



**Universidade do Estado do Rio de Janeiro**  
Centro Biomédico  
Instituto de Biologia Roberto Alcântara Gomes

Fernanda Mathias de Miranda Barros

**Composição genética de populações do Sudeste Asiático:  
análise dos principais haplogrupos da região e contribuição para o  
entendimento das teorias das rotas de migração**

Rio de Janeiro  
2019

Fernanda Mathias de Miranda Barros

**Composição genética de populações do Sudeste Asiático:  
análise dos principais haplogrupos da região e contribuição para o  
entendimento das teorias das rotas de migração**

Tese apresentada como requisito parcial para a obtenção do título de Doutor, ao Programa de Pós-Graduação em Biociências na Universidade do Estado do Rio de Janeiro.

Orientadora: Prof.<sup>a</sup> Dra. Maria Leonor Rodrigues de Sousa Botelho de Gusmão

Rio de Janeiro

2019

CATALOGAÇÃO NA FONTE  
UERJ/REDE SIRIUS/BIBLIOTECA CB-A

B277 Barros, Fernanda Mathias de Miranda

Composição genética de populações do Sudeste Asiático: análise dos principais haplogrupos da região e contribuição para o entendimento das teorias das rotas de migração / Fernanda Mathias de Miranda Barros. – 2019.

111 f.

Orientadora: Maria Leonor Rodrigues de Sousa Botelho de Gusmão.

Tese (Doutorado) - Universidade do Estado do Rio de Janeiro, Instituto de Biologia Roberto Alcântara Gomes. Pós-graduação em Biociências.

1. Genética de populações – Ásia, Sudeste - Teses. 2. Cromossomo Y. 3. População - Aspectos genéticos - Teses. 4. Variedade genética - Teses. I. Gusmão, Maria Leonor Rodrigues de Sousa Botelho de. II. Universidade do Estado do Rio de Janeiro. Instituto de Biologia Roberto Alcântara Gomes. III. Título.

CDU 575.17

Bibliotecária: Angela da Silva Velho - CRB7/4780

Autorizo apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta tese, desde que citada a fonte.

---

Assinatura

---

Data

Fernanda Mathias de Miranda Barros

**Composição genética de populações do Sudeste Asiático:  
análise dos principais haplogrupos da região e contribuição para o  
entendimento das teorias das rotas de migração**

Tese apresentada como requisito parcial para a  
obtenção do título de Doutor, ao Programa de  
Pós-Graduação em Biociências na Universidade  
do Estado do Rio de Janeiro.

Aprovada em 27 de novembro de 2019.

Banca Examinadora: \_\_\_\_\_

Prof.<sup>a</sup> Dra. Maria Leonor Rodrigues de Sousa Botelho de  
Gusmão (Orientadora)

Instituto de Biologia Roberto Alcântara Gomes - UERJ

\_\_\_\_\_  
Prof.<sup>a</sup> Dra. Cíntia Barros Santos-Rebouças

Instituto de Biologia Roberto Alcântara Gomes - UERJ

\_\_\_\_\_  
Prof.<sup>a</sup> Dra. Verônica Marques Zembrzuski

Fundação Oswaldo Cruz

\_\_\_\_\_  
Prof. Dr. Mário Campos Junior

Fundação Oswaldo Cruz

Rio de Janeiro

2019

## DEDICATÓRIA

Dedico essa tese a minha família: Caio, Nuno e Lia.

## AGRADECIMENTOS

Gostaria de agradecer a todos que, de alguma forma, contribuíram para que essa tese se concretizasse.

Em primeiro lugar gostaria de agradecer aos meus filhos Nuno e Lia que, no meio de tanta pressão, me fazem lembrar diariamente a simplicidade e alegria da vida.

Aos meus pais, Mário e Malvina, por sempre me ensinarem a importância e valorização dos estudos.

Aos meus irmãos, Márcia e Mário Sérgio, por serem meu porto seguro. É muito confortável saber que sempre que precisar posso “correr para casa”.

À minha sogra, por proporcionar momentos divertidos aos meus filhos na minha ausência para o trabalho.

À minha orientadora. É uma honra e orgulho ter trabalhado com você. Um exemplo de simplicidade, compreensão e competência. Preocupada não só com a realização do trabalho, mas principalmente com o ser humano. A nossa convivência foi um aprendizado diário para mim.

À toda equipe do LDD, principalmente a Silvia pelo apoio quando eu estava grávida.

Às amigas, Holy e Rossana, pelos momentos descontraídos e de tantas risadas.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da minha bolsa de estudos.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Meu agradecimento mais profundo só poderia ser dedicado a uma pessoa, meu marido Caio. Difícil escrever em tão poucas palavras o tanto que eu gostaria de agradecer. Obrigada pela calma e compreensão nos momentos de tensão, que não foram poucos. Você fez parte de cada pedaço dessa minha conquista.

O senhor saiba: eu toda a minha vida pensei por mim, forro, sou nascido diferente. Eu sou é eu mesmo. Diverjo de todo o mundo... Eu quase que nada não sei. Mas desconfio de muita coisa. O senhor concedendo, eu digo: para pensar longe, sou cão mestre - o senhor solte em minha frente uma idéia ligeira, e eu rastreio essa por fundo de todos os matos. Amén!

*João Guimarães Rosa*

## RESUMO

BARROS, Fernanda Mathias de Miranda. *Composição genética de populações do Sudeste Asiático: análise dos principais haplogrupos da região e contribuição para o entendimento das teorias das rotas de migração*. 2019. 111 f. Tese (Doutorado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

A caracterização genética de uma população é de suma importância, pois permite conhecer a história/evolução de um determinado país/continente, além da demografia dos diferentes subgrupos étnicos presentes numa dada região. Pesquisas realizadas em Genética Populacional têm focado, principalmente, em regiões europeias, sendo a Ásia sub-representada, apesar de ser o maior e mais populoso continente. O presente trabalho teve por objetivo aprofundar o conhecimento da composição genética de dois países do Sudeste Asiático, Vietnã e Filipinas, contribuindo para o estudo filogenético da região e, conseqüentemente, para o entendimento das teorias das rotas de migração no Sudeste Asiático. Para isso, foi feita, primeiramente, uma análise da estrutura genética de ambas as populações quanto às suas linhagens paternas. Logo, foram selecionadas (de forma aleatória) 318 amostras vietnamitas e 120 amostras de filipinos, de indivíduos do sexo masculino. Para essas amostras, foi feita a extração do DNA e a genotipagem de marcadores específicos do cromossomo Y (Y-SNP e Y-STR, os quais diferem quanto a taxa de mutação). Assim, foi possível avaliar a diferenciação populacional em diferentes escalas temporais. Em relação às populações vietnamitas, foi visto que há uma diferença maior entre populações de uma mesma região do que entre populações localizadas em diferentes regiões, indicando que não existe uma boa correlação entre Geografia e Genética. No entanto, quando utilizados critérios de etnicidade, verificou-se uma diferenciação genética dos diferentes grupos étnicos, demonstrando ser este o melhor critério para agrupamento das subpopulações do Vietnã. Dos nove grupos étnicos do Vietnã incluídos neste estudo, o Mong mostrou ser o mais diferenciado, apresentando diferenças genéticas significativas com as etnias restantes. Nas Filipinas, observou-se uma pequena distância genética entre os três principais grupos de ilhas (Luzon, Mindanao e Visayas), sendo esta não significativa. Os resultados obtidos neste estudo demonstraram que existe uma grande diversidade genética entre indivíduos de uma mesma região e pouca diferença entre populações de diferentes regiões geográficas. Os dados obtidos para as populações Vietnamitas e Filipinas foram ainda comparados com os de outras populações do Sudeste Asiático, para as quais existiam dados disponíveis para os mesmos marcadores genéticos. Os resultados obtidos condizem com a existência de duas rotas de migração no continente do Sudeste Asiático, tendo as mesmas se bifurcado próximo ao Paquistão e seguido uma para Norte e outra para Sul. Nas ilhas do Sudeste Asiático, os resultados sugerem também a existência de duas rotas: uma no sentido de Taiwan a Indonésia e a outra no sentido contrário.

Palavras-chave: Genética de Populações. Sudeste Asiático. Vietnã. Filipinas. Cromossomo Y. Linhagens paternas.



## ABSTRACT

BARROS, Fernanda Mathias de Miranda. *Genetic composition of Southeast Asian species: analysis of the main groups of regions and contribution to the understanding of migration route theories*. 2019. 111 f. Tese (Doutorado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

The genetic characterization of a population is extremely important, as it allows to know the history/evolution of a given country/continent, as well as the demographics of different ethnic subgroups present in certain regions. Population Genetics research has focused mainly on European regions, with Asia being underrepresented, despite being the largest and most populous continent. The present work aimed to deepen the knowledge of the genetic composition of two Southeast Asian countries, Vietnam and the Philippines, contributing to the phylogenetic study of the region and, consequently, to the understanding of the migration route theories in Southeast Asia. To this end, a first analysis of the genetic structure of both populations was performed regarding their paternal lineages. Therefore, 318 Vietnamese and 120 Filipino samples were randomly selected from male subjects. For these samples, DNA extraction and genotyping with specific Y chromosome markers (Y-SNP and Y-STR, which differ in mutation rate) were performed. Thus, it was possible to evaluate population differentiation at different time scales. Concerning Vietnamese populations, larger differences were observed between populations of the same region than between populations from different regions, indicating a weak correlation between Geography and Genetics. However, when ethnicity criteria were used, there was a genetic differentiation of the ethnic groups, which proved to be the best criterion for grouping subpopulations of Vietnam. Of the nine ethnic groups in Vietnam included in this study, Mong proved to be the most differentiated, presenting significant genetic differences with the remaining ethnic groups. In the Philippines, a small genetic distance was observed between the three main island groups (Luzon, Mindanao and Visayas), which was not significant. The results obtained in this study demonstrated that there is a great genetic diversity among individuals from the same region and a small differentiation between populations from different geographic regions. The data obtained for the Vietnamese and Philippine populations were further compared with those of other Southeast Asian populations, for which data from the same genetic markers were available. The results obtained are consistent with the existence of two migration routes on the continent of Southeast Asia, which branched close to Pakistan and followed one to the north and one to the south. On the islands of Southeast Asia, the results also suggest the existence of two routes: one going from Taiwan to Indonesia and another going in the opposite direction.

Keywords: Population Genetics. Southeast Asia. Vietnam. Philippines, Y chromosome, Male lineages

## LISTA DE FIGURAS

Figura 1 – Mapa representativo da hipótese das duas rotas de migração da África para a Eurásia. ....	17
Figura 2 – Representação da hipótese da existência de apenas uma rota de migração.....	18
Figura 3 – Representação da hipótese da existência de duas rotas de migração. .	20
Figura 4 – Representação das hipóteses das rotas de migração das Ilhas do Sudeste Asiático.....	21
Figura 5 – Localização geográfica das Filipinas.....	22
Figura 6 – Mapa da população da linguagem Austronésia. ....	23
Figura 7 – As três etapas do processo de formação do povo vietnamita. ....	25
Figura 8 – Localização geográfica do Vietnã e suas respectivas etnias .....	27
Figura 9 – Comparação de diferentes modos de transmissão genética, incluindo a do cromossomo Y. ....	33
Figura 10 – Marcadores Y-STR mais utilizados. ....	35
Figura 11 – Árvore filogenética do cromossomo Y.....	37
Figura 12 – Distribuição geográfica dos Haplogrupos C, D, N e O. ....	39
Figura 13 – Mapa do Vietnã. ....	43
Figura 14 – Mapa das Filipinas. ....	44
Figura 15 – Árvore filogenética dos haplogrupos do cromossomo Y, definidos com base em Y-SNP. ....	47
Figura 16 – Estratégia de tipagem no <i>Multiplex</i> Um usando gel de Poliacrilamida. ....	49
Figura 17 – Descrição da estratégia utilizada na genotipagem dos quinze Y-STRs.....	50
Figura 18 – Multiplex Ásia Dois.....	51
Figura 19 – Representação gráfica dos haplogrupos encontrados nas amostras do Vietnã. ....	56
Figura 20 – Representação gráfica dos haplogrupos encontrados nas amostras das Filipinas. ....	63
Figura 21 – Mapa e Rotas de Migração do continente do Sudeste Asiático .....	69
Figura 22 – Escalonamento Multidimensional das análises de Y-STR ( $F_{ST}$ ) para as amostras do continente do Sudeste Asiático. ....	70

Figura 23 – Mapa e Rotas de Migração nas ilhas do Sudeste Asiático.....	71
Figura 24 – Escalonamento multidimensional das análises de Y-STR ( $F_{ST}$ ) para as amostras das ilhas do Sudeste Asiático.....	72

## LISTA DE QUADROS

Quadro 1 – Classificação vietnamita de acordo com a língua.....	29
Quadro 2 – Haplogrupos com os relativos marcadores e localização.....	36
Quadro 3 – <i>Multiplex</i> Ásia Um.....	48
Quadro 4 – Lista dos <i>primers</i> de PCR inclusos nos dois <i>Multiplex</i> . <i>Multiplex</i> Um representado em vermelho e o Dois em azul.....	52
Quadro 5 – Lista dos <i>primers</i> de extensão (SBE) selecionados. Representado em vermelho os <i>primers</i> do <i>Multiplex</i> Um e em azul os do <i>Multiplex</i> Dois.....	53
Quadro 6 – Classificação das etnias, segundo a literatura. ....	61

