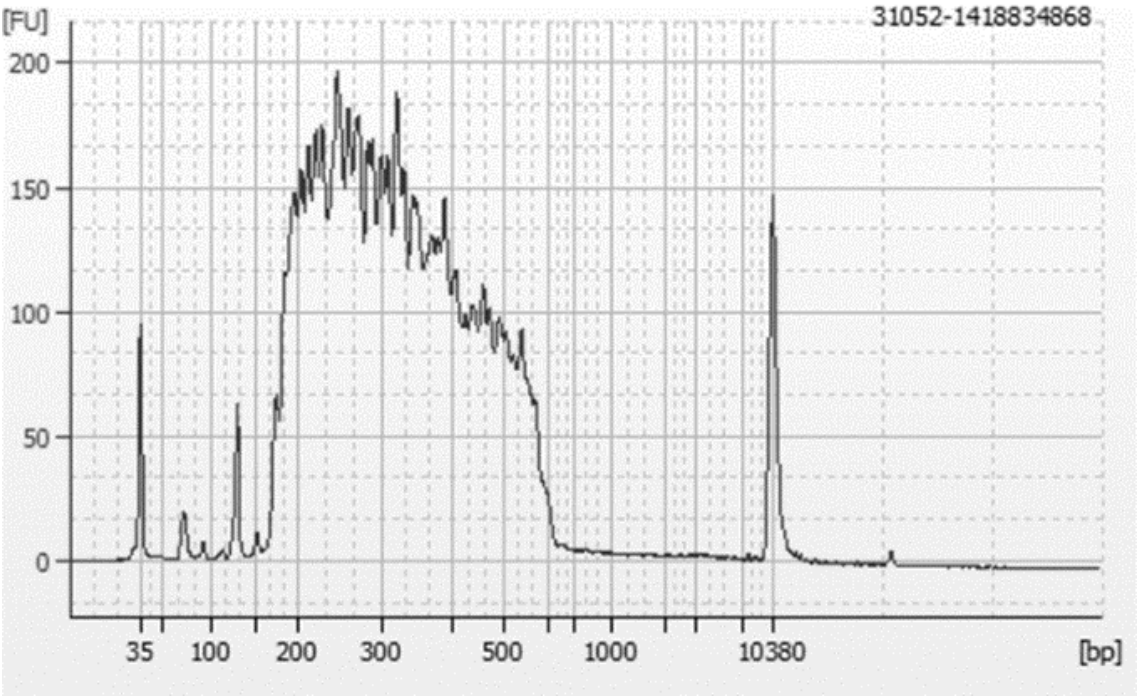


Figura 8: Perfil de picos esperado no eletroferograma para as bibliotecas de fragmento deste ensaio. Tal eletroferograma é produzido pela *Agilent Technologies* durante a avaliação do kit *Haloplex* customizado por meio de ensaio *in silico* de desenho de sondas para captura de DNA alvo utilizando o aplicativo *SureDesign*. Para leitura correta no *2100 Bioanalyzer* é crítico que os 13 picos previstos no *ladder* sejam identificados, que os picos do menor (35pb) e do maior (10380pb) marcador apresentem um tempo de corrida correspondente entre as amostras analisadas e o *ladder* e que as linhas de base sejam planas (Cher, 2011).

6. SEQUENCIAMENTO

Como descrito no item anterior, a partir das concentrações obtidas no *2100 Bioanalyzer* e *Qubit*, prepara-se soluções a 4nM de biblioteca, de maneira que cada biblioteca tenha a mesma representatividade. O *2100 Bioanalyzer* é utilizado também para quantificação deste *pool* final de bibliotecas. O *pool* final de bibliotecas pode ser adicionalmente quantificado por fluorescência (*Qubit™ dsDNA BR Assay*, *Life Technologies*) ou PCR em tempo real (*ViiA™ 7 Real-Time PCR System*, *Applied Biosystems*). Na PCR quantitativa em tempo real (qPCR), através da construção de uma curva padrão, é possível determinar a quantidade absoluta de DNA nas amostras a serem sequenciadas (*Applied Biosystems ViiA™ 7 Real-Time PCR System. User Guide. Life Technologies*, 2010, 262p).

A seguir, 5µL do *pool* de bibliotecas a 4nM são desnaturados pela adição de 5µL de NaOH (0,2M) e 990µL *de Hybridization Buffer* (HT1) para obtenção de uma biblioteca a 20pM, que é diluída com HT1 para que se chegasse à 600µL de uma solução de concentração

16pM. Essa solução é inserida no cartucho de sequenciamento do *MiSeq Personal Sequencer* (*Preparing Libraries for Sequencing on the MiSeq*, California, USA, *Illumina*, Inc, 2013, 14p).

Antes de iniciar o sequenciamento é preciso carregar uma *SampleSheet* ou planilha de amostras (*.csv) contendo as informações necessárias para configurar a corrida, as quais correspondem à lista de amostras a serem sequenciadas e às sequências dos índices utilizados em cada amostra.

Durante o sequenciamento, as moléculas de DNA fita simples se ligam à superfície da *flow cell* por meio dos adaptadores presentes em suas extremidades. O sequenciamento ocorre a partir da síntese da fita complementar ao fragmento ligado na *flow cell* utilizando didesoxinucleotídeos marcados com fluoróforos específicos para cada base (A, T, C e G). O *software* Bustard relaciona então a fluorescência emitida após cada incorporação ao didesoxinucleotídeo correspondente com o *cluster* de fragmentos fixados à região que emitiu o sinal. No caso do sequenciamento usando reagentes *MiSeq Reagent Kit V3* (*600 cycle*), são realizados 301 ciclos em leituras do tipo *paired-end* (cada fragmento é sequenciado tanto na fita *forward* quanto na *reverse*) perfazendo um total de 2x301 ciclos de leitura, além de 8 ciclos referentes à leitura do índice incorporado em cada fragmento capturado, o que relaciona de maneira unívoca o fragmento a uma determinada amostra.

7. REFERÊNCIAS

Cher, C. (Field A. S. (2011). Bioanalyzer Applications for Next Gen Sequencing: Updates and Tips, 1–49.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*(3), 1215. https://doi.org/10.1093/nar/16.3.1215

Anexo D – **Protocolo de processamento computacional de dados de sequenciamento de nova geração**

Laboratório de Pesquisas Forenses e Genômicas
Departamento de Química – FFCLRP/USP

Protocolo de processamento computacional de dados de sequenciamento de nova geração

Oliveira MLG, Marcorin L, Pereira ALE, Debortoli G,

Fracasso NCA, Silva GV, Castelli EC, Mendes-Junior CT

Versão 2 – abril de 2018

Adaptado de: Oliveira MLG. Sequenciamento de nova geração do gene *IRF4*: identificação de variações associadas a fenótipos de pigmentação na população brasileira. Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo. 2016. 151p.

1. ANÁLISE DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO 201

1.1. Análise primária 201

1.2. Análise secundária 202
1.2.1. Avaliação da qualidade 202
1.2.2. Trimagem 202
1.2.3. Alinhamento ao genoma de referência 204
1.2.3.1. SAM 206
1.2.3.2. BAM 207

1.3. Análise terciária 208
1.3.1. Determinação de variantes 208
1.3.1.1. VCF 212
1.3.2. Remoção de genótipos de baixa qualidade 214

2. REFERÊNCIAS 216

1. ANÁLISE DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO

O fluxograma de análises no NGS pode ser sumarizado em três fases principais: a fase primária, que consiste na geração dos dados no formato FastQ, a fase secundária, que compreende a aplicação de filtros de qualidade e o alinhamento a um genoma de referência e a fase terciária, que compreende o processo de determinação das variantes através de comparação das sequências alvo a um genoma de referência.