## LISTA DE TABELAS

Tabela 1 – Resultado do teste estatístico AMOVA para diferentes províncias vietnamitas.....	57
Tabela 2 – Resultado das análises de distância genética par a par para Y-SNPs e Y-STRs das províncias .....	58
Tabela 3 – Diversidade haplotípica. ....	60
Tabela 4 – Resultados das análises de distância genética das etnias do Vietnã. ....	62
Tabela 5 – Resultado do teste estatístico AMOVA entre os grupos das Filipinas .....	64
Tabela 6 – Resultado das análises de distância genética do grupo de ilhas das Filipinas.....	65
Tabela 7 – Comparação par a par das distâncias genéticas $F_{ST}$ entre amostras populacionais das Filipinas. ....	66
Tabela 8 – Comparação par a par das distâncias genéticas $R_{ST}$ entre amostras populacionais das Filipinas. ....	66
Tabela 9 – Comparação par a par de distâncias genéticas $F_{ST}$ entre amostras populacionais das Filipinas. ....	67
Tabela 10 – Comparação par a par das distâncias genéticas $R_{ST}$ entre amostras populacionais das Filipinas. ....	68

## LISTA DE ABREVIATURAS E SIGLAS

AMOVA	Análise de Variância Molecular
DNA	Ácido desoxirribonucleico (do inglês, <i>Deoxyribonucleic Acid</i> )
InDels	Polimorfismo de Inserção e Deleção (do inglês, <i>Insertion Deletion Polymorphisms</i> )
mtDNA	DNA Mitocondrial (do inglês, <i>Mitochondrial DNA</i> )
MDS	Escalonamento Multidimensional (do inglês, <i>MultiDimensional Scaling</i> )
MSY	Região específica masculina do cromossomo Y (do inglês, <i>Male Specific Y region</i> ).
NRY	Região Não-Recombinante do Cromossomo Y (do inglês, <i>Non Recombining region of the Y chromosome</i> ).
PCR	Reação em cadeia da polimerase (do inglês, <i>Polymerase Chain Reaction</i> )
SBE	<i>Primer</i> de Extensão (do inglês, <i>Single Base Extension</i> )
SNP	Polimorfismo de base única (do inglês, <i>Single-Nucleotide Polymorphism</i> )
STR	Repetição curta em <i>tandem</i> (do inglês, <i>Short Tandem Repeat</i> )
YCC	(do inglês, <i>Y Chromosome Consortium</i> )
YHRD	(do inglês, <i>Y-Chromosome STR Haplotype Reference Database</i> )

## SUMÁRIO

	<b>INTRODUÇÃO</b> .....	15
1	<b>REVISÃO DA LITERATURA</b> .....	16
1.1	<b>Migrações humanas para o Sudeste Asiático</b> .....	16
1.2	<b>Filipinas</b> .....	21
1.3	<b>Vietnã</b> .....	24
1.4	<b>Marcadores genéticos</b> .....	29
1.5	<b>Cromossomo Y</b> .....	31
2	<b>OBJETIVOS</b> .....	40
2.1	<b>Objetivo geral</b> .....	40
2.2	<b>Objetivos específicos</b> .....	40
3	<b>MATERIAIS E MÉTODOS</b> .....	42
3.1	<b>Amostragem e obtenção das amostras de DNA</b> .....	42
3.2	<b>Extração de DNA</b> .....	44
3.3	<b>Genotipagem de marcadores Y-STRs</b> .....	46
3.4	<b>Genotipagem de Y-SNPs</b> .....	47
3.4.1	<u>Protocolos utilizados com <i>Multiplex Um e Dois</i></u> .....	51
3.4.1.1	Seleção das sequências alvo e desenho dos <i>primers</i> .....	51
3.4.1.2	<i>Multiplex Um e Dois</i> .....	53
3.4.1.3	Análise estatística dos resultados .....	54
4	<b>RESULTADOS E DISCUSSÃO</b> .....	55
4.1	<b>Composição genética do Vietnã</b> .....	55
4.1.1	<u>Análise da subestrutura genética das províncias</u> .....	56
4.1.2	<u>Análise da diversidade das províncias</u> .....	59
4.1.3	<u>Análise da subestrutura genética por etnias</u> .....	60
4.2	<b>Composição genética das Filipinas</b> .....	62
4.2.1	<u>Análise da subestrutura genética</u> .....	64
4.2.2	<u>Comparação com outras amostras das Filipinas na literatura</u> .....	66
4.2.2.1	Comparação com 21 Y-STRs .....	66
4.2.2.2	Comparação com 17 Y-STRs .....	67

4.2.3	<u>Comparação entre as populações do Sudeste da Ásia</u> .....	68
4.2.3.1	Análise das rotas de migração continentais .....	68
4.2.3.2	Escalonamento multidimensional entre as amostras populacionais das ilhas do Sudeste Asiático.....	71
	<b>CONCLUSÕES</b> .....	73
	<b>REFERÊNCIAS</b> .....	75
	<b>APÊNDICE A - Tabela das amostras do Vietnã</b> .....	83
	<b>APÊNDICE B - Tabela das amostras das Filipinas</b> .....	91
	<b>APÊNDICE C - Referências das amostras do Sudeste Asiático</b> .....	95
	<b>APÊNDICE D - Análise das amostras do Continente Asiático</b> .....	97
	<b>APÊNDICE E - Análise das amostras das Ilhas Asiáticas</b> .....	98
	<b>ANEXO - Artigos publicados</b> .....	99



## INTRODUÇÃO

A Genética de Populações é uma ciência dedicada ao estudo da diversidade genética das populações e sua variação ao longo do tempo. Através do estudo da distribuição das frequências alélicas e genotípicas numa população, e sua alteração ao longo das gerações, é possível determinar as forças evolutivas responsáveis pela diversidade que é hoje encontrada dentro e entre populações, da mesma ou de distintas espécies (PYERITZ; KORF; GRODY, 2019). Um estudo mais aprofundado da história das populações humanas nos dá um melhor entendimento da diversidade genética existente e sua distribuição a nível das populações atuais. Da mesma maneira, a determinação dos padrões de variabilidade genética em populações humanas nos ajuda a compreender como as populações humanas evoluíram até aos dias de hoje (RAMACHANDRAN, 2010). O conhecimento da genética populacional é importante, portanto, por nos permitir inferir aspectos relacionados com a história e demografia das populações humanas e seus respectivos grupos étnicos.

Esta dissertação teve por objetivo a caracterização genética das linhagens paternas de populações localizadas no Sudeste Asiático, com a finalidade de investigar relações entre diferentes grupos étnicos e tentar contribuir para um melhor entendimento da sua história.

## 1 REVISÃO DA LITERATURA

A caracterização dos padrões de variação genética em populações africanas atuais em comparação com os encontrados em outros continentes, possibilitou uma reconstrução da história evolutiva das populações humanas (CAMPBELL; TISHKOFF, 2008; CAVALLI-SFORZA; FELDMAN, 2003; LOPEZ; VAN DORP; HELLENTHAL, 2015). Análises usando diversos tipos de marcadores autossômicos, por diferentes técnicas de genotipagem e sequenciamento, demonstraram que os níveis de diversidade mais elevados se encontram na África e que essa diversidade diminuiu à medida que nos distanciamos deste continente. A existência de uma maior variabilidade genética na África, em relação aos demais continentes, foi também observada usando-se marcadores de linhagem paterna (cromossomo Y) e materna (DNA mitocondrial). Assim, estudos filogenéticos relacionando os polimorfismos do DNA mitocondrial (mtDNA) e estudos focando na variação da região não recombinante do cromossomo Y (NRY) mostraram que as linhagens mais antigas se encontram na África, representando os ancestrais de todas as outras linhagens até agora encontradas (BELLWOOD; NESS, 2014). Além dos estudos genéticos, dados craniométricos também corroboraram a origem africana do homem moderno (ou *Homo sapiens*) (MATSUMURA *et al.*, 2019), portanto, existe atualmente um consenso sobre a descendência africana do homem moderno. Contudo, existem ainda algumas controvérsias em relação aos padrões de dispersão fora deste continente, os quais continuam a serem debatidos (REYES-CENTENO *et al.*, 2014).

### 1.1 Migrações humanas para o Sudeste Asiático

Ao mapear a frequência dos marcadores genéticos nos povos modernos conseguimos entender como os povos antigos se movimentaram pelo mundo. Essas grandes migrações levaram os descendentes de um pequeno grupo de africanos a ocupar diversos locais na Terra (CAVALLI-SFORZA; FELDMAN, 2003). Entretanto, continua sendo discutido como essas grandes migrações ocorreram. Nas últimas décadas surgiram várias hipóteses com modelos em debate divergindo quanto ao

número de dispersões e rotas geográficas, bem como ao tempo histórico, das mesmas (MATSUMURA *et al.*, 2019).

Segundo alguns autores (TIMMERMANN; FRIEDRICH, 2016), os primeiros *Homo sapiens* migraram alcançando a linha costeira nordeste da África e Bal-el-Mandeb, em baixa densidade, e em direção à Europa, Índia, sul da China e Sudeste da Ásia. Baseando-se em dados fósseis e arqueológicos, esses pesquisadores citam a hipótese de êxodos do *Homo sapiens* da África em direção à Eurásia em vários episódios orbitalmente estimulados, em que uma insolação de verão boreal somadas as chuvas intensas no norte de África e Península Arábica geraram corredores habitáveis do tipo savanas, permitindo assim a dispersão do homem moderno.

Uma questão recente é se teriam havido duas ou mais dispersões separadas do homem moderno da África para a Eurásia (LOPEZ; VAN DORP; HELLENTHAL, 2015). Uma hipótese seria a existência de duas rotas geográficas diferentes no nordeste da África. A primeira rota seria uma dispersão ao norte através do vale do Nilo e Península do Sinai para o sudoeste da Ásia e Europa, e a segunda ao sul passando pelo estreito de Bab-el-Mandeb em direção ao leste, costas do sul e Sudeste da Ásia para Austrália (MELLARS, 2006) (Figura 1).

Figura 1 – Mapa representativo da hipótese das duas rotas de migração da África para a Eurásia.



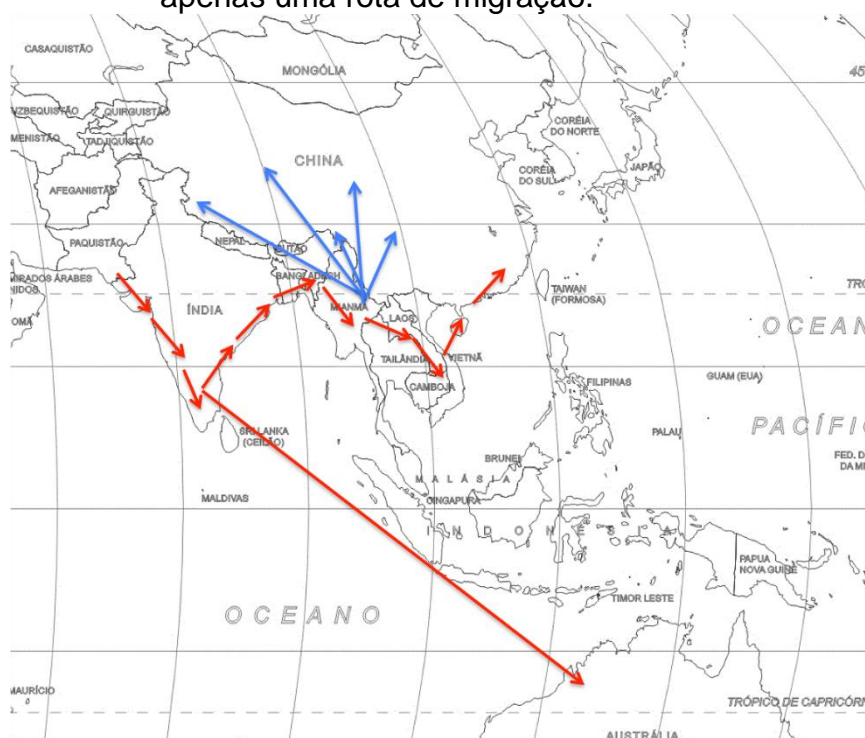
Fonte: (MELLARS, 2006).

Pesquisas arqueológicas, paleontológicas e genéticas têm auxiliado os questionamentos sobre as possíveis rotas de migrações humanas existentes, porém as dispersões humanas ocorridas no Sudeste Asiático e Oceania têm sido muito

discutidas, uma vez que esse tema continua em aberto e controverso (REICH *et al.*, 2011).

Evidências genéticas e arqueológicas vêm sendo utilizadas para fortalecer a hipótese de uma única onda migratória, que teria ocorrido através da costa do sul Asiático, e esta teria resultado no povoamento inicial do Sudeste da Ásia e Oceania (REICH *et al.*, 2011). Pesquisas realizadas com mtDNA (de herança materna) corroboraram a existência de uma única rota de migração que seguiu pela Índia, Sudeste da Ásia e Austrália (Figura 2). Além disso, a mesma teria seguido em direção ao interior do continente Asiático, num processo rápido e gradativo (FORSTER, 2005).

Figura 2 – Representação da hipótese da existência de apenas uma rota de migração.



Legenda: As setas vermelhas representam a rota pela costa do Sudeste Asiático e as setas azuis representam a dispersão para o interior do continente.

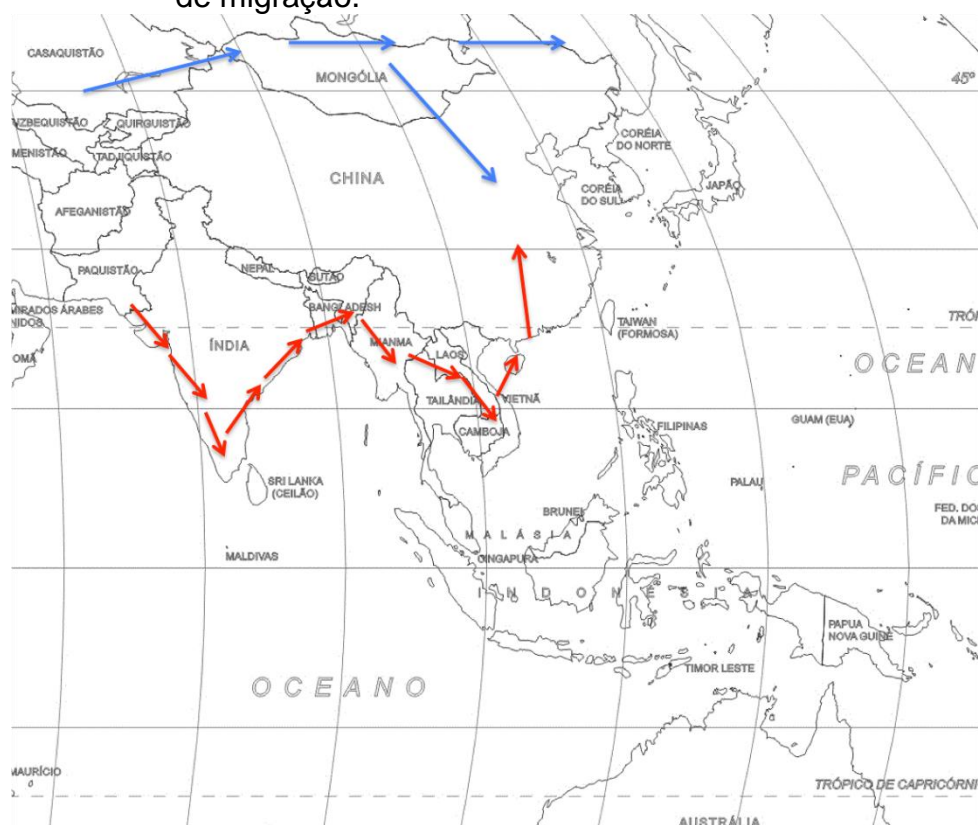
Fonte: A Autora, 2019.

Segundo Mellars (MELLARS, 2006), o foco dos argumentos apresentados para afirmar que existiu uma única migração, seria a existência de uma diversidade genética das populações modernas europeias e asiáticas muito menor que das africanas, diferença essa que seria impossível conciliar com um modelo de dois eventos de dispersão com rotas distintas.

Em 2009, a Organização do Genoma Humano (HUGO) Pan-Asian SNP Consortium trabalhou em conjunto com 11 países e regiões, realizando um mapeamento de 55.000 polimorfismos de nucleotídeo único (SNP, do inglês, *Single Nucleotide Polymorphism*), num projeto que englobou amostras de 73 populações asiáticas e afirmou a existência de apenas uma rota de migração. Os autores deste trabalho concluíram que os humanos migraram ao longo de uma rota costeira do Oriente Médio para o Sudeste Asiático, e daí para o norte, buscando adaptar-se gradualmente ao clima mais severo (NORMILE, 2009).

Em contrapartida, estudos feitos em cromossomo Y de habitantes do Leste Asiático apoiam tanto a importância da rota do sul, ao desempenhar um papel no estabelecimento humano moderno da Ásia oriental, quanto a existência também de expansões demográficas em direção ao Leste da Ásia, através de uma rota do norte (CONSORTIUM, 2009). Algumas pesquisas apontam para a existência de duas grandes ondas de migração do Oriente Médio (Figura 3), ao norte, de imigrantes que viajaram para o leste, através do estepe da Eurásia; e uma outra, distinta e separada, costeira, em direção ao sul, ao redor da borda da Índia, e que seguiria de ilha em ilha para toda a Indonésia, Malásia e Filipinas até ao Pacífico (ZHONG, 2011).

Figura 3 – Representação da hipótese da existência de duas rotas de migração.



Legenda: Rota pela costa do Sudeste da Ásia (setas vermelhas). Rota mais ao norte em direção ao leste Asiático (setas azuis).

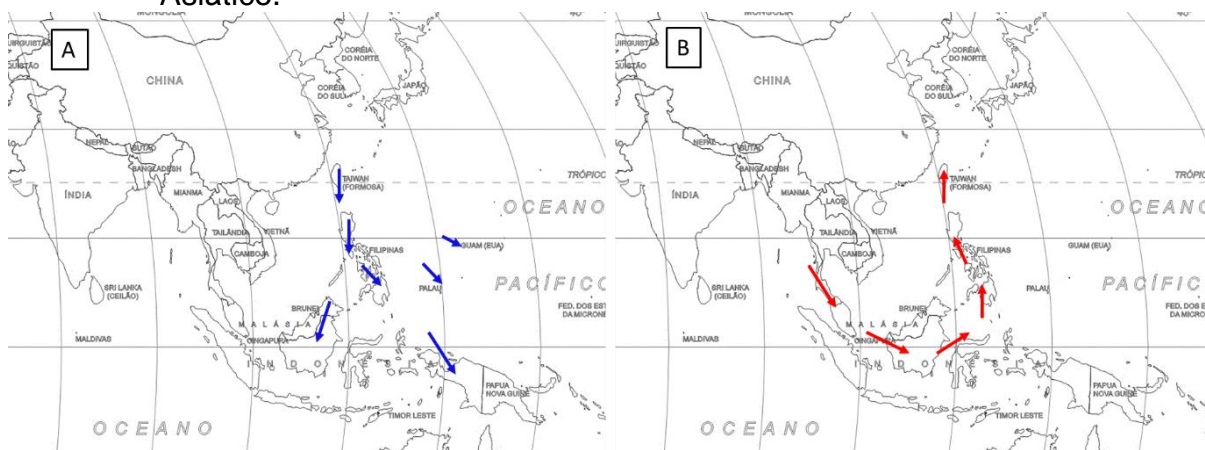
Fonte: A Autora, 2019.

Outro ponto que também vem sendo discutido é como ocorreram as migrações para as ilhas do Pacífico. Com Taiwan como provável ponto de partida, uma teoria, também conhecida como “*Out of Taiwan*”, sugere que esta rota se expande para o sul, ao longo das ilhas do Sudeste Asiático e ilhas da Oceania, trazendo com ela a agricultura, pecuária caseira e as línguas Austronésias. Este modelo de migração assume uma rápida expansão de Taiwan à Polinésia, com pouca ou nenhuma mistura entre as populações em expansão e as já existentes (JINAM *et al.*, 2012). Por outro lado, existe também o modelo “Sudeste Asiático”, o qual indica uma origem Austronésia, provavelmente do leste da Indonésia ao invés de Taiwan. Há, por último, uma hipótese de migração mais complexa, onde se propõe que teriam ocorrido as duas ondas descritas sobre o povoamento das ilhas do Sudeste Asiático e não apenas uma delas (JINAM *et al.*, 2012).

Trejaut e colaboradores (TREJAUT *et al.*, 2014) investigaram o padrão de distribuição da diversidade específica das linhagens paternas em Taiwan, Filipinas e Indonésia, para tentar esclarecer se os mesmos seriam consistentes com a teoria da

migração “Sudeste Asiático” ou “*Out of Taiwan*”, ou até mesmo se teria ocorrido algo mais complexo como dispersões nos dois sentidos por todas as ilhas do Sudeste Asiático. Os resultados sustentam a existência de duas rotas de dispersão, ambas com origem na região continental: a primeira com direção nordeste, passando por Taiwan, e a outra no sentido Sudeste, Indochina, passando pela Indonésia. Estes autores observam ainda uma alta diversidade do cromossomo Y nas Filipinas, que atribuem ao fluxo de genes vindo de comerciantes de vela da costa Sudeste da China ou Indochina Oriental (Vietnã).

Figura 4 – Representação das hipóteses das rotas de migração das Ilhas do Sudeste Asiático.



Legenda: A) Hipótese *Out Of Taiwan*. B) Hipóteses do Sudeste Asiático

Fonte: A Autora, 2019.

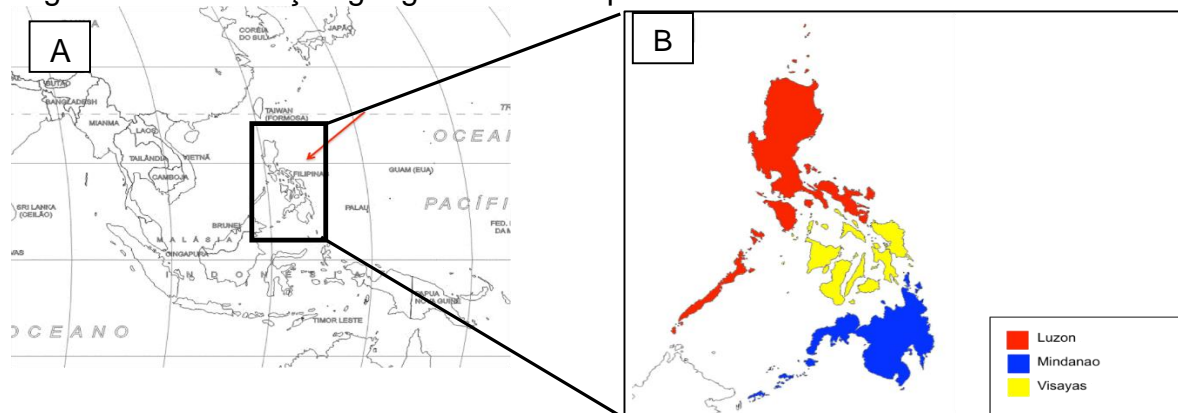
## 1.2 Filipinas

As Filipinas, oficialmente conhecidas como a República das Filipinas, são o sétimo país mais populoso da Ásia com cerca de 93 milhões de habitantes (DETROIT *et al.*, 2013). O arquipélago das Filipinas está localizado no Oceano Pacífico Ocidental, sendo composto por 7.107 ilhas e categorizado sob três principais divisões geográficas: Luzon, Visayas e Mindanao (Figura 5).

Luzon e Mindanao são os maiores grupos de ilhas, representando 95 % das milhas quadradas que compõem as Filipinas atualmente (DELFIN *et al.*, 2011). A ilha de Luzon tem ganhado destaque na literatura e estudos arqueológicos, devido a importância dos fósseis encontrados em suas cavernas. Nesta ilha, estão presentes

as cavernas de Tabon e Callao. Na primeira, localizada no Quezon, foi encontrada uma calota craniana, considerada por muito tempo como o registro mais antigo de fóssil humano das Filipinas (SKALETSKY *et al.*, 2003). Na segunda, na região de Cagayan, foi encontrado um metatarso humano, que é a evidência direta mais antiga da presença humana nas Filipinas (SKALETSKY *et al.*, 2003). Estudos interligados de Arqueologia e Genética podem contribuir diretamente para o preenchimento da lacuna que existe na história entre o período dos fósseis e as populações existentes (SKALETSKY *et al.*, 2003). Nas Filipinas, destacam-se dois eventos demográficos significativos, ocorridos na região da Ásia-Pacífico. O primeiro foi a colonização inicial de Sunda e Sahul, cerca de 40-50 mil anos atrás, para os registros fósseis citados acima, e o segundo, refere-se à expansão Austronésia (DELFIN *et al.*, 2011).

Figura 5 – Localização geográfica das Filipinas.



Legenda: A) Mapa do Sudeste Asiático com destaque nas ilhas das Filipinas. B) Ilhas das Filipinas ampliadas mostrando os três principais grupos (Luzon em vermelho, Visayas em amarelo e Mindanao azul).

Fonte: A Autora, 2019.

A grande maioria (senão toda) da população indígena das Filipinas fala filipino, língua pertencente à família linguística Austronésia (BELLWOOD; FOX; TRYON, 2006). Apesar da existência de outras línguas anteriores nas ilhas do Sudeste Asiático, a Austronésia tornou-se dominante, enquanto as línguas ancestrais das ilhas foram desaparecendo. Existem algumas hipóteses sobre como ocorreu a expansão dessa família linguística nessas ilhas, sendo sua origem amplamente discutida (KLAMER, 2019). A hipótese aceita pela grande maioria dos pesquisadores linguísticos, e suportada por dados genômicos, é conhecida por “*Express Train*”. Esta hipótese indica uma origem em Taiwan (JOBILING; HURLES; TYLER-SMITH, 2013), onde observa-se uma maior diversidade das línguas da família Austronésia (WEI *et*

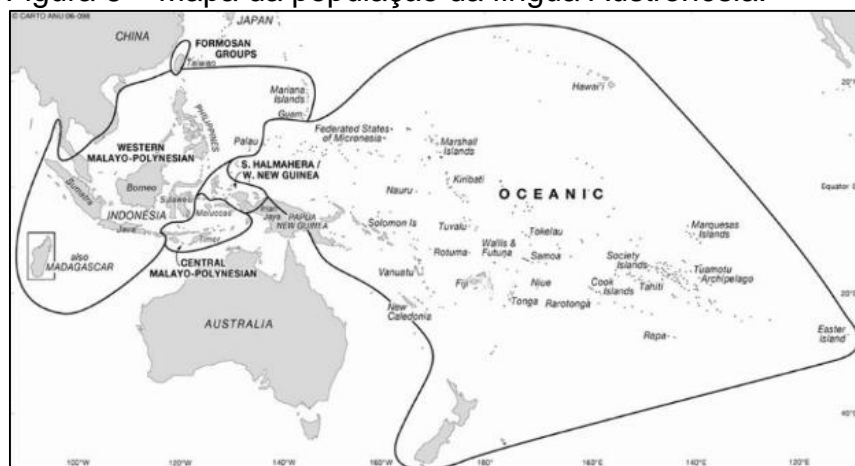


*al.*, 2017). Segundo essa hipótese, durante os últimos milhares de anos, houve uma migração rápida de pessoas de linguagem Austronésia, partindo de Taiwan em direção as ilhas do Pacífico (KAYSER *et al.*, 2000). Já a hipótese conhecida por “*Entangled-bank*” não considera apenas uma única onda migratória rápida, mas várias ondas longas, além de uma complexa interação entre Sudeste Asiático, Melanésia e Polinésia (KAYSER *et al.*, 2000) dando origem a uma das maiores famílias de línguas faladas pelo mundo, que abrange aproximadamente 351 milhões de pessoas (WAHAB *et al.*, 2018).