1.1. Análise primária

O processamento dos dados segue passos gerais pré-estabelecidos, de acordo com o objetivo pretendido, e tem início no *base calling* (ou a interpretação da imagem obtida a partir do sequenciador), quando são gerados os dados brutos do sequenciamento no formato *base call* (*.bcl). Estes dados são então disponibilizados através do ambiente de computação em nuvem da *Illumina*, *BaseSpace* (https://basespace.illumina.com), no formato FastQ (*.fastq.gz), o qual possui uma pontuação de qualidade associada (*Q-score*) a cada base sequenciada.

Essa pontuação de qualidade, também denominada *Phred-score* por ter sido introduzida como convenção pelo programa Phred (Ewing, Hillier, Wendl, & Green, 1998), se relaciona à probabilidade de erro na determinação de uma base. A medida se baseia na fórmula a seguir, onde dado a determinação de uma base X, a probabilidade de que X não seja verdadeira (P(X)) é expressa pela relação $Q(X) = -10 \log_{10} (P(X))$ ou $P(X) = 10^{-Q(X)/10}$. Assim, um Q30 indica probabilidade de ocorrência de um erro a cada 1.000 bases determinadas.

Cada arquivo FastQ é nomeado como no exemplo a seguir, onde PIG0001 representa o nome da amostra, S1 representa o número da amostra e indica neste exemplo que esta é a primeira amostra listada na *SampleSheet*, L001 indica o número da linha que a amostra ocupa na *flow cell*, sendo sempre o mesmo, já que a *flow cell* utilizada possui apenas uma linha, R1 e R2 indicam que o arquivo é composto por *reads forward* e *reverse*, respectivamente, enquanto o último segmento 001 sempre recebe essa denominação:

```
PIG0001_S1_L001_R1_001.fastq.gz
PIG0001_S1_L001_R2_001.fastq.gz
```

1.2. Análise secundária

Essa segunda etapa consiste na aplicação de filtros de qualidade e posterior alinhamento dos fragmentos sequenciados a um genoma de referência.

1.2.1. Avaliação da qualidade

Em um primeiro momento, é importante verificar a qualidade das *reads* sequenciadas, a fim de garantir que os dados não apresentem problemas que poderiam afetar o passo seguinte de mapeamento. Para tanto, pode ser empregado o *software* FastQC (Andrews S., 2010), o qual utiliza um arquivo FastQ para gerar um relatório completo da qualidade das sequências (Figura 1a). A partir dos resultados apresentados é possível decidir por estratégias que melhorem a qualidade dos dados gerados.

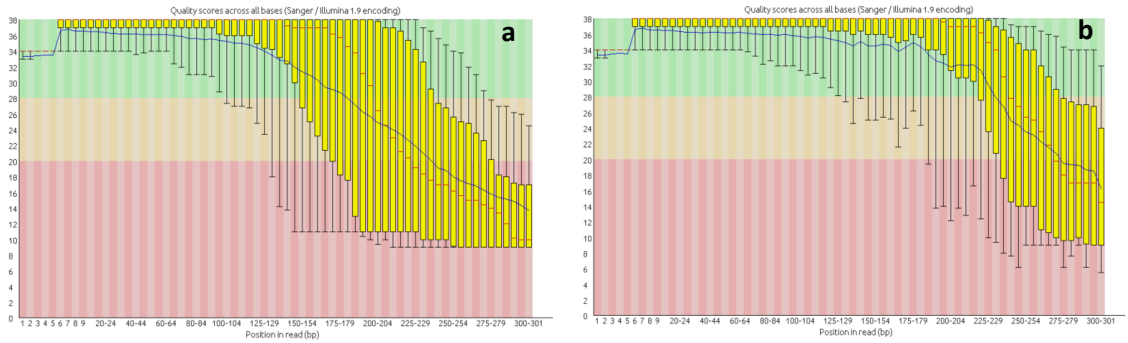


Figura 1: Gráfico de distribuição da qualidade por base no programa FastQC. Cada coluna apresenta um *boxplot* com os valores de qualidade média das bases incorporadas em todas as *reads* em um determinado ciclo. Como pode ser observado em (a) ocorre uma diminuição da qualidade (Q<30) das *reads* não-processadas a partir dos ciclos 150-154. O eixo Y apresenta os *scores* de qualidade, sendo dividido conforme cores que indicam: verde (boa qualidade), laranja (qualidade razoável) e vermelho (baixa qualidade). (b) Após processadas pelo *cutadapt*, foi obtida uma maior proporção das *reads* com qualidade desejável, ainda que a partir dos ciclos 225-229 tenha ocorrido uma diminuição da qualidade, o que, no entanto, é comum para ensaios da *Illumina*.

1.2.2. Trimagem

Sequências de adaptadores, bem como bases diagnosticadas com baixo *Q-score* devem ser retiradas das extremidades das *reads*, em um processo referido como trimagem. As sequências obtidas podem ser trimadas de diversas maneiras: por tamanho, qualidade ou a partir de sequências conhecidas.

A trimagem por qualidade parte do princípio de que é preferível trabalhar com sequências menores de maior qualidade, removendo sequências de bases na extremidade da *read* que tenham uma qualidade abaixo do limite estabelecido como adequado. A partir do que pôde ser observado no relatório gerado pelo FastQC optou-se por realizar a trimagem apenas pelos adaptadores *paired-end TruSeq da Illumina* utilizados no sequenciamento.

No processo de sequenciamento *Illumina*, todos os *clusters* são sequenciados seguindo o número de ciclos previstos pelo kit de reagentes utilizado (no caso, 2x301 ciclos). Mesmo que um *cluster* inclua um pequeno fragmento a ser sequenciado, os ciclos de sequenciamento não são interrompidos, continuando até alcançar o número total de ciclos. Como consequência, as *reads* geradas como *output* podem conter, além das sequências alvo (caso essas sejam menores que o número de ciclos), as sequências dos adaptadores do outro extremo do fragmento. Caso ainda restem ciclos após o sequenciamento dos adaptadores, leituras adicionais continuam sendo feitas, embora com qualidade extremamente baixa, uma vez que nucleotídeos inexistentes são aleatoriamente identificados (possivelmente devido à fluorescência de diferentes nucleotídeos não incorporados e não retirados durante a lavagem). Logo, é preciso encontrar e remover esses adaptadores, o que resultará, consequentemente, na remoção das bases incorporadas aleatoriamente.

Em função da construção realizada pelo *Haloplex* (*PCR primer* --- *Illumina adaptor* --- *TARGET* --- *Illumina adaptor* --- *Barcode* --- *PCR primer*), ao remover as sequências dos adaptadores de uma *read* podemos remover tudo o que existe depois dessas sequências, neste caso os *primers* da PCR e o *barcode,* quando presente.

A trimagem pelos adaptadores pode ser realizada com uso do *cutadapt* (Martin, 2011) utilizando a seguinte linha de comando (Tabela 1):

```
cutadapt -e 0.2 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -o
sample.trimmed.r1.fastq -p sample.trimmed.r2.fastq
input.raw.r1.fastq input.raw.r2.fastq
```

Tabela 1: Descrição dos componentes da linha de comando executada no *cutadapt*

Comandos utilizados no *cutadapt*	**Função do comando executado**
-e 0.2	indica o nível máximo de tolerância de erros (*mismatches*, inserções e deleções) entre as sequências dos adaptadores e sequências presentes nas *reads*, para que uma sequência adaptadora seja efetivamente trimada.
-a	indica que o parâmetro a seguir se refere à sequência do adaptador *forward.*
-A	indica que o parâmetro a seguir se refere à sequência do adaptador *reverse.*
-o	indica o arquivo *output* da primeira *read* do par.
-p	indica o arquivo *output* da segunda *read* do par.

Tal procedimento resultou em melhorias significativas na qualidade média das *reads* processadas (Figura 1b).