A família Austronésia possui cerca de 1200 idiomas e está amplamente representada em diversas ilhas do Sudeste Asiático (Figura 6), inclusive nas Filipinas.

O Tagalog é uma língua falada por 20 milhões de pessoas nas Filipinas e pertencente ao ramo Malaio Polinésio da família Austronésia (<https://www.ethnologue.com/>). A história das linguagens das Filipinas está fortemente ligada a uma história de colonização e uma busca de uma unificação da língua. Em 1937, o Instituto Nacional de Linguagem (Surian ng Wikang Pambansa), na tentativa de unificar as ilhas Filipinas como um Estado-nação, declarou o Tagalog como língua oficial (LESADA, 2017), por ser uma língua falada/entendida em todas as regiões, por estar extremamente presente na literatura e também ser a linguagem de Manila (capital econômica e política) (BARRIOS, 2014).

Figura 6 – Mapa da população da língua Austronésia.



Fonte: (PAWLEY, 2006).

### 1.3 Vietnã

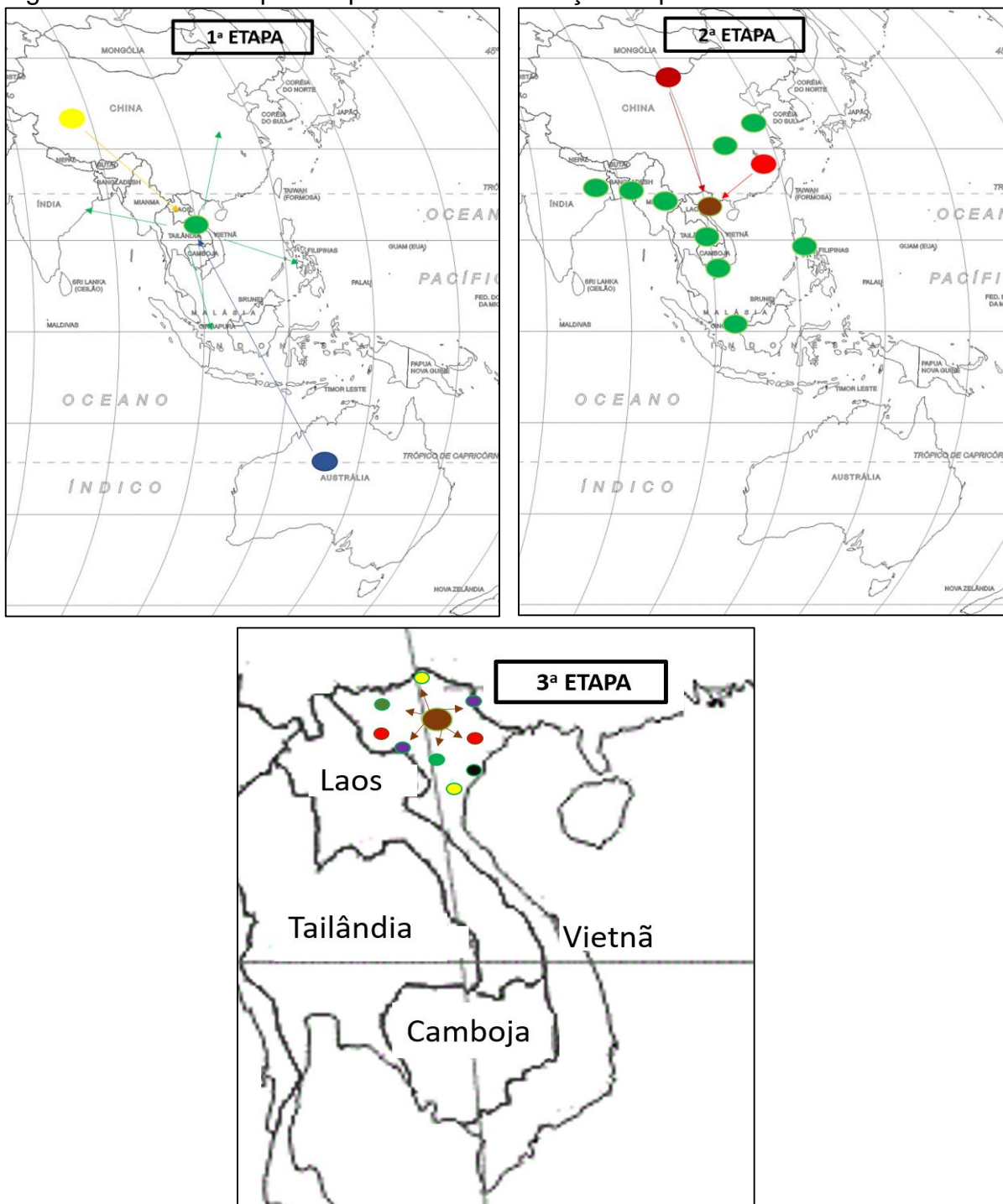
A República Socialista do Vietnã é um país tropical localizado no Sudeste Asiático que faz fronteira com a China (Norte), Laos (Noroeste) e Camboja (Sudoeste). A população do Vietnã tem cerca de 89 milhões de habitantes, classificando-a entre os países com maiores densidades populacionais do mundo (DANG, 2012; JOBLING; TYLER-SMITH, 2003). Do ponto de vista étnico, o Vietnã representa uma das populações mais diversas e complexas de toda a Ásia, com 54 comunidades étnicas oficiais (JOBLING; TYLER-SMITH, 2003; PARKER, 2005).

O processo de formação do povo vietnamita pode ser dividido em 3 etapas (Figura 7).

Na segunda etapa, cerca de 5 mil anos atrás, no norte do Vietnã, essa população Malaia teve em contato frequente com populações que migraram do Norte da China e sul do Rio Yantze, formando assim uma nova população.

Na terceira etapa, essa nova população foi dividida em uma série de grupos, chamados pelos chineses de Bach Viet (<https://www.alotrip.com/about-vietnam-people/origin-vietnamese-people>).

Figura 7 – As três etapas do processo de formação do povo vietnamita.



Legenda: Na primeira etapa, uma população do Tibete (círculo amarelo) e uma da Austrália (círculo azul) migraram (setas coloridas) para a Indochina dando origem ao grupo Malaio (círculo verde na segunda etapa, no Norte do Vietnã, essa população Malaia teve contato com outras do Norte da China (vermelha escura) e do Sul do Rio Yantze (vermelha clara), formando uma nova população que, na terceira etapa, dividiu-se em diversos grupos.

Fonte: A Autora, 2019.

De tempos em tempos, esses processos de divisão e misturas formavam grupos étnicos ao mesmo tempo que modelavam a composição genética do Sudeste Asiático (PISCHEDDA *et al.*, 2017). Hoje em dia, a literatura classifica politicamente a

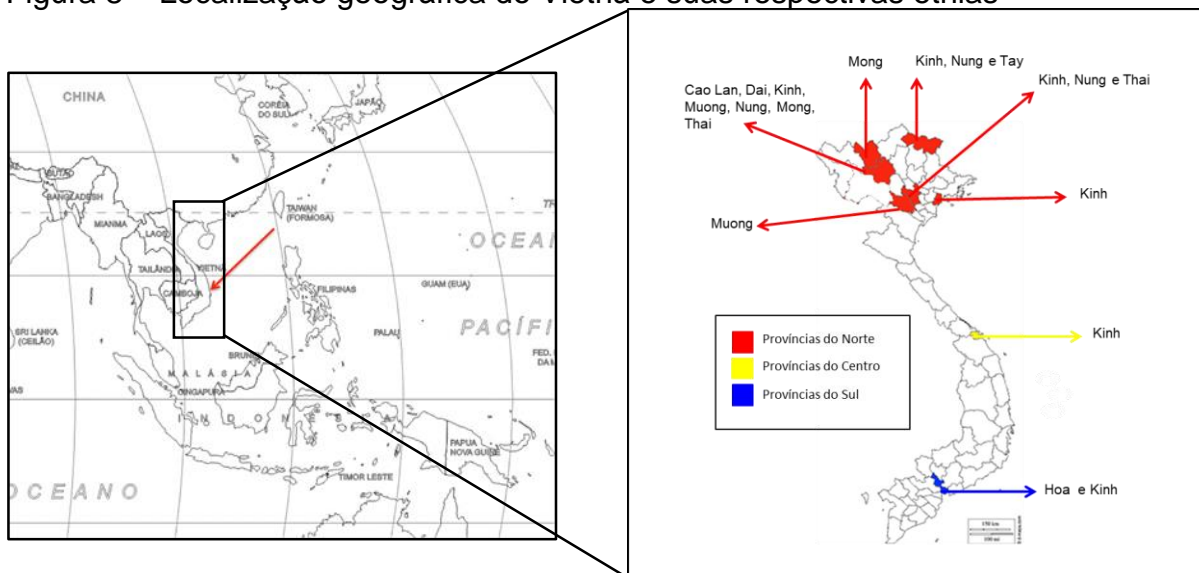
composição étnica do Vietnã em dois grupos, separando o “grupo étnico majoritário” dos “grupos étnicos minoritários”.

O grupo étnico majoritário é composto pela etnia Kinh e é o mais beneficiado, com acesso, por exemplo, a educação e benefícios de programas governamentais (DANG, 2010). Essa é a maior e principal etnia vietnamita e vive em planícies do rio Vermelho, onde ficou conhecida como Kinh (que significa “capital”), em referência à capital vietnamita Hanoi, por onde passa grande parte desse rio (TAYLOR, 2013). De acordo com o Censo de População e Habitação Vietnã de 2009 (<http://www.gso.gov.vn>), este grupo étnico tem uma população de 73.594.427 indivíduos, representando 85,7 % do total da população do Vietnã. Pessoas Viet ou Kinh se originaram de uma área geográfica que hoje faz parte do norte do Vietnã e o sul da China e podem ser encontradas tanto por todo o país como em outros países do Sudeste Asiático (KEYES, 2002).

Pessoas da etnia Kinh se concentram principalmente em áreas de planície, nas quais essa geografia favorece a cultura de arroz (seu principal meio de subsistência) (PARKER, 2005) e são relativamente homogêneos nos quesitos linguagem, alimentação e outros aspectos culturais (BANISTER; CENSUS, 1985). A língua mais importante do país é o vietnamita, que é a língua nativa dos Kinh (JOBBLING; TYLER-SMITH, 2003).

O segundo grupo (que inclui os “grupos étnicos minoritários”) tem taxas de pobreza e mortalidade infantil muito maiores que as do grupo majoritário, além do aumento do analfabetismo, saúde precária e falta de acesso aos demais serviços públicos (DANG, 2010). Este grupo é composto pelos demais grupos étnicos vietnamitas e no presente trabalho estão representados os grupos minoritários Tay, Thai, Mong, Cao Lan, Dao, Hoa, Muong e Nung, do Norte do Vietnã, e Hoa, pertencente a um grupo minoritário do Sul (Figura 8).

Figura 8 – Localização geográfica do Vietnã e suas respectivas etnias



Legenda: A etnia Kinh se encontra distribuída em todas as regiões. As etnias Dao, Mong, Muong, Nung, Cao Lan, Thai, Tay estão nas províncias da região norte e Hoa está em uma província da região sul.

Fonte: A Autora, 2019.

Os Hoa são um grupo de pessoas que vêm da China e que, de acordo com a classificação da República Popular da China, pertencem ao grupo étnico Han. No Vietnã, eles representam uma população de 823.071 pessoas (0,96 %), que se concentram na cidade de Ho Chi Minh City (414.045 habitantes, representando 50,3 % do total da população Hoa, no Vietnã).

Os Muong representam uma população de 1.268.963 pessoas que habitam a região montanhosa no Norte do Vietnã, sendo geralmente encontrados nas províncias de Hoa Binh e Than Hoa (<https://www.refworld.org/docid/5bb5db7b7.html>) e também no país Laos, onde os Muong migraram no período da Guerra do Vietnã (SCHLIESINGER, 2015).

O Tay é o segundo maior grupo étnico no Vietnã, com uma população de 1.626.392 pessoas (1,895 %) (<http://www.gso.gov.vn>) sendo que a sua maioria vive nas províncias montanhosas do norte do rio Vermelho (*Ethnic groups in Viet Nam : an analysis of key indicators from the 2009 Viet Nam population and housing census / UNFPA, 2011*).

Os Nung têm uma relação estreita com os Tay e os Choang (Zhuang). Este grupo tem uma população de 968.800 pessoas (1,123 %) (<http://www.gso.gov.vn>) que está concentrada, principalmente, na província do nordeste da região de Tonkin, ao longo da fronteira com a China.

O grupo Hmong ou Mong, migrou da China para o Vietnã no século XIX e, segundo o censo de 2009 (<http://www.gso.gov.vn>) é o sexto maior grupo no Vietnã com uma população de 1.068.189 pessoas (1,244 %) e considerado um dos mais desprivilegiados economicamente de todos os povos vietnamitas. Os Mong vivem dispersos em áreas montanhosas do Norte do Vietnã, a altitudes de 800 a 1500 m acima do nível do mar, em regiões fronteiriças com a China. Essa etnia é representada por vários grupos como como os Black, Flowers Hmong, White, Green, cada um vestindo uma versão levemente diferente da vestimenta vietnamita tradicional (PARKER, 2005).

Os Dao também têm uma origem chinesa, tendo migrado para a Indochina, onde ficaram conhecidos por Dao ao Norte do Vietnã, Lantien Yao em Laos e também na Tailândia por Mien (JONSSON, 2001). Eles possuem uma população de 751.067 habitantes (MICHAUD; SWAIN; BARKATAKI-RUSCHEWEYH, 2016).

O grupo étnico Cao Lan ou Sán Chay vive ao norte do Vietnã e fala a língua Tai. Apesar dessa etnia pertencer ao grupo Central da família de língua Tai, tiveram pouco ou nenhum contato com as comunidades Tai localizadas na Tailândia (TIMSUKSAI; RAMBO, 2016).

O grupo étnico Thai possui cerca de 1.550.423 habitantes (2009) concentrados na parte noroeste e oeste do Norte do Vietnã (<https://www.refworld.org/docid/5bb5db7b7.html>).

Segundo o Portal do Governo do Vietnã (<http://chinhphu.vn/portal/page/portal/English>), a comunidade étnica vietnamita também pode ser classificada em 8 grupos diferentes de acordo com seu idioma (Quadro 1). Segundo esta classificação, as 9 populações incluídas neste estudo pertencem a quatro grupos diferentes, nomeadamente *Việt-Mường*, *Tày-Thái*, *Mông-Đao* e *Chinese*.

Quadro 1 – Classificação vietnamita de acordo com a língua.

Việt-Mường	Grupo: Chút, <b>Kinh</b> , <b>Muong</b> , hồ
Tày-Thái	Grupo: BỐ Y, Giáy, Lào, Lự, <b>Nung</b> , <b>Cao Lan</b> , <b>Tay</b> , <b>Thai</b>
Mon-Khmer	Grupo: Ba na, Brâu, Bru-Vân kiều, Chơ-ro, Co, Cơ-ho, Cơ-tu, Cơ-tu, Gié-triêng, Hrê, Kháng, Khmer, Kho mú, Mạ, Mảng, M'Nông, O'-đu, Rơ-măm, Tà-ôi, Xinh-mun, Xơ-đăng, Xtiêng
Mông-Dao	Grupo: <b>Dao</b> , <b>Mong</b> , Pà Thên
Kađai	Grupo: Cờ Lao, La Chí, La ha, Pu Péo
Austro-Polynesian	Grupo: Chăm, Chu-ru, Ê đê, Gia-rai, Ra-glai
Chinese	Grupo: <b>Hoa</b> , Ngái, Sán Dìu.

Legenda: Os grupos étnicos do presente trabalho estão destacados em vermelho: Kinh, Muong, Nung, Cao Lan (San Chay), Tay, Thai, Dao, Mong e Hoa. Línguas e grupos escritos em cor preta possuem as letras originais do alfabeto vietnamita. Nos grupos destacados em vermelhos algumas letras foram substituídas por correspondentes próximas do alfabeto português.

Fonte: A Autora, 2019.

#### 1.4 Marcadores genéticos

O polimorfismo genético surge nas populações como resultado de mutações. Uma mutação é uma alteração na sequência do DNA que pode ocorrer por erros na incorporação dos nucleotídeos durante a replicação e podem ser herdadas dos pais ou adquiridas ao longo da vida (KLEIN *et al.*, 2019; SCHUMACHER; VIJG, 2019).

Um *locus* diz-se polimórfico numa população sempre que existirem no mínimo dois alelos com frequências superiores a 0,01 (1 %) (KARKI *et al.*, 2015; LAPPALAINEN *et al.*, 2019). Além disso, existem alelos que por apresentarem uma frequência inferior a 1 % na população são denominados alelos raros, sendo estes presentes na maioria dos genes em populações grandes (RAYCHAUDHURI, 2011).

Os polimorfismos do DNA podem ser empregados como marcadores genéticos. A utilização desses marcadores vai depender de alguns aspectos, como sua localização cromossômica (e conseqüentemente seu modo de transmissão), tipo de variação que representam e distribuição da variabilidade intra e interpopulacional (BAYE, 2011).

Os polimorfismos podem ser caracterizados quanto a sua sequência ou quanto ao seu tamanho (comprimento) (LIU, 2018). Os polimorfismos de sequência surgem através da substituição de um ou mais nucleotídeos em uma determinada região do DNA, diferindo apenas na composição de bases (LIU *et al.*, 2014). Já os polimorfismos de comprimento são gerados através da inserção/deleção de um ou mais nucleotídeos na sequência do DNA, resultando, conseqüentemente, em variações de aumento ou diminuição na cadeia de DNA (MILLS *et al.*, 2006).

Os SNPs são os polimorfismos de sequência mais frequentes e consistem na troca de uma única base por outra (LANDER *et al.*, 2001; LAPPALAINEN *et al.*, 2019). De acordo com o tipo de base, os SNPs podem ser classificados em transições e transversões. As transições resultam de trocas observadas entre duas bases púricas (C por T ou vice-versa) ou entre duas bases pirimídicas (A por G ou vice-versa). Já as transversões vão envolver a troca entre bases púrica e pirimídicas, isto é, de C por G ou A, de T por A ou G, ou vice-versa (ROBERT; PELLETIER, 2018).

Os STRs (do inglês, *Short Tandem Repeats*), também conhecidos como microssatélites, vão se caracterizar por um menor número de bases que formam o motivo repetitivo (geralmente de 2-6 pb) (GYMREK *et al.*, 2017). Esses marcadores tendem a apresentar taxas de mutação mais elevadas que as mutações que ocorrem em regiões não repetitivas, com isso, apresentam geralmente valores altos de diversidade intrapopulacional (BUGOYE; MULIMA; MISINZO, 2018; ECKERT; HILE, 2009; ZHAO *et al.*, 2015).

Os InDels (do inglês, *Insertion deletion polymorphisms*) são variações de comprimento (na ordem das unidades) presentes no genoma gerado por inserções ou deleções de nucleotídeos (LAPPALAINEN *et al.*, 2019). Estes marcadores, tal como os SNPs, são geralmente bialélicos e apresentam taxas de mutação mais baixas que as encontradas para os microssatélites ou STRs. Como resultado, os marcadores bialélicos tendem a apresentar uma menor diversidade intrapopulacional, quando comparados com os STRs (ECKERT; HILE, 2009; XIE *et al.*, 2018).

Dependendo da presença ou ausência de recombinação, os diferentes tipos de genomas disponíveis podem ser separados em dois grupos principais: genomas recombinantes, que permitem a identificação de um único indivíduo (marcadores autossômicos e do cromossomo X) e genomas não recombinantes, que permitem discriminar linhagens parentais e fornecer informações compartilhadas por um grupo



de indivíduos relacionados (cromossomo Y - na maior parte da sua extensão - e genoma mitocondrial) (BUTLER, 2014; COMAS; LUISELLI; RICKARDS, 2018).

Os marcadores de linhagem são transmitidos, de geração para geração, da mãe para o filho/filha (mtDNA) ou do pai para o filho (Cromossomo Y). Além de estudos sobre genética de populações, esses marcadores também são muito utilizados em casos forenses em que a evidência é limitada, para encontrar pessoas desaparecidas, identificar vítimas de desastre, e em algumas situações complexas de parentesco (BUCKLETON; KRAWCZAK; WEIR, 2011; BUTLER, 2014).

O mtDNA é utilizado como ferramenta para testes de identidade genética devido à ausência de recombinação e ao seu alto número de cópias. A tipagem do mtDNA baseia-se na região controle ou sequência do genoma total, sendo muito usada em análises de uma variedade de amostras forenses como ossada, dentes e outras amostras biológicas (AMORIM; FERNANDES; TAVEIRA, 2019). Apesar dos marcadores de linhagem terem um importante papel na identificação humana, eles não são tão usados na rotina dos laboratórios moleculares como os marcadores autossômicos. O surgimento e evolução dos kits comerciais para o estudo de marcadores de linhagem, bem como a criação de bancos de dados de haplótipos (Y-STRs e mtDNA), permitiram aumentar muito o número de amostras populacionais estudadas e, com isso, a confiabilidade de uma análise estatística (BUTLER, 2014).

Atualmente, devido a sua haploidia, os marcadores do cromossomo Y são muito populares nas pesquisas de ancestralidade biogeográfica. O conhecimento sobre esse cromossomo e seu uso na genealogia genética resultou em aplicações nas mais diversas disciplinas, como por exemplo, história familiar, demografia, antropologia, ciências forenses, genética de populações e evolução de cromossomos sexuais, sendo importante em pesquisas multidisciplinares futuras (CALAFELL; LARMUSEAU, 2017).

## **1.5 Cromossomo Y**

O cromossomo Y é o segundo menor cromossomo humano, contendo em média 60 milhões de pares de bases (Mb) (MORTON, 1991). Apesar do seu tamanho,

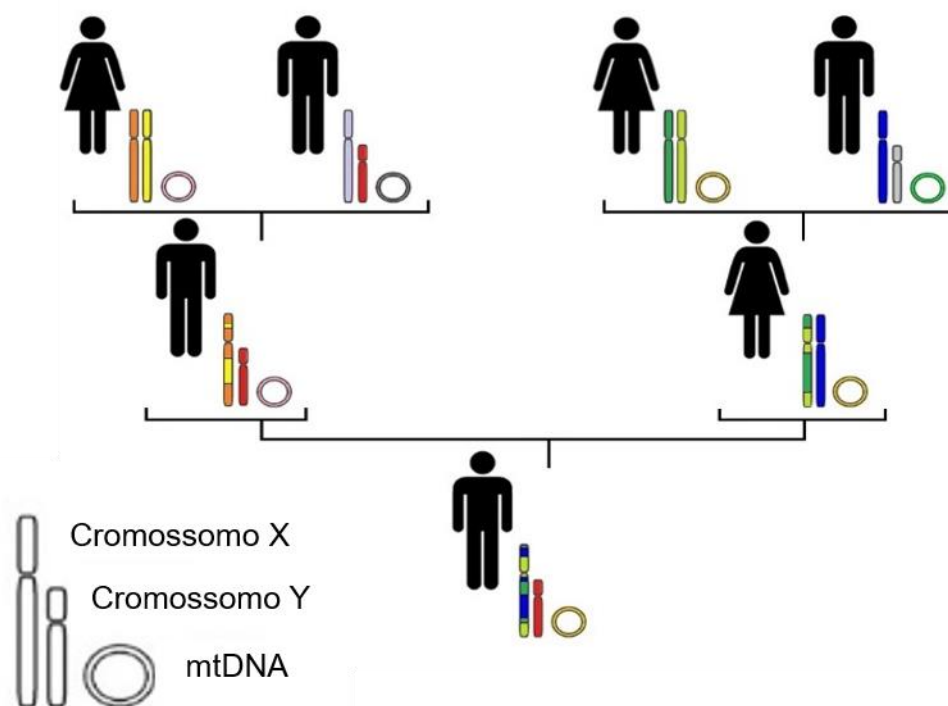
este cromossomo contém os genes necessários para a determinação do sexo masculino (PEREIRA e GUSMÃO 2019).

Pesquisas em genômica revelaram que este cromossomo contém uma região, compreendendo 95 % do seu comprimento, onde não há homologia com o cromossomo X e nem ocorre recombinação. Esta região passou a ser conhecida como Região Não-Recombinante do Cromossomo Y (NRY, do inglês, *Non Recombining region of the Y chromosome*). Porém, com a descoberta de uma elevada recombinação intra-cromossômica na NRY (SKALETSKY *et al.*, 2003), foi proposta uma designação mais apropriada: região masculina específica do cromossomo Y (MSY, do inglês, *Male Specific Y region*).

Em estudos de genética populacional, os “marcadores do cromossomo Y” são atribuídos especificamente àqueles que se localizam na região MSY, excluindo aqueles em regiões pseudoautosômicas (PAR), as quais apresentam homologia com o cromossomo X (JORDE *et al.*, 2000; UNDERHILL *et al.*, 2001). A MSY apresenta certas características que ditam a sua utilidade em estudos populacionais, nomeadamente, está presente apenas em indivíduos do sexo masculino, sendo transmitida aos descendentes do mesmo sexo (Figura 9).

Devido à ausência de recombinação e à presença de apenas um cromossomo Y, a MSY é uma região haploide (JOBBLING, 2001), transmitida em bloco de geração para geração, sendo a mutação a única fonte de variabilidade. O acúmulo de mutações faz com que os cromossomos Y divirjam, dando origem a distintas linhagens. Estas características fazem com que o cromossomo Y possa ser usado para definir linhagens masculinas em populações (JOBBLING; TYLER-SMITH, 2003).

Figura 9 – Comparação de diferentes modos de transmissão genética, incluindo a do cromossomo Y.