1.2.3. Alinhamento ao genoma de referência

Estando as *reads* processadas, o próximo passo é mapeá-las contra um genoma de referência. A maior parte dos programas de alinhamento disponíveis usam um dentre os seguintes algoritmos de mapeamento: baseado em *hash* ou na transformação de Burrows-Wheeler (BWT). Estes algoritmos irão construir um arquivo de alinhamento com base nos dados de *single-end* e em uma análise seguinte utilizarão os dados de *paired-end* para refinar os alinhamentos gerados.

Para fazer o alinhamento pode ser utilizado o BWA (Li & Durbin, 2009), o qual utiliza um algoritmo do tipo BWT, que tem como principal vantagem o menor consumo de memória. O BWA é um pacote de softwares, que consiste em três algoritmos, que são chamados por diferentes comandos: `aln/samse/sampe` para o BWA-backtrack, `bwasw` para o BWA-SW e `men` para o BWA-MEM.

Antes de realizar o alinhamento, porém, é preciso indexar o genoma de referência por meio do índice FM (Ferragina, Manzini, Veli, & Navarro, 2004). Neste caso, utilizamos um genoma já indexado (*Homo sapiens*, UCSC, hg19), disponibilizado pela *Illumina* através do *iGenomes* em: https://support.illumina.com/sequencing/ sequencing_software/igenome.html. Pode-se também indexar localmente o genoma de interesse por meio de ferramentas

específicas. Para tanto, é necessário gerar três diferentes arquivos, arquivo índice, arquivo fasta índice e um dicionário de sequência, os quais podem ser obtidos por meio dos programas BWA, Samtools e Picard, respectivamente.

O algoritmo utilizado para realizar o alinhamento é o BWA-MEM em função da sua especificidade para sequências com tamanho entre 70 e 1Mb e por ser mais rápido e acurado que o BWA-SW. A seguinte linha de comando deve ser utilizada (Tabela 2):

```
bwa mem -t 6 -R
'\@\R\G\\tID:sampleID\tLB:sampleLB\tSM:tag\tPL:illumina\tPU:sampleP'
/path/to/reference/hg19.fa
sample.trimmed.r1.fastq sample.trimmed.r2.fastq
> mapped.sample.sam
```

Tabela 2: Descrição dos componentes da linha de comando executada no BWA.

Comandos utilizados no BWA	**Função do comando executado**
-t	*number of threads* (número de processos com um endereço/espaço compartilhado no *Linux*).
-R	indica que o parâmetro a seguir se refere à linha do cabeçalho de determinado grupo de *reads.*
\t	comando convertido em TAB no output SAM, a fim de separar adequadamente os campos.
\@\R\G\\	identifica o cabeçalho de cada grupo de *reads*. Permite ainda que o arquivo BAM possa ser lido pelo GATK.
ID	identifica a qual grupo de *reads* cada *read* pertence. O ID permite que ao invés de ter que lidar com múltiplos grupos de dados, se passe a lidar com apenas um. Assim, a informação sobre o grupo de *reads* permite a identificação dos dados de diferentes experimentos, ainda que eles estejam combinados em um arquivo único. Cada linha @RG deve conter um ID único. Logo, todas as *reads* de um mesmo grupo são consideradas como parte da mesma corrida de sequenciamento e indicam a linha que ocupam nessa corrida. Por exemplo: ID: FLOWCELL1. LANE2.
LB	identificador da biblioteca de DNA.
SM	nome da amostra sequenciada em determinado grupo de *reads*. Todo dado que tiver um mesmo valor SM será tratado como pertencendo a uma mesma amostra.

Continuação Tabela 2: Descrição dos componentes da linha de comando executada no BWA.

Comandos utilizados no BWA	Função do comando executado
PL	identifica a plataforma de sequenciamento utilizada.
PU	denominação alternativa do grupo de *reads*. Mantém a informação sobre a linha onde a amostra se encontra na *flow cell*. Essa designação não é exigida pelo GATK, mas prevalece sobre o ID caso seja executada uma recalibração de base.
/path/to/reference/hg19.fa	identifica o caminho a ser seguido para consulta ao genoma de referência
sample.trimmed.r1.fastq	*input* para o mapeamento, contendo a primeira *read* do par
sample.trimmed.r2.fastq	*input* para o mapeamento, contendo a segunda *read* do par
> mapped.sample.sam	indica o redirecionamento do resultado do mapeamento para o arquivo SAM especificado.

Como *output* do comando acima é gerado um arquivo SAM (*.sam). Porém, uma vez que os arquivos SAM são arquivos de texto que ocupam muito espaço no disco rígido, e para que as informações sobre mapeamento pudessem ser visualizadas, esse arquivo pode ser convertido para o formato BAM (*.bam).

1.2.3.1. SAM

O formato SAM (*Sequencing Alignment/Map*) (Li et al., 2009) é um arquivo baseado em texto, utilizado como *output* por diversos programas de alinhamento. Cada linha no arquivo SAM armazena informações sobre uma *read* mapeada contra um genoma de referência.

Arquivos SAM possuem uma seção com o cabeçalho (que é opcional) e uma seção de alinhamento. As linhas no cabeçalho iniciam com "@" seguido por um código de duas letras. Por sua vez, cada *read* alinhada à referência é representada em uma linha da seção de alinhamento.

Podem estar presentes ainda linhas opcionais com marcações predefinidas (*The SAM/BAM Format Specification Working Group, Sequence Alignment/Map Format Specification*, 2015, http://samtools.github.io/hts-specs/SAMv1.pdf). Algumas destas marcações são geradas pelo BWA (aquelas começando com "X").

1.2.3.2. BAM

O formato BAM (*Binary Alignment/Map*) é a versão binária compactada de um arquivo SAM, indexável e amplamente utilizado para representar as sequências após o alinhamento. A compactação em blocos (chamada BGZF) de um arquivo BAM permite além da diminuição do espaço de armazenamento o carregamento parcial do arquivo, o que é bastante utilizado em programas de visualização como o IGV ou o *UCSC Genome Browser*.

A conversão do arquivo SAM em BAM gera um arquivo com as mesmas informações e pode ser realizada através do programa *SAMtools* (http://samtools.sourceforge.net/), o qual fornece diversas ferramentas para manipulação desses arquivos.

Para visualizar o arquivo BAM e permitir sua utilização pelos *softwares* de determinação de variantes, é preciso ordenar os alinhamentos de acordo com a posição genômica. O comando a seguir é utilizado para realizar a conversão de SAM em BAM e a ordenação do arquivo BAM:

```
samtools sort mapped.sample.sam sorted.mapped.sample.bam
```

O nome do arquivo gerado pelo comando seguiu o formato a seguir, onde PIG0001 representa o nome da amostra e *sorted* indica que o arquivo BAM gerado se encontra ordenado:

```
PIG0001.sorted.bam
```

Arquivos BAM armazenam as mesmas informações que um arquivo SAM e possuem uma seção de cabeçalho, que contém informações gerais sobre o arquivo como nome, tamanho da amostra, método de alinhamento utilizado e uma seção com o alinhamento

propriamente dito contendo nome, sequência e qualidade da *read*, além de marcações personalizadas.

Os arquivos BAM devem ainda ser indexados, permitindo uma pesquisa rápida no arquivo ordenado. Um índice de correspondência do arquivo BAM em questão em relação às coordenadas de referência será identificado pela terminação (*.bam.bai). O comando a seguir permitiu gerar um arquivo indexado:

```
samtools index sorted.mapped.sample.bam
```

O *output* do comando acima pode ser nomeado como a seguir:

```
PIG0001.sorted.bam.bai
```

Uma vez ordenados e indexados, os arquivos BAM puderam ser visualizados através de um visualizador externo. Neste trabalho utilizamos o *Integrative Genomics Viewer* (IGV) 2.3.67, um programa em java disponibilizado pelo *Broad Institute* (Robinson et al., 2011; Thorvaldsdottir, Robinson, & Mesirov, 2013).