Legenda: Os cromossomos Y (cromossomos menores) e mtDNA (círculos) descendem apenas de uma linhagem dos avós, representando a herança patrilinear e matrilinear, respectivamente. Em contraste, os autossomos descendem de todos os avós, gerando segmentos cromossômicos que sofreram recombinações em todas as gerações.

Fonte: (PEREIRA e GUSMÃO, 2019).

O cromossomo Y possui diversos tipos de polimorfismos, tanto de sequência como de comprimento. Marcadores do tipo STR e SNP no cromossomo Y (Y-STR e Y-SNP) são utilizados no estudo de linhagens masculinas. Uma combinação desses marcadores, pode nos fornecer dados importantes, como a origem continental do doador masculino, além de nos permitir inferir padrões de migração em populações humanas, mediados por indivíduos do sexo masculino.

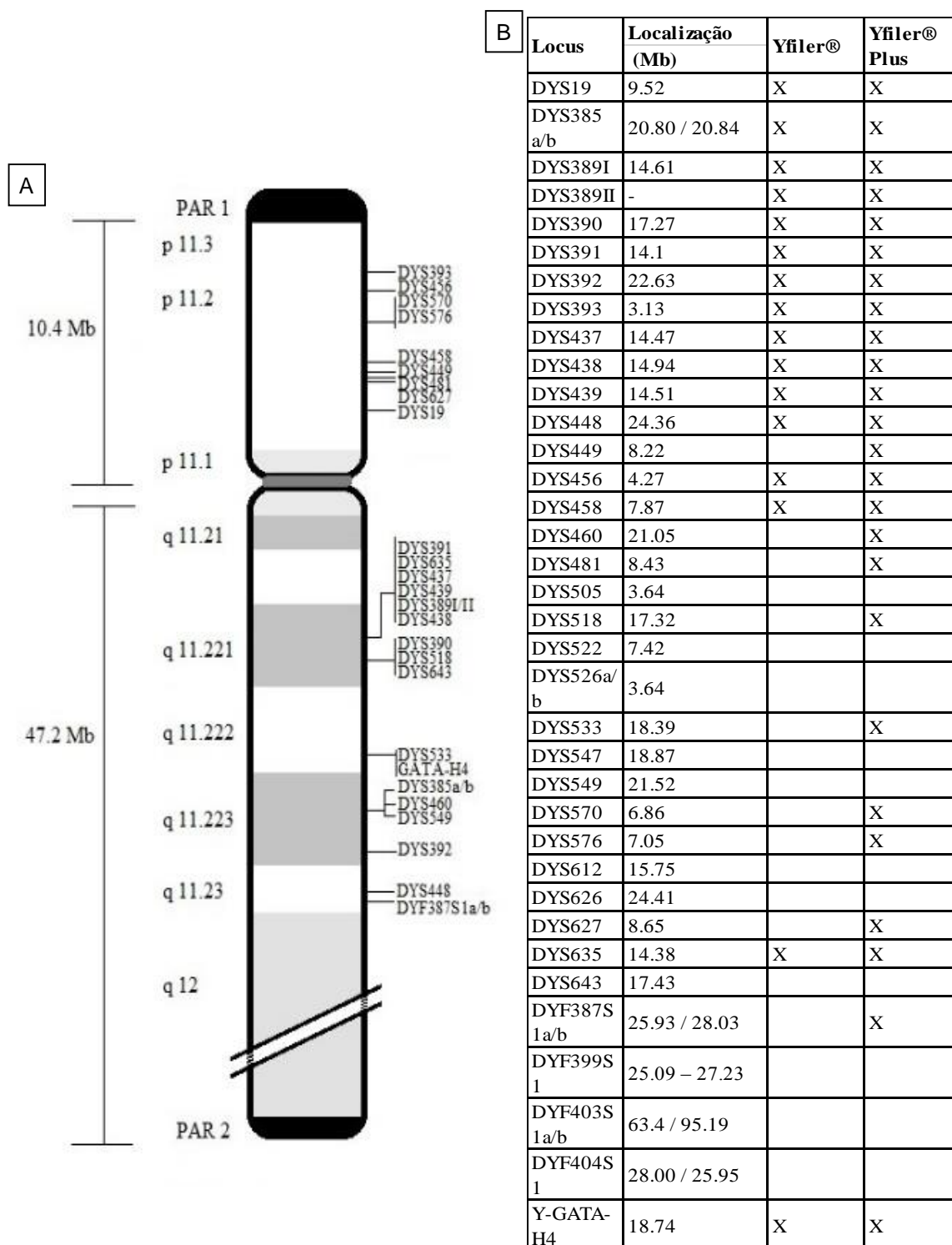
O conjunto de informações obtidas com base em Y-STRs é denominado de haplótipo, os quais, quando analisados em um número elevado de marcadores (*multiplex*), tendem a ser altamente diversos nas populações. Esses *Multiplex* são encontrados em forma de kits, sendo utilizados em diversas estratégias e diferenciados de acordo com o número de marcadores (Figura 10) (BELEZA *et al.*, 2003; BUTLER, 2014; LI, 2015).

Com a necessidade de se conhecer a distribuição das frequências, haplotípicas a nível mundial, para aplicações forenses, foi criado um banco de dados *online*, de

livre acesso, chamado YHRD (*Y-Chromosome STR Haplotype Reference Database*; [www.yhrd.org](http://www.yhrd.org)) pertencente ao Instituto de Medicina Legal, na Universidade Humboldt de Berlim (Alemanha). Esse banco de dados contém perfis de haplótipos dos marcadores mais utilizados em populações humanas e, também, mapas que fornecem informações sobre a distribuição geográfica dos haplótipos pesquisados.

O YHRD é uma ferramenta avançada, tendo como principal característica a função de busca, que fornece aos usuários uma estimativa das frequências dos haplótipos (Y-STRs) em todas as populações, suficientemente, grandes e homogêneas. Sendo assim, é bastante aplicável em amostras do continente Europeu, mas ainda não foi abordado continente Asiático (LESSIG *et al.*, 2003).

Figura 10 – Marcadores Y-STR mais utilizados.



Legenda: A) Marcadores Y-STR mais utilizados em estudos de genética populacional e forense, bem como sua localização no cromossomo Y. B) Dois *Multiplex* utilizados no presente trabalho e seus marcadores específicos.

Fonte: (VÂNIA PEREIRA, 2019).

Os Y-STRs apresentam uma taxa de mutação elevada, permitindo determinar eventos de diversificação populacional mais recentes, sendo por isso mais eficazes no estudo de grupos populacionais com uma história comum mais próxima. Já os Y-SNPs, por apresentarem uma taxa de mutação muito menor que a dos Y-STRs, permitem estabelecer os denominados haplogrupos (ou linhagens masculinas), contendo diferentes haplótipos que divergiram por mutação.

Diversos estudos sobre os SNPs permitiram a construção de uma árvore filogenética detalhada. Em 2002, o YCC (*Y Chromosome Consortium*) publicou a primeira filogenia correlacionando as linhagens do cromossomo Y, sendo atualizado subsequentemente em 2003. Com base nesta filogenia, o grupo descreveu 19 haplogrupos, nomeando-os com letras de A a R. Esse Consórcio estabeleceu um conjunto de regras visando a unificação da nomenclatura, permitindo a inclusão de mutações e haplogrupos ainda a serem descobertos (CONSORTIUM, 2002). Desde então, milhares de outros polimorfismos bialélicos foram descobertos, sendo adicionadas mais duas letras (S e T), em 2008 (Quadro 2) (KARAFET *et al.*, 2008). A filogenia e nomenclatura mais atual e utilizada é a sugerida por Van Oven e colaboradores, em 2014 (VAN OVEN *et al.*, 2014), que nomeia os haplogrupos utilizando a mutação mais estável que dá origem à ramificação (Figura 11).

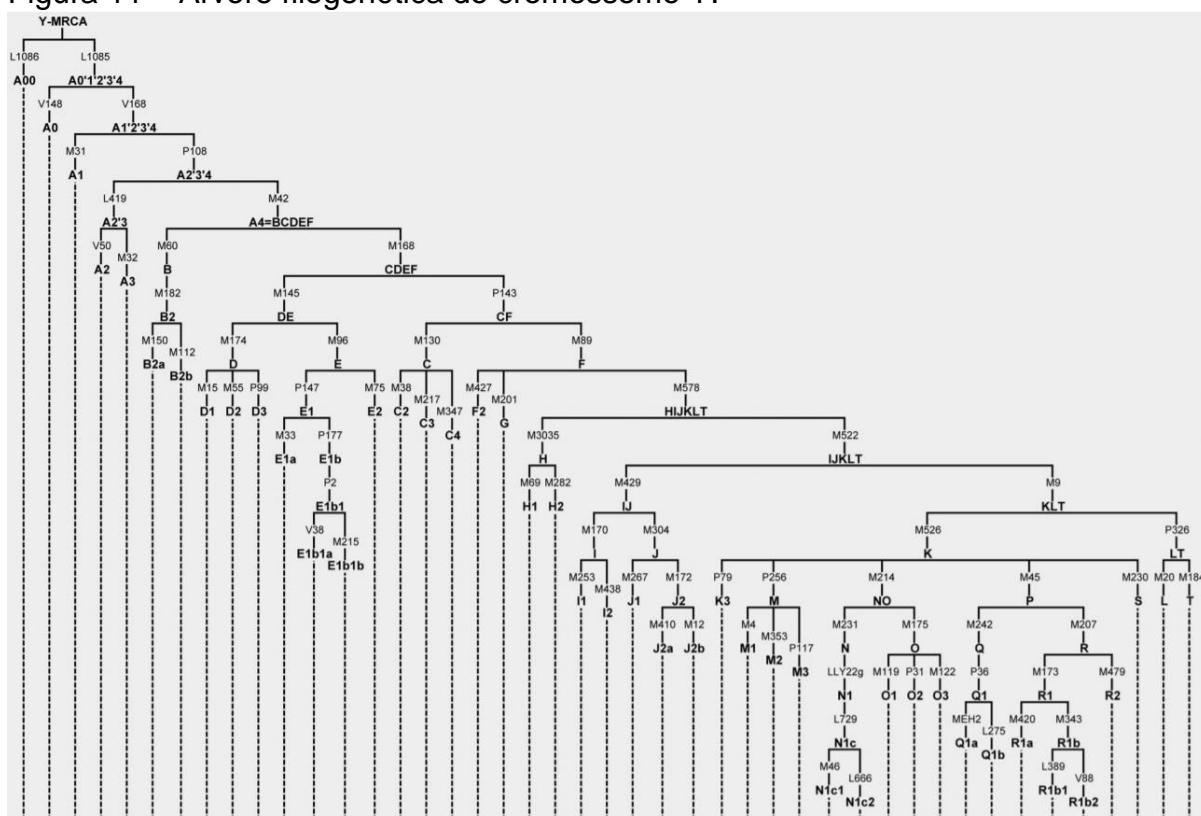
Quadro 2 – Haplogrupos com os relativos marcadores e localização.

Paragruppo	Marcadores	Localização
A	M91 e P97	Africano
B	M60, M181, P85 e P90	África, região subsariana
C	RPS4Y711, M216, P184, P255 e P260	Ásia, Austrália, and Oceania
D	M174 e JST021355	Ásia Central (Tibet), Japão, Sudeste Asiático e ilhas de Andamão
E	M96, P29, P150, P152, P154, P155, P156, P162, P168, P169, P170, P171, P172, P173, P174, P175	África, Oriente Médio, Sul da Europa, Sul e Centro da Ásia
F	F-P91/P104, F-M427/M428, F-P96, e F-P254	Leste da Ásia e Países Baixos
G	M285 e M342	Oriente Médio, Mediterrâneo e Montanha Causcasiana
H	M69	Índia
I	P30, P40, M253, M307, e M450	Europa
J	2F2a, M304, e P209	Oriente Médio, Norte da África, Europa, Ásia Central, Paquistão, e Índia
K	M9, P128, P131, e P132	Indonésia e Austrália
L	M11, M20, M22, M61, M185, e M295	África, Oriente Médio, Ásia Central, Norte África, Europa e costa do Mediterrâneo
M	M4, M5=P73, M106, M186, M189, M296 e P35	Oceania e Leste da Indonésia
N	M231	Norte da Eurásia
O	(M175, P186, P191, e P196)	Leste da Ásia, Ásia Central e Oceania
P	Distribuído entre linhagem Q e R	
Q	M242	Norte da Eurásia, Sibéria, Europa, Leste da Ásia e Oriente Médio
R	M207, M306, P224, P227, P229, P232, P280, e P285	Europa
S	M230, P202, e P204	Oceania e Indonésia
T	M70, M193, M272 e USP9Y + 3178 = M184	Oriente Médio, África e Europa

Legenda: Destaque em vermelho dos haplogrupos utilizados para as análises do presente trabalho.

Fonte: Quadro adaptado do artigo (KARAFET *et al.*, 2008).

Figura 11 – Árvore filogenética do cromossomo Y.



Legenda: Árvore filogenética mínima evidenciando os marcadores (SNP) mais basais.

Fonte: (VAN OVEN *et al.*, 2014).

Alguns dos haplogrupos estão confinados a áreas ou continentes específicos (ex. os haplogrupos A e B na África), enquanto outros mostram extensa representação geográfica. Essa condição tem sido usada para explorar padrões de migração, subestrutura da população e mistura entre diversas populações; permitindo, além disso, fazer inferências sobre a possível origem geográfica de uma amostra de interesse, de origem desconhecida (GUSMÃO; BRIÓN; GOMES, 2007).