1.3. Análise terciária

1.3.1. Determinação de variantes

Em um projeto de sequenciamento, o passo seguinte ao mapeamento é a detecção das diferenças entre as regiões genômicas sequenciadas e o genoma de referência (determinação de variantes). Um programa amplamente utilizado para análise dos dados de NGS é o *Genome Analysis Toolkit* (GATK) v.3.5, disponibilizado pelo *Broad Institute* (DePristo et al., 2011; McKenna et al., 2010; Van der Auwera et al., 2013).

O programa apresenta uma estrutura modular e é dividido em diversas análises, as quais são reunidas junto a um *framework* em um arquivo java (*.jar). Para ter acesso à lista de análises disponíveis é utilizado o comando:

```
java -jar GenomeAnalysisTK.jar
```

Já para realizar uma análise específica é utilizado o comando:

```
java -jar GenomeAnalysisTK.jar -T <analysis> <arguments*>
```

Para realizar a determinação de SNPs e indels as ferramentas *HaplotypeCaller* com o parâmetro --emitRefConfidence GVCF e *UnifiedGenotyper* do GATK podem ser utilizadas. O *HaplotypeCaller* considera simultaneamente todas as amostras a serem analisadas e, ao encontrar uma região apresentando sinais de variação (região ativa), o programa descarta o mapeamento preexistente e realiza uma remontagem das *reads* naquela região. O programa realinha então cada sequência contra a sequência de referência com o objetivo de identificar potenciais sítios de variação.

Comparado ao *UnifiedGenotyper,* o *HaplotypeCaller* possibilita a nomeação de SNPs e indels de forma simultânea. Isso permite que o *HaplotypeCaller* seja mais preciso quando lidando com regiões reconhecidamente problemáticas, como por exemplo, onde são encontrados diferentes tipos de variações próximas umas das outras (Van der Auwera et al., 2013).

Por outro lado, o *UnifiedGenotyper* possui filtros menos restringentes que possibilitam encontrar variantes com frequências populacionais mais baixas, principalmente em áreas de baixa cobertura, como *singletons* e alelos raros em polimorfismos multialélicos. Porém, essa característica pode resultar em uma taxa alta de falsos positivos. Além disso, essa ferramenta faz a leitura das sequências ponto a ponto, não sendo aplicável à identificação de indels. Por essas e outras razões, o próprio site do GATK descontinuou as atualizações e aposentou a ferramenta *UnifiedGenotyper* recomendando o uso do *HaplotypeCaller* em seu lugar.

Devido a utilização do parâmetro --emitRefConfidence GVCF, é gerado para cada amostra separadamente um arquivo gVCF (do inglês, *genomic variant calling format*), onde estão dispostas um registro para cada posição (ou intervalo de interesse) independentemente de uma variante ser detectada naquele sítio ou não. O uso do gVCF é importante pois permite a adição de outros conjuntos de amostra, caso necessite realizar uma análise conjunta. Desta forma, pode-se adicionar os arquivos gVCFs de diferentes coortes e realizar a análise de

chamada de variantes simultaneamente nestes conjuntos de amostra, aumentando a acurácia nesta determinação.

As seguintes linhas de comando são utilizadas para realizar a nominação das variantes (Tabela 3):

HaplotypeCaller:

```
java -jar GenomeAnalysisTK.jar \
     -R reference.fasta \
     -T HaplotypeCaller \
     -I listofsamples.list \
     --emiteRefConfidence GVCF \
     --dbsnp dbSNP.vcf \
     -L targets.interval_list \
     -o output.raw.snps.indels.g.vcf \
     --dontUseSoftClippedBases \
     -drf DuplicateRead
```

UnifiedGenotyper:

```
java -jar GenomeAnalysisTK.jar \
     -T UnifiedGenotyper \
     -R reference.fasta \
     -I sample1.bam [-I sample2.bam ...] \
     --dbsnp dbSNP.vcf \
     -o snps.raw.vcf \
     -stand_call_conf [50.0] \
     -stand_emit_conf 10.0 \
     [-L targets.interval_list]
```

Tabela 3: Descrição dos componentes da linha de comando executada no GATK

Comandos utilizados no GATK	**Função do comando executado**
-R	indica o arquivo com o genoma de referência
-T	indica qual ferramenta do GATK será utilizada, isto é, *HaplotypeCaller* ou *UnifiedGenotyper*
-I	indica arquivo de texto com lista de amostras (formato BAM) a ser utilizado pelo GATK
--dbsnp	indica arquivo (dbSNP.vcf) contendo a identificação dos rs de cada ponto de variação oficialmente descrito, a ser utilizada para preenchimento da coluna ID do *output*
-stand_call_conf	é o limiar mínimo de confiança (na escala Phred) utilizado para separar determinações de genótipo realizadas com maior e menor confiança. Apenas determinações de genótipo emitidas com confiança ≥ limiar mínimo são

	consideradas. O valor default recomendado é de 30

Continuação Tabela 3: Descrição dos componentes da linha de comando executada no GATK

Comandos utilizados no GATK	Função do comando executado
-stand_emit_conf	permite emitir determinações de genótipo com baixa qualidade como resultados filtrados
--emitRefConfidence GVCF	indica a produção do arquivo g.vcf para o *HaplotypeCaller*
-L	direciona o GATK a restringir o processamento dos dados aos intervalos genômicos especificados em um arquivo "*targets.interval_list*", que indica o cromossomo e as coordenadas das posições de início e fim em relação ao genoma de referência
-o	indica o *output*, que é um arquivo VCF ou gVCF (*Variant Call Format*) contendo os sítios que o algoritmo (*UnifiedGenotyper* ou *HaplotypeCaller*, respectivamente) identificou como sendo variáveis
--dontUseSoftClippedBases	direciona o programa a não analisar bases *soft clipped* (bases não alinhadas nas extremidades das *reads*), o que diminui a determinação de falso-positivos
-drf	desabilita a aplicação de filtros à análise (*DuplicateRead* neste caso).

Após gerar os arquivos g.vcf para cada amostra um passo é realizado para combinar os arquivos em um único arquivo vcf com a linha de comando:

```
java -jar GenomeAnalysisTK.jar \
     -T GenotypeGVCFs \
     -R reference.fasta \
     -I listofsamples.list \
     -o cohort.g.vcf
```

Tabela 4: Descrição dos componentes da linha de comando executada no GATK

Comandos utilizados no GATK	Função do comando executado
-R	indica o arquivo com o genoma de referência
-T	indica qual ferramenta do GATK será utilizada, isto é, *GenotypeGVCFs*
-I	indica arquivo de texto com lista de amostras (formato BAM) a ser utilizado pelo GATK
-o	indica o *output*, que é um arquivo gVCF (*Variant Call Format*) contendo os sítios que o algoritmo (*HaplotypeCaller*) identificou como sendo variáveis

1.3.1.1. VCF

O VCF é um formato de arquivo que contém informações sobre variantes encontradas em posições específicas em relação a um genoma de referência. O arquivo VCF possui um cabeçalho, bastante útil, uma vez que contém uma descrição para os campos contidos no arquivo (Tabela 5):

Tabela 5: Descrição dos componentes da seção de cabeçalho de um arquivo VCF

Informações presentes no cabeçalho	Descrição das abreviações apresentadas
Fileformat	versão do formato de arquivo e a versão do *variant caller* utilizado
FILTER	identifica quais filtros são aplicados aos dados
FORMAT e INFO	essas linhas definem as anotações contidas nas colunas correspondentes a essas informações no arquivo VCF
GATKCommandLine	contém a linha de comando utilizada pelo GATK para realizar a genotipagem das amostras. Essa linha de comando especifica todos os parâmetros utilizados pelo *variant caller*, incluindo os nomes do arquivo com o genoma de referência e do arquivo BAM
Contig lines e Reference	contém informações como os nomes dos fragmentos, tamanho e genoma de referência que é utilizado como *input* para gerar o arquivo BAM

Além do cabeçalho, um arquivo VCF contém as linhas com os dados das amostras, onde estão contidas em cada linha informações sobre uma única variante. As linhas com os dados são listadas seguindo o designado no cabeçalho do arquivo VCF. As oito primeiras colunas do arquivo (até INFO) representam as propriedades observadas dos sítios variantes (ou invariantes). Informações específicas à amostra são apresentadas na coluna FORMAT e nas seguintes.

As nove primeiras colunas são necessárias para o formato VCF, ainda que possam estar vazias. As informações contidas em cada coluna estão apresentadas na Tabela 6.

Tabela 6: Descrição dos componentes do corpo de um arquivo VCF

Informações presentes no corpo do arquivo	**Descrição das abreviações apresentadas**
CHROM e POS	contém as coordenadas genômicas nas quais as variantes ocorrem
ID	campo opcional, cujo conteúdo se baseia nas informações contidas em um banco de dados de referência como o dbSNP
REF e ALT	identificam o alelo referência e alternativo(s) observado(s) na(s) amostra(s). No caso de inserções, o alelo ALT fornece a sequência inserida, bem como a base anterior à inserção. Para deleções, o alelo ALT é representado pela base anterior à deleção
QUAL	é a probabilidade de que um polimorfismo (REF/ALT) exista de fato neste local indicado. No entanto, por ser este valor dependente da quantidade de dados analisada ele não representa uma medida útil para avaliação da qualidade de determinação de uma variante
FILTER	esse campo contém os filtros aplicados aos dados analisados. Se a variante passou pelos filtros aplicados o valor indicado é PASS. Caso nenhum filtro tenha sido aplicado será indicado "."
INFO	informações descrevendo cada variante
DB	associado ao dbSNP
H3	associado ao HapMap3
VALIDATED	validado por experimentos contínuos
NA	número total de alelos nos genótipos identificados
AC	contagem alélica dentre os genótipos do alelo ALT, na mesma ordem em que listados
SVTYPE	tipo de variante estrutural (DEL para deleção, DUP para duplicação, INV para inversão)
END	posição final ocupada pela variante
IMPRECISE	indica que a posição ocupada pela variante não é precisa
CIPOS/CIEND	intervalo de confiança em torno da posição e posição final para variantes localizadas com baixa precisão
FORMAT	fornece a ordem das informações contidas nas colunas seguintes (amostras):
GT	indica o genótipo da amostra em cada sítio. Em diploides, são indicados os dois alelos de cada indivíduo, sendo "0" indicativo do alelo

	referência e "1" do alelo alternativo. O separador (/) indica que os genótipos não estão faseados e (|) indica genótipos em fase
AD	representa o número de *reads* que dá suporte a cada um dos alelos (alternativo e/ou referência) reportados
DP	indica o número total de *reads* que representam o genótipo reportado
GQ	indica qual a confiança (escala Phred) da determinação de genótipo realizada. Valores muito baixos são indicativos de pouca confiança no genótipo identificado
PL	é a probabilidade normalizada dos genótipos possíveis ou a probabilidade de o genótipo não estar correto. Assim, quanto menor o valor apresentado aqui, melhor
HP	identificadores de haplótipos eventualmente determinados pelo *ReadBackedPhasing*
PQ	qualidade de faseamento de haplótipos eventualmente determinados pelo *ReadBackedPhasing*

1.3.2. Remoção de genótipos de baixa qualidade

Alguns dos genótipos considerados pelo *HaplotypeCaller* e *UnifiedGenotyper* podem não ser verdadeiros, principalmente quando a genotipagem se refere a um segmento com baixa cobertura de sequenciamento ou quando, no caso de genótipos heterozigotos, um dos alelos é pouco representado.

Visando diminuir a taxa de falso-positivos, genótipos de baixa cobertura são interrogados, tomando como *input* o arquivo VCF obtido pelo *HaplotypeCaller* e *UnifiedGenotyper*.

Neste processo, o arquivo VCF é tratado pela ferramenta VCFx (versão 1.0) (Castelli et al., 2015), disponível em http://www.castelli-lab.net/apps/apps_vcfx.php, que interroga alelos segundo as regras:

- Para genótipos homozigotos inferidos em segmentos cobertos por oito ou menos *reads,* é introduzido um alelo interrogado (vcfx *alpha* = 8). A definição final desse tipo de genótipo (homozigoto ou heterozigoto) será inferida posteriormente.
- Genótipos heterozigotos em que um dos alelos está altamente subrepresentado (proporção de *reads* igual ou menor que 1%), são considerados homozigotos para o alelo mais representado (vcfx *beta* = 0.01). Esse procedimento minimiza a influência de *reads* mapeadas incorretamente e da alta taxa de erros de sequenciamento geralmente introduzidos em dados de NGS. Essa correção só pode ser aplicada em situações de alta cobertura (100 ou mais *reads*).

- Para genótipos heterozigotos em que um dos alelos está relativamente subrepresentado (proporção de *reads* entre 1% e 20%), um alelo interrogado é introduzido representando o alelo subrepresentado (vcfx *delta* = 0.20). A definição final desse tipo de genótipo (homozigoto ou heterozigoto) será inferida posteriormente.
- Para genótipos heterozigotos inferidos com no máximo cinco *reads* em que um dos alelos está subrepresentado (Proporção de *reads* ente 20% e 40%), um alelo interrogado é inserido representando o alelo subrepresentado. No entanto, quando os dois alelos têm a mesma representatividade, ambos são considerados interrogados (vcfx *gamma* = 0.50). A definição final desse tipo de genótipo (homozigoto ou heterozigoto) será inferida posteriormente.

A linha de comando utilizada no VCFx é indicada a seguir:

```
vcfx checkpl input=input.vcf output=output.vcf
```

Após a análise pelo VCFx, se forem realizadas análises por ambas as ferramentas (*HaplotypeCaller* e *UnifiedGenotyper*) as variantes que apresentaram genótipos distintos nos dois arquivos VCF obtidos através do *HaplotypeCaller* e *UnifiedGenotyper*, são analisadas manualmente para (a) verificar qual dos dois algoritmos fez a identificação correta do genótipo e (b) nos casos de genótipos interrogados por somente um dos algoritmos, para resolver variantes que são interrogadas por conta do número de *reads* consideradas pelos algoritmos, principalmente no caso do *HaplotypeCaller* que, devido aos filtros extremamente rigorosos, acaba descartando um número considerável de *reads* de suas análises. Para isso, é preciso analisar os arquivos BAM das amostras nos pontos interrogados. Como descrito anteriormente, o programa IGV pode ser utilizado para a visualização, avaliação de qualidade e contagem das *reads* alternativas que cobrem determinada posição. Considerando somente as bases chamadas com qualidade mínima Q30 em *reads* mapeadas com qualidade mínima Q30, as regras empregadas pelo VCFx são novamente aplicadas, agora manualmente para determinar se essas variantes continuariam interrogadas ou se seriam resolvidas.

Após essa etapa de maturação do arquivo, as informações contidas nos dois arquivos são unidas em um único arquivo VCF (VCF consenso) que é utilizado nas análises posteriores.

Após a análise pelo VCFx, se apenas a ferramenta *HaplotypeCaller* com o parâmetro --emitRefConfidence GVCF foi utilizada, os pontos de variação interrogados das amostras devem ser visualizados no programa IGV. É preciso analisar os arquivos BAMs destas amostras para avaliar a qualidade e contagem de *reads* alternativas que cobrem determinada

posição (interrogada pelo programa VCFx). Isto porque o *HaplotypeCaller* aplica filtros extremamente rigorosos, descartando um número considerável de *reads* para suas análises. Portanto, analisar ponto a ponto os arquivos BAM muitas vezes pode auxiliar na compreensão do porquê as variantes foram interrogadas pelo programa VCFx.

Após a maturação manual o arquivo VCF é atualizado manualmente nas posições onde houveram alelos interrogados reavaliados e resolvidos.