No Sudeste da Ásia, há 4 haplogrupos dominantes: O-M175, C-M130, D-M174 e N-M231 (ZHONG *et al.*, 2010).

O haplogrupo O (Figura 12) deriva-se do NO-M214, um dos ramos mais difundidos na Eurásia. Esse haplogrupo é relativamente antigo e abrange a grande maioria das linhagens masculinas do leste e Sudeste da Ásia. O haplogrupo O-M175 deu origem a três principais sub-haplogrupos (O1a-M119, O2-L690 e O3-M122), totalizando 60 % dos cromossomos Y entre as populações do Leste Asiático (WANG; LI, 2013).

O sub-haplogrupo O1a-M119 é prevalente ao longo da costa Sudeste da China, ocorrendo em altas frequências em pessoas que falam Daic e em aborígenes de Taiwan (WANG; LI, 2013).

O sub-haplogrupo O2a1-M95, derivado de O2-L690, é o subclado mais frequente de O2, que é o principal haplogrupo da Indochina, e também encontrado em muitas populações localizadas no sul da China e no leste da Índia (WANG; LI, 2013).

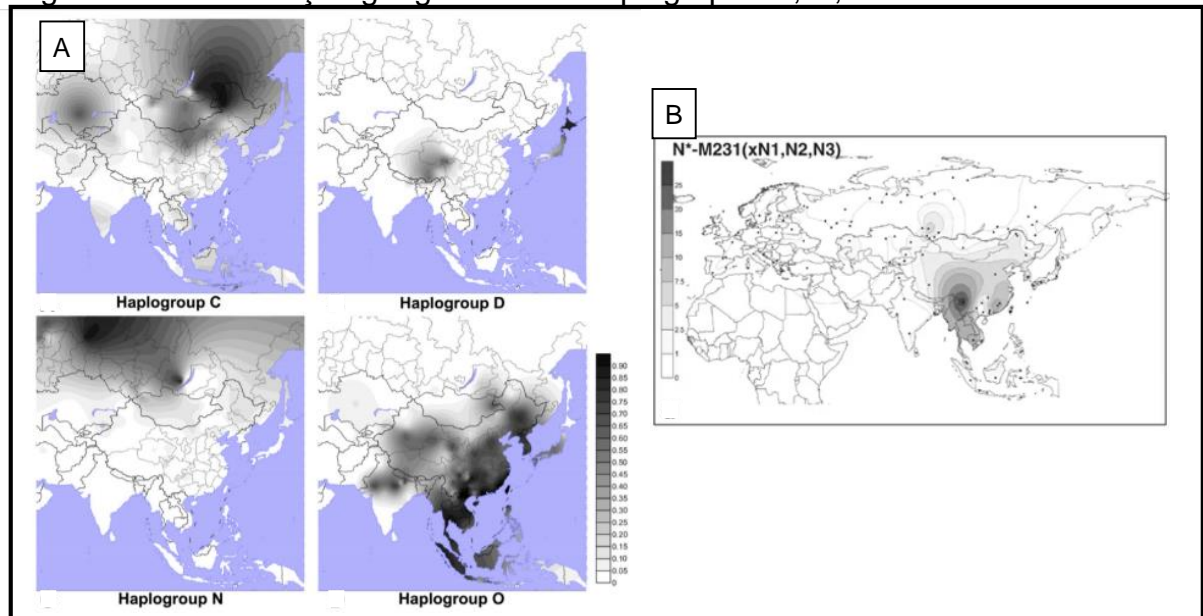
O sub-haplogrupo O-P201 é derivado de O3-M122 e conecta Taiwan e Polinésia, embora este apresente uma frequência relativamente baixa na Polinésia (BELLWOOD; NESS, 2014).

No haplogrupo N (também derivado do NO-M214), apesar de alguns estudos o citarem como presente predominantemente no Norte da Ásia, sua mutação M231(xN1,N2,N3) pode ser encontrada no Sudeste Asiático, incluindo Indochina (Figura 12).

O haplogrupo C-M130 possui uma alta frequência na Ásia, Oceania, baixa na Europa e América e é ausente na África (Figura 12). Por ser ausente na África, essa linhagem é altamente informativa para estudos de rotas de migração na saída do homem da África na pré-história. Estudos recentes realizados na Austrália e Nova Guiné mostraram que o haplogrupo C é provavelmente um dos primeiros tipos fundadores fora da África (ZHONG *et al.*, 2010).



Figura 12 – Distribuição geográfica dos Haplogrupos C, D, N e O.



Legenda: O Haplogrupo N é citado na literatura como encontrado tanto no Norte da Ásia (A), quanto no sul próximo a Indochina (B).

Fonte: A) (WANG; LI, 2013) B) (ROOTSI *et al.*, 2007).

Todas as teorias sobre rotas de migração têm sido abordadas através do estudo de marcadores genéticos com características muito diversas, incluindo genomas recombinantes e não recombinantes. Os marcadores de linhagem permitem inferir mais facilmente filogenias, sendo por isso especialmente úteis para traçar rotas de migração. A maioria dos estudos relativos a este tipo de marcadores, em populações do Sudeste Asiático, incluíram marcadores de ancestralidade materna (mtDNA), sendo mais escassos os estudos relativos a marcadores de ancestralidade paterna (cromossomo Y).

## 2 OBJETIVOS

### 2.1 Objetivo geral

Este projeto teve por objetivo principal aprofundar o conhecimento relativo à composição genética de populações do Sudeste Asiático, através do estudo das principais linhagens de origem paterna presentes na região, e assim contribuir para o entendimento das teorias das rotas de migração.

### 2.2 Objetivos específicos

1. Determinar os haplótipos para marcadores Y-STR, em amostras selecionadas das principais regiões vietnamitas (incluindo indivíduos dos grupos étnicos Kinh, Tay, Thai, Muong, Hoa, Nung, Mong, Dao e Cao Lan), coletadas de várias províncias/municípios do norte ao sul do país e em amostras selecionadas das principais regiões das Filipinas (Luzon, Visayas e Mindanao). Calcular as frequências haplotípicas, comparar os perfis genéticos obtidos e verificar se existe estratificação entre diferentes regiões e/ou grupos étnicos, tanto do Vietnã como das Filipinas.
2. Otimizar novas metodologias através de um conjunto de *Multiplex* de marcadores SNP de cromossomo Y, para classificação dos principais Haplogrupos Asiáticos.
3. Determinar os haplogrupos presentes nas amostras do Vietnã e das Filipinas, por genotipagem dos marcadores Y-SNP, previamente selecionados.
4. Calcular as frequências dos haplogrupos e comparar os perfis genético obtidos, verificando se existe estratificação nas diferentes regiões.
5. Calcular parâmetros de diversidade genética intra e interpopulacional, tanto nas amostras estudadas neste trabalho como em outras

previamente publicadas para marcadores do cromossomo Y no Sudeste Asiático.

6. Analisar os resultados obtidos referentes às linhagens masculinas presentes em populações do Sudeste Asiático considerando às diferentes hipóteses de colonização desta região, com o intuito de elucidar aspectos relacionados à história destas populações.

### **3 MATERIAIS E MÉTODOS**

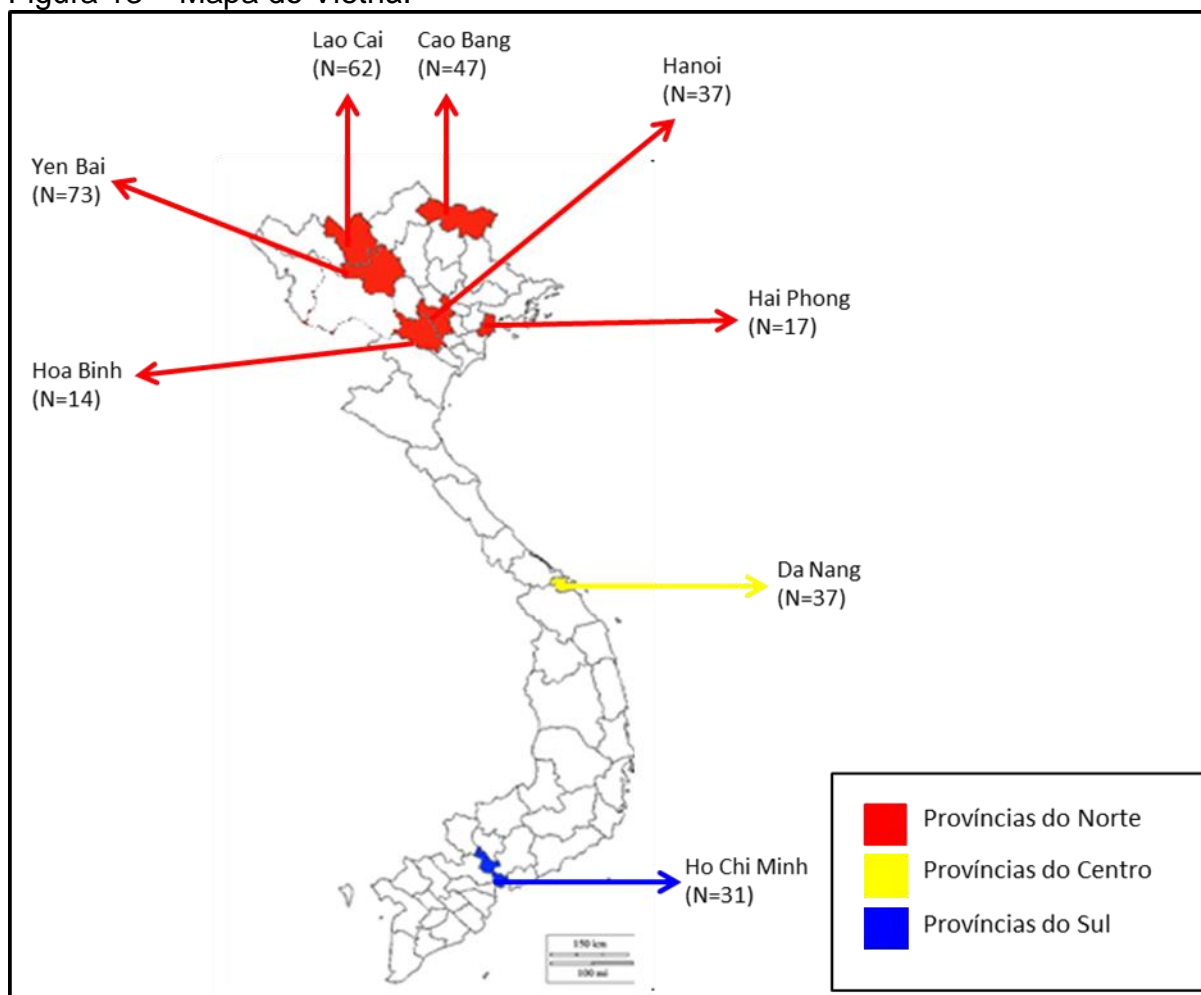
#### **3.1 Amostragem e obtenção das amostras de DNA**

No âmbito dessa tese, foram estudadas duas amostras populacionais: Vietnamitas e Filipinos.

Para o estudo da população do Vietnã, foram selecionadas, de forma aleatória, 318 amostras de pessoas do sexo masculino (não aparentados até a terceira geração) residentes em diferentes regiões vietnamitas (Figura 13). Nomeadamente, foram selecionadas amostras:

- Do Sul, província Ho Chi Minh, compreendendo indivíduos dos grupos étnicos Kihn e Hoa;
- Do Centro, província Da Nang, todos pertencem a etnia Kihn;
- Do Norte, província Hai Phong (todos também da etnia Kihn), Hanói (a maioria Kihn), bem como das áreas rurais em Lao Cai (etnia Mong), Cao Bang (etnias Tay, Nung e Kinh), Yen Bai (etnias Cao Lang, Dao, Mong, Muong, Nung, Thai, Kinh), Hoa Bihn (etnia Muong).

Figura 13 – Mapa do Vietnã.



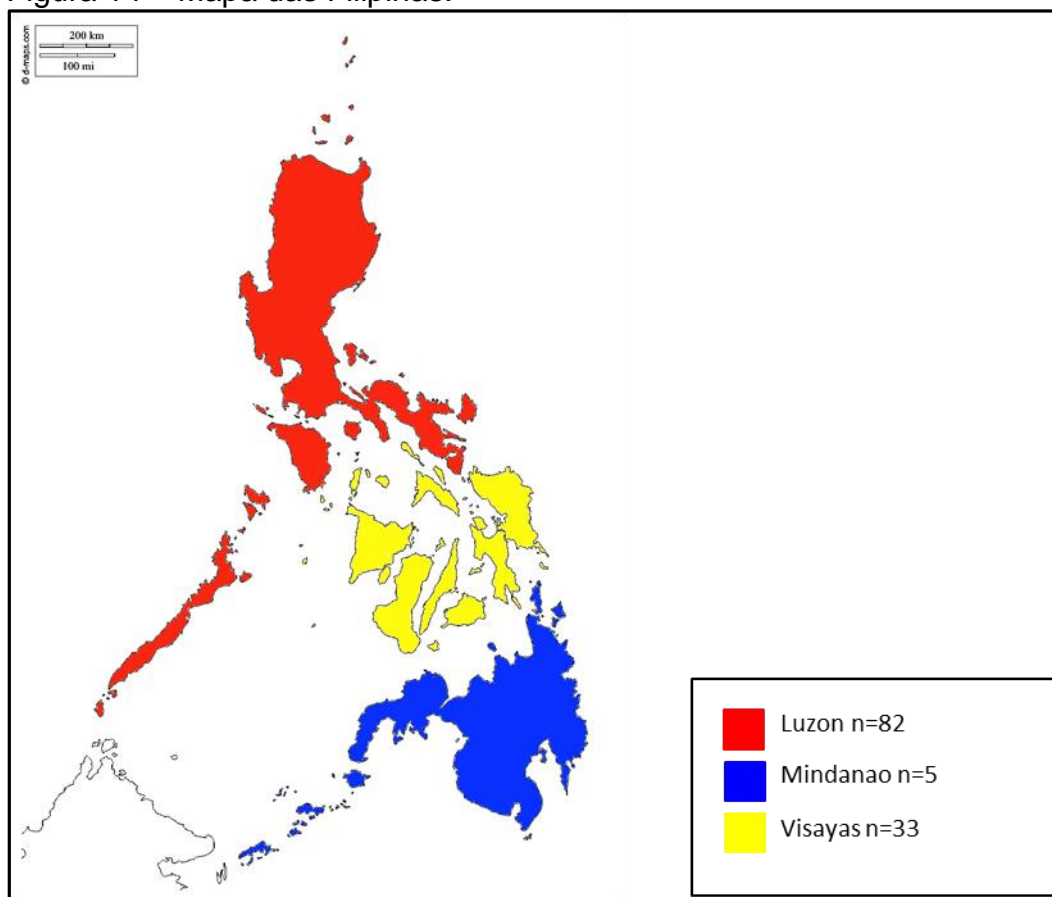
Legenda: Em destaque, as províncias utilizadas neste trabalho bem como seus respectivos números de amostras.