2. REFERÊNCIAS

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Castelli, E. C., Mendes-Junior, C. T., Sabbagh, A., Porto, I. O. P., Garcia, A., Ramalho, J., … Donadi, E. A. (2015). HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. *Human Immunology*, *76*(12), 945–953. https://doi.org/10.1016/j.humimm.2015.06.016

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–8. https://doi.org/10.1038/ng.806

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research*, *8*(3), 175–185. https://doi.org/10.1101/gr.8.3.175

Ferragina, P., Manzini, G., Veli, M., & Navarro, G. (2004). An Alphabet-Friendly FM-Index, 150–160.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. https://doi.org/10.1093/bib/bbs017

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In *Current Protocols in Bioinformatics* (Vol. 11, p. 11.10.1-11.10.33). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471250953.bi1110s43

Anexo E – **Imputação**

Laboratório de Pesquisas Forenses e Genômicas
Departamento de Química – FFCLRP/USP

Protocolo de Imputação de Dados Genômicos

Marcorin L, Carrato TMT, Moraes VMS, Mendes-Junior CT

Versão 1 – Julho de 2022

Preparação do *input*

Para imputação no Sanger Imputation Server, o servidor pede um arquivo de input no formato VCF, no qual todos os alelos estejam na fita forward, e também estejam anotados no painel de referência para a imputação (no nosso caso, 1000 Genomes Phase 3), alinhados com o genoma de referência hg19. O Michigan Imputation Server pede apenas um arquivo no formato VCF e faz as conferências por si só (exclusão de SNPs que não estão no painel de referência e troca de fitas dos alelos). Porém, para fins de padronização, utilizaremos o mesmo input VCF para os dois programas, e, portanto, precisamos prepará-lo para atender aos requisitos do Sanger. Para o IMPUTE2 é necessário a conversão do VCF para GEN, a partir do software 'QCTOOL v2', para sistemas UNIX (disponível em https://www.well.ox.ac.uk/~gav/qctool_v2/documentation/download.html

A ferramenta HRC-1000G-check-bim-pl v3.0, (disponível em https://www.well.ox.ac.uk/~wrayner/tools), é utilizada para correção nos arquivos VCF preparados anteriormente. Um arquivo no formato LEGEND, contendo as frequências alélicas da população de referência, é necessário para sua utilização, esses dados baseados no painel do 1000 *Genomes Project* estão disponíveis no mesmo site. Utilizamos apenas os dados da população de Porto Rico, devido sua semelhança à população brasileira. Avaliamos a semelhança da amostra brasileira estudada aqui com as populações continentais Europeia, Africana, Leste Asiática e Americana do painel *1000 Genomes Project* e a população americana do *Human Genome Diversity Project* utilizando análise de componente principal (PCA). A **Figura E1** mostra que a população americana do *1000 Genomes Project* é mais

semelhante à nossa amostra. Porém, dentro desse grupo há populações muito diversas, então avaliamos também a semelhança da amostra brasileira com as subpopulações americanas desse painel. Como mostra a **Figura E2**, a população Porto-Riquenha é a mais semelhante, seguida pela população Colombiana. A configuração padrão do programa filtra as posições que não estão presentes na referência, posições com frequências que apresentem diferenças maiores do que 20% com relação a referência e posições palindrômicas (alelos A/T e C/G) com *minor alelle frequency* acima de 40%.

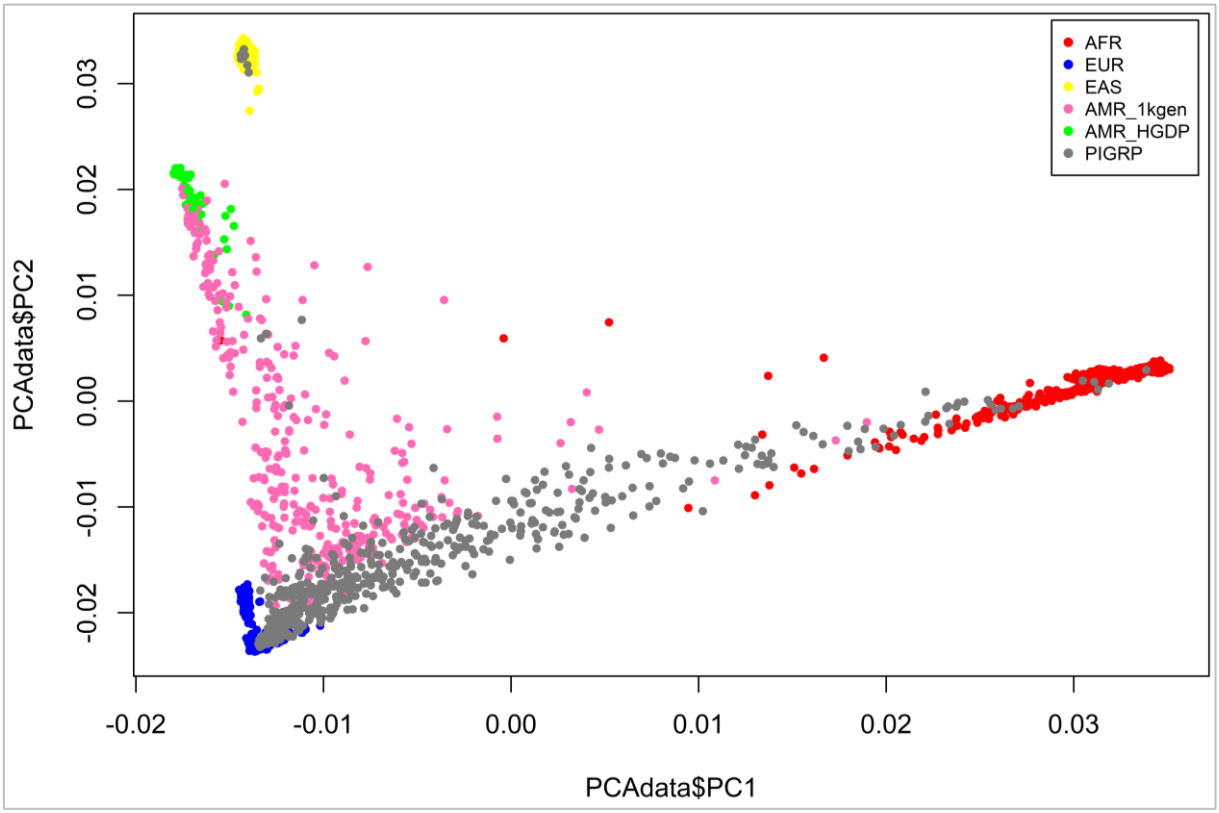


Figura E1. Análise de componente principal (PCA) para amostra brasileira (PIGID) e as populações continentais europeia (EUR), africana (AFR), leste asiática (EAS) e americana (AMR_1kgen) do *1000 Genomes Project* e americana (AMR_HGDP) do *Human Genome Diversity Project*.

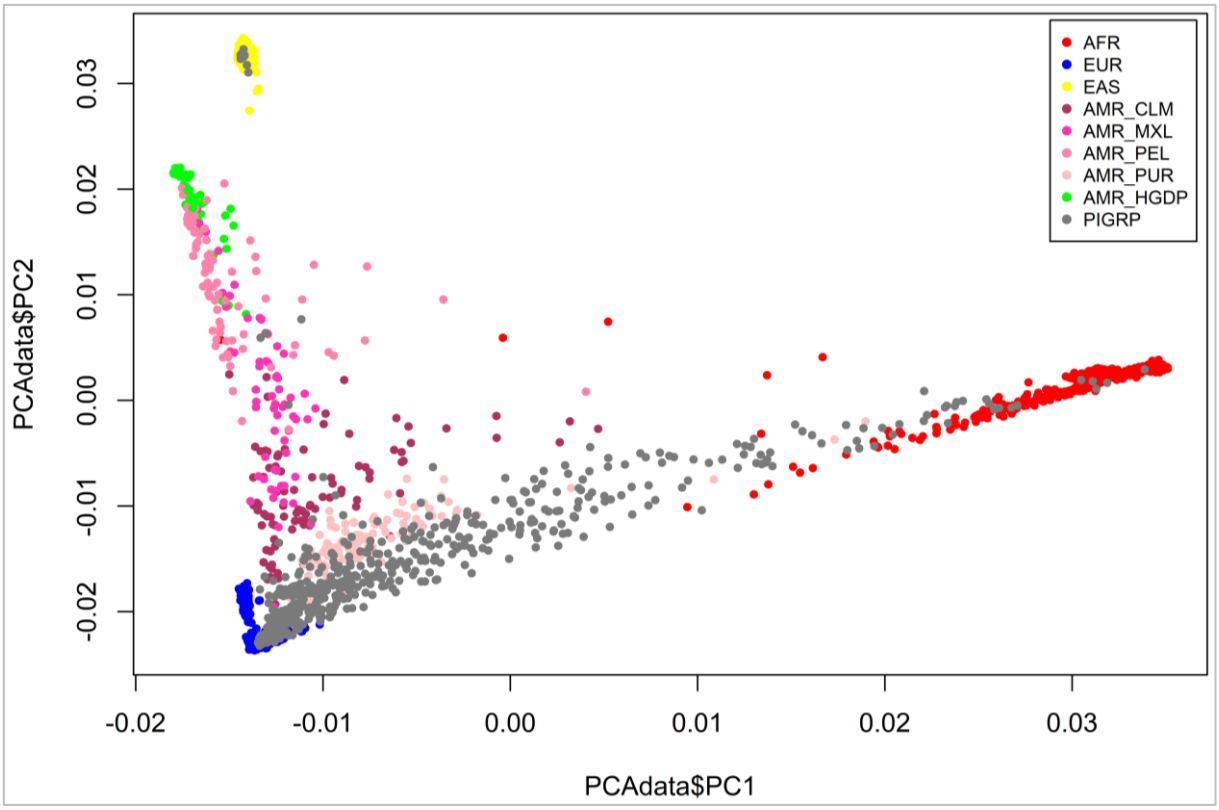


Figura E2. Análise de componente principal (PCA) para amostra brasileira (PIGID) e as populações europeia (EUR), africana (AFR), leste asiática (EAS), subpopulações colombiana (AMR_CLM), mexicana (AMR_MXL), peruana (AMR_PEL) e porto-riquenha (AMR_PUR) do *1000 Genomes Project* e população americana do *Human Genome Diversity Project.*

Sanger Imputation Server

O *Sanger Imputation Server* é uma plataforma online para faseamento e imputação de grandes volumes de dados (disponível em https://imputation.sanger.ac.uk/). O serviço é prestado pelo *Wellcome Sanger Institute* (https://www.sanger.ac.uk/) e é gratuito.

Após ter finalizado as etapas de correção do arquivo VCF, é preciso registrar um trabalho na plataforma do *Sanger Imputation Server*. Nessa etapa é necessário inserir uma conta Globus (um serviço de transferência de grandes volumes de dados da universidade de Chicago, disponível em https://www.globus.org/), escolher o painel de referência e decidir se os dados devem ser somente faseados, somente imputados (nesse caso, o arquivo deve ter sido previamente faseado), ou faseados e imputados. O servidor oferece duas opções de ferramentas para faseamento, EAGLE (recomendado) e SHAPEIT2, e quatro opções de painéis de referência: *Haplotype Reference Consortium (release 1.1)*, *African Genome Resources*, *1000 Genomes Phase 3*, UK10K, e a junção de UK10K e *1000 Genomes Phase 3*. Não é possível escolher subpopulações dentro dois painéis. Para esse estudo escolhemos fasear com EAGLE e imputar utilizando o painel do *1000 Genomes Project* como referência.

Com o projeto registrado, é enviado um e-mail com o endereço para transferência do arquivo, que deve conter todas as regiões de interesse compactadas (.vcf.gz), para a plataforma Globus. Assim, foi enviado um arquivo VCF contendo todas as posições dos 22 cromossomos autossomos. Após terminado o projeto, o servidor envia um novo e-mail com o endereço para transferência dos resultados, que consistem em uma pasta com arquivos VCF para cada cromossomo faseado e imputado. Embora nenhum relatório de qualidade seja emitido pela plataforma, o *imputation score* pode ser obtido no campo INFO do arquivo VCF.

Michigan Imputation Server

O *Michigan Imputation Server* é um servidor online (disponível em https://imputationserver.sph.umich.edu/index.html) e gratuito para imputação de genótipos utilizando o software Minimac4. Para usá-lo, é necessário ter um cadastro no site.

Como input, é necessário um arquivo VCF compactado (vcf.gz) contendo informações para apenas um cromossomo. Assim, foram preparados 22 arquivos VCFs de input referentes aos 22 cromossomos autossomos.

Para começar um trabalho de imputação, na aba "*Run*" (após escolher "*Genotype Imputation* (Minimac 4)"), primeiramente deve-se nomear o projeto (etapa opcional) e escolher qual painel de referência usar para a imputação, sendo a fase 3 do projeto 1000 *Genomes* (1000G Phase 3 v5) o painel escolhido. Posteriormente, devem ser fornecidos ao sistema os 22 arquivos VCF compactados (vcf.gz). Em cada um deles, os sítios de variação devem estar organizados por posição genômica. Em "*Array Build*" deve-se selecionar o genoma de referência com o qual nossos dados estão alinhados (em nosso caso, GRCh37/hg19). Em "*rsq Filter*", deve-se indicar se deseja que o programa faça algum filtro de qualidade. Visto que esse filtro pode ser aplicado manualmente posteriormente, foi marcada a opção padrão "*off*". Em "*Phasing*", o algoritmo Eagle v2.4 é escolhido para faseamento do arquivo submetido. Além de ser a indicação do servidor, é também o algoritmo utilizado pelo Sanger. No campo "*Population*", também deve-se escolher dentre os grupos populacionais europeu (EUR), africano (AFR), americano miscigenado (AMR), leste asiático (EAS) ou sul asiático (SAS) do projeto 1000 Genomes apenas para que seja realizado um controle de qualidade das frequências, ou seja, para que as frequências do VCF submetido sejam comparadas com as frequências dessa população de referência. Como nossos arquivos já foram conferidos em sua preparação, podemos marcar o campo "*Other/Mixed*", e o controle

de qualidade das frequências nesse caso não é realizado. Em "*Mode*", escolhe-se o que deseja que o programa faça (em nosso caso, foi escolhida a opção de Default "*Quality Control & Imputation*").

Confirmando os termos de serviço, pode-se dar início à imputação. O servidor verifica se o arquivo de input está valido, gerando informações como o número de amostras, o cromossomo imputado e o número de SNPs presentes. Também é realizado o controle de qualidade em relação ao genoma de referência, no qual as variantes são checadas e excluídas caso (a) estejam em duplicatas, (b) sejam INDEL, (c) sejam sítios monomórficos, ou (d) tenham um SNP call rate abaixo de 90%. Como os arquivos haviam sido previamente verificados, nenhuma variante precisou ser excluída nessa etapa. Após isso, todos os sítios de variação contidos no intervalo de SNPs do VCF fornecido serão imputados.

Como resultado, o servidor disponibiliza para download um arquivo ZIP contendo o VCF compactado com os genótipos imputados e um arquivo INFO com as informações sobre a imputação em cada ponto. Esses arquivos ficam disponíveis para download no site do Michigan por uma semana. É enviado por e-mail uma chave de acesso para extração dos arquivos.

A **Tabela E1** nos traz as quantidades de posições imputadas por cada uma das metodologias utilizadas. Os valores de INFO *score*, e de Rsq indicam a confiabilidade de cada posição imputada para cada servidor. O servidor *Sanger* imputou um total de 81.706.022 posições com 30,98% dessas posições dentro do INFO *score* acima de 0,99, enquanto que 33,49% dessas posições encontram-se abaixo do *score* de 0,30. O servidor *Michigan* imputou um total de 47.114.758 posições com 5,82% dessas posições apresentando o Rsq acima de 0,99, enquanto que 42,6% dessas posições encontram-se abaixo do valor de 0,30. O software Impute2 imputou um total de 76.211.542 posições com 6,42% dessas posições dentro do INFO *score* acima de 0,99, enquanto que 60,84% dessas posições encontram-se abaixo do *score* de 0,30.

Tabela E1. Total de posições imputadas pelas ferramentas *Sanger Imputation Server*, *Michigan Imputation Server* e Impute2 e as proporções de posições imputadas com confiabilidade (valores de INFO *score*) de 30%, 50%, 80% e 99% em cada cromossomo autossômico.

	Sanger Imputation Server					**Michigan**					**Impute2**				
Crom.	**Nº posições**	**% INFO ≥ 0,99**	**% INFO ≥ 0,80**	**% INFO ≥ 0,50**	**% INFO ≥ 0,30**	**Nº posições**	**% Rsq ≥ 0,99**	**% Rsq ≥ 0,80**	**% Rsq ≥ 0,50**	**% Rsq ≥ 0,30**	**Nº posições**	**% INFO ≥ 99**	**% INFO ≥ 80**	**% INFO ≥ 50**	**% INFO ≥ 30**
1	6.500.358	31,08	49,40	58,95	66,26	3.738.816	5,50	38,17	51,37	56,48	6.039.976	6,61	25,97	34,25	38,96
2	7.117.614	31,58	50,17	59,37	66,29	4.058.008	6,39	39,96	52,58	57,34	7.117.614	6,44	25,95	33,92	38,46
3	5.862.629	30,82	49,73	58,89	65,66	3.356.270	6,79	41,23	53,70	58,37	5.862.629	6,88	26,57	34,49	38,96
4	5.763.673	30,58	50,22	59,52	66,37	3.338.491	6,28	41,73	54,20	58,87	1.116.134	6,61	28,16	36,61	41,08
5	5.293.915	30,72	50,03	59,29	66,13	3.032.697	6,24	41,24	53,79	58,49	5.293.915	6,64	26,69	34,59	39,09
6	5.051.641	31,07	50,59	59,75	66,44	2.954.714	8,00	42,69	54,83	59,37	5.051.642	8,17	28,20	35,99	40,45
7	4.741.583	30,88	50,30	59,90	66,88	2.753.762	5,76	39,72	52,89	57,85	4.741.586	6,56	26,67	34,99	39,58
8	4.622.575	31,62	50,68	59,87	66,85	2.651.746	6,00	40,40	52,91	57,65	4.622.576	6,35	26,28	34,28	38,84
9	3.579.889	31,46	50,26	59,93	67,46	2.063.313	5,20	38,37	51,59	56,47	3.579.890	5,77	25,54	33,94	38,55
10	4.013.458	31,54	50,68	60,19	67,23	2.334.319	5,96	40,06	52,86	57,77	4.013.459	6,64	26,81	35,09	39,64
11	4.067.158	30,99	49,76	58,95	65,73	2.333.614	6,34	39,94	52,60	57,48	4.067.160	6,92	26,42	34,40	38,90
12	3.889.061	31,53	50,39	59,88	66,85	2.243.031	5,91	39,76	52,92	57,88	3.591.949	6,99	26,45	34,59	39,15
13	2.872.968	31,77	51,18	60,51	67,44	1.661.804	6,46	41,50	54,43	59,21	2.872.968	7,05	27,30	35,45	39,98
14	2.669.300	30,90	49,63	59,24	66,53	1.535.757	5,39	39,04	51,92	56,74	2.669.300	6,44	26,18	34,35	39,01
15	2.437.477	30,76	49,42	59,24	66,68	1.404.321	4,34	36,82	50,63	55,92	2.348.020	5,37	25,16	33,96	38,76
16	2.713.871	30,68	48,79	59,05	66,50	1.549.545	4,00	34,68	48,74	54,16	2.713.871	4,54	23,88	33,05	37,93
17	2.341.782	30,08	47,91	58,30	66,02	1.346.121	3,80	34,24	49,31	55,04	2.341.782	5,07	24,43	33,95	38,85
18	2.279.214	30,95	50,07	59,97	67,26	1.319.711	5,17	38,94	52,78	58,03	2.279.215	6,05	26,23	35,04	39,78
19	1.843.199	29,25	46,91	58,34	66,37	1.084.846	3,42	31,58	47,65	54,10	1.843.199	4,98	24,04	34,82	39,94
20	1.822.225	30,24	48,79	58,93	66,57	1.047.713	4,54	37,12	51,65	57,05	1.822.225	5,42	25,12	34,23	39,06
21	1.112.215	29,57	48,33	59,30	66,93	653.843	3,95	35,68	50,10	55,13	1.112.215	5,23	25,09	34,42	39,37
22	1.110.217	29,27	46,69	57,74	66,19	652.316	3,74	32,64	47,39	53,13	1.110.217	4,94	23,70	33,85	39,47
Total	**81.706.022**	**30,98**	**49,86**	**59,41**	**66,51**	**47.114.758**	**5,82**	**39,37**	**52,45**	**57,40**	**76.211.542**	**6,42**	**26,18**	**34,53**	**39,16**

Impute2

Impute2 é um programa de utilização livre para fins acadêmicos, disponível para download em https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#download. O arquivo VCF utilizado no programa foi o mesmo aplicado aos outros dois servidores. Seguindo o manual de "*Best Practices"* e utilizando os arquivos de referência "1000 *Genomes Phase 3"* disponibilizados no mesmo site, seguiu-se todo o protocolo padrão, incluindo o *pre-phasing,* sendo este indicado mesmo em dados já corrigidos por outros programas. As seguintes linhas de comandos foram utilizadas

Pre-Phasing:

```
./impute2 \
     -prephase_g \
     -m ./Example/example.chr22.map \
     -g ./Example/example.chr22.study.gens \
     -int 20.4e6 20.5e6 \
     -Ne 20000 \
     -o ./Example/example.chr22.prephasing.impute2
```

Imputação:

```
./impute2 \
     -use_prephased_g \
     -m ./Example/example.chr22.map \
     -h ./Example/example.chr22.1kG.haps \
     -l ./Example/example.chr22.1kG.legend \
     -known_haps_g
./Example/example.chr22.prephasing.impute2_haps \
     -strand_g ./Example/example.chr22.study.strand \
     -int 20.4e6 20.5e6 \
     -Ne 20000 \
     -o ./Example/example.chr22.one.phased.impute2
     -phase
```

Os comandos significam então:

- `-m`: o mapa de recombinação fornecido pelo site, junto ao download dos arquivos do 1000 *Genomes;*

- `-g:` o arquivo contendo os dados dos quais se quer realizar a imputação;
- `-int:` o intervalo genômico a se realizar a imputação, de acordo com o manual é indicado que se realize em partes (*chunks)*, de $5x10^6$ bases, sendo então cada cromossomo dividido em vários pedaços de 5 Mb.
- `-Ne:` tamanho efetivo populacional. O valor sugerido pelos desenvolvedores para estudos modernos de imputação é de 20000.
- `-h:` arquivo contendo todos os haplótipos conhecidos. Fornecidos pelos desenvolvedores do programa.
- `-l:` arquivo LEGEND, com as informações dos SNPs presentes no arquivo fornecido em `-h`. Fornecido pelos desenvolvedores.
- `-o:` nome do *output*, definido pelo usuário.
- `-known_haps_g:` arquivo resultado da função `-prephase_g`.
- `-strand_g:` arquivo demonstrando a orientação da fita onde encontram-se os SNPs fornecidos em `-g`.