Fonte: A Autora, 2019.

As amostras do Vietnã foram obtidas em colaboração, já estabelecida, com os pesquisadores Carlos Vullo (*Laboratorio de Genética Forense, Equipo Argentino de Antropología Forense\_EAAF*), Nguyen Duc Nhu (Instituto de Medicina Forense do Ministério da Saúde de Hanói, Vietnã) e Tran Duc Phan (Instituto de Medicina Forense do Ministério da Saúde de Hanói, Vietnã).

Para o estudo da população das Filipinas foram selecionadas 120 amostras, de forma aleatória, de pessoas do sexo masculino (não aparentados até a terceira geração), pertencentes a etnia Tagalog, naturais das principais regiões das Filipinas: Luzon (n=82), Mindanao (n=5) e Visayas (n=33) (Figura 14). No âmbito deste estudo, foram recolhidas informações da naturalidade dos pais e avós dos doadores, não existindo, no entanto, informação relativa ao grupo étnico dos mesmos.

Figura 14 – Mapa das Filipinas.



Legenda: Em destaque as três principais ilhas das Filipinas.  
Fonte: A Autora, 2019.

As amostras das Filipinas foram obtidas em colaboração com o pesquisador Dr. Antonio Gonzalez-Martín, do Departamento de Zoologia e Antropologia Física da Faculdade de Ciências Biológicas da Universidade Complutense de Madri, Espanha.

As amostras foram coletadas com consentimento informado e enviadas para o LDD-UERJ de forma codificada, após remoção dos dados pessoais. O presente estudo faz parte de um projeto mais amplo que visa o estudo de outros marcadores genéticos para além dos incluídos neste trabalho, tendo sido aprovado pelos comités de ética das instituições locais.

### 3.2 Extração de DNA

A extração de DNA das amostras das Filipinas foi realizada através do método de extração orgânica (fenol/clorofórmio) pelos colaboradores na Faculdade de Ciências Biológicas da Universidad Complutense em Madri, Espanha.

Nas amostras Vietnamitas, a extração foi feita utilizando-se o método de extração orgânica por fenol/clorofórmio ou o método de Chelex®.

O DNA foi extraído a partir de manchas de sangue em papel de FTA. Para cada mancha, cortou-se um pedaço de 0,5 a 1 cm<sup>2</sup> (de acordo com a intensidade e quantidade de sangue na mancha) e colocou-se em um microtubo de 1,5 µL.

Na extração por fenol/clorofórmio, ao tubo contendo a mancha de sangue, foi adicionado 1 µL de SSC 1x (Saline-Sodium Citrate) e, depois de 15 minutos, o papel foi macerado para que o sangue contido neste fosse liberado. Depois, a amostra foi centrifugada por 3 minutos a 13.000 rpm (12.000 g) e o sobrenadante foi eliminado. Acrescentaram-se 400 µL de Tampão de Extração, 30 µL de proteinase K (10 mg/µL) e 20 µL de DTT (1 M) e incubou-se a 37°C por toda a noite. No dia seguinte, adicionaram-se 200 µL de Fenol/Clorofórmio/Isoamílico (25:24:1) para separação das fases orgânica e aquosa. Agitou-se o tubo em vortex, até se formar uma emulsão leitosa, e centrifugou-se por 3 minutos a 13.000 rpm (12.000 g). O sobrenadante foi transferido para outro tubo, onde se adicionaram 200 µL de Clorofórmio/Isoamílico (24:1) seguido de centrifugação por 3 minutos a 13.000 rpm (12.000 g). Adicionaram-se 30 µL de Acetato de sódio (3 M) e 1000 µL de Etanol 100 % (a frio) e colocou-se o tubo a -20°C por 12 h. Na manhã seguinte, o material no tubo foi centrifugado por 15 minutos a 10.000 rpm (15.000 g) em uma centrífuga refrigerada. Após a centrifugação, o material foi decantado, e foram adicionados 1000 µL de Etanol 70 % (gelado), centrifugando por 3 minutos a 13.000 rpm (12.000 g), e decantado novamente. Por último, adicionaram-se 30 µL da solução tampão TE (Tris-HCL e EDTA). O produto extraído foi armazenado a -20°C, até utilização.

Na extração por Chelex®, foi seguido o método descrito por Lareu e colaboradores (LAREU *et al.*, 1994). Ao tubo contendo a mancha de sangue, foi adicionado 1 µL de água milli-Q. Como essa imersão em água não é suficiente para a retirada do sangue contido no papel, acrescentaram-se 400 µL do Buffer de Extração (citado acima no protocolo do Fenol/clorofórmio) e incubou-se a temperatura ambiente por 30 minutos. Depois a amostra foi centrifugada por 5 minutos a 13.000 rpm (12.000 g). O sobrenadante foi descartado, adicionaram-se 500 µL de Chelex® 100 (100-200 mesh, solução a 5 %) e agitou-se em vortex. Logo após, foram feitas

duas incubações sucessivas: a primeira a 56°C durante 20 minutos e a segunda a 100°C durante 8 minutos. Centrifugou-se a amostra novamente por 5 minutos a 13.000 rpm (12.000 g) e armazenou-se a 4°C para uso imediato ou a -20°C para estoque prolongado.

### 3.3 Genotipagem de marcadores Y-STRs

As amostras extraídas foram amplificadas e genotipadas com dois kits comerciais. O primeiro, AmpFISTR® Y-Filer™ (*ThermoFisher Scientific*, USA), inclui 17 marcadores Y-STR: DYS389I, DYS389II, DYS390, DYS456, DYS19, DYS385 (incluindo os *loci* DY385a e DY385b), DYS458, DYS391, DYS392, DYS393, DYS439, DYS635, DYS437, DYS438, DYS448 e GATAH4. Já o segundo, o Yfiler™ Plus (*ThermoFisher Scientific*), além dos já citados, inclui ainda os marcadores DYS576, DYS627, DYS460, DYS518, DYS570, DYS449, DYS481, DYS533 e DYF387S1. Os experimentos foram realizados de acordo com o indicado no manual dos kits, disponível no site do fabricante.

A PCR foi realizada num volume final de 5 µL contendo 0,16 µL de AmpliTaqGold DNA polymerase [5 U/µL] (*ThermoFisher Scientific*) 1 µL de um *mix* de *primer*, 1,84 µL do Reaction mix (ambos fornecidos com o kit AmpF/ STR Y-Filer (*ThermoFisher Scientific*), 1,5 µL de água e 0,5 µL de DNA (2 ng/µL). Esta reação foi realizada no termociclador Veriti Thermal Cycler® (*ThermoFisher Scientific*) sob as seguintes condições de amplificação: desnaturação inicial a 95°C por 11 minutos, seguida de 28 ciclos de desnaturação a 94°C por 1 minuto, anelamento a 61°C por 1 minuto e extensão a 72°C por 1 minuto, seguida de extensão final a 60°C por 80 minutos.

Após a PCR, as amostras amplificadas foram misturadas com 8,8 µl de formamida (*Hi-Di Formamide*, *ThermoFisher Scientific*) e 0,2 µl de *GeneScan*™ 500 *LIZ*™ *dye Size Standard* (*ThermoFisher Scientific*), e aplicadas em placas de 96 poços. As amostras foram submetidas à eletroforese capilar num aparelho 3500 *Genetic Analyser* (*ThermoFisher Scientific*) e analisadas utilizando-se o *software GeneMapper V4.1* (*ThermoFisher Scientific*). As designações dos alelos foram feitas por comparação com escalas alélicas padronizadas pelo fabricante. Após a

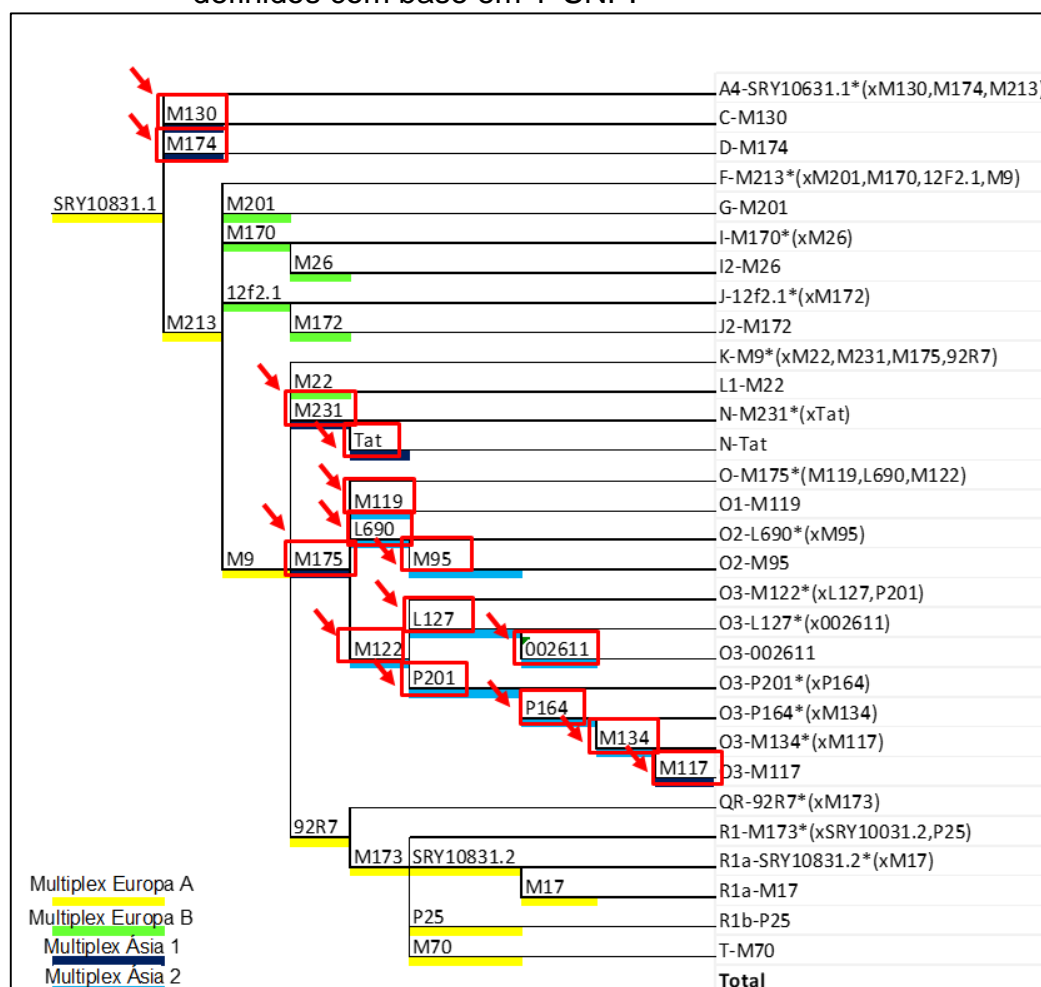


confirmação das genotipagens, os resultados foram exportados para o programa Excel.

### 3.4 Genotipagem de Y-SNPs

Com base em estudos prévios e em dados provenientes da história da colonização do Sudeste Asiático, foram selecionados 15 Y-SNPs (Figura 15) que permitem a classificação das principais linhagens masculinas esperadas nesta região. Os Y-SNPs selecionados definem haplogrupos característicos de populações do Sudeste Asiático, incluindo linhagens pertencentes aos haplogrupos O, D, N e C (KARAFET *et al.*, 2008).

Figura 15 – Árvore filogenética dos haplogrupos do cromossomo Y, definidos com base em Y-SNP.



Legenda: As marcações em vermelho (setas e retângulos) indicam os 15 marcadores Y-SNPs selecionados para o presente trabalho.

Fonte: A Autora, 2019.

A estratégia traçada incluiu dois protocolos de genotipagem em *multiplex* dos 15 Y-SNPs selecionados. O primeiro *multiplex* (*Multiplex Ásia Um*) foi montado incluindo os SNPs mais basais dos haplogrupos selecionados (Quadro 3), bem como dois outros SNPs que permitem discriminar sublinhagens dentro dos haplogrupos N e O.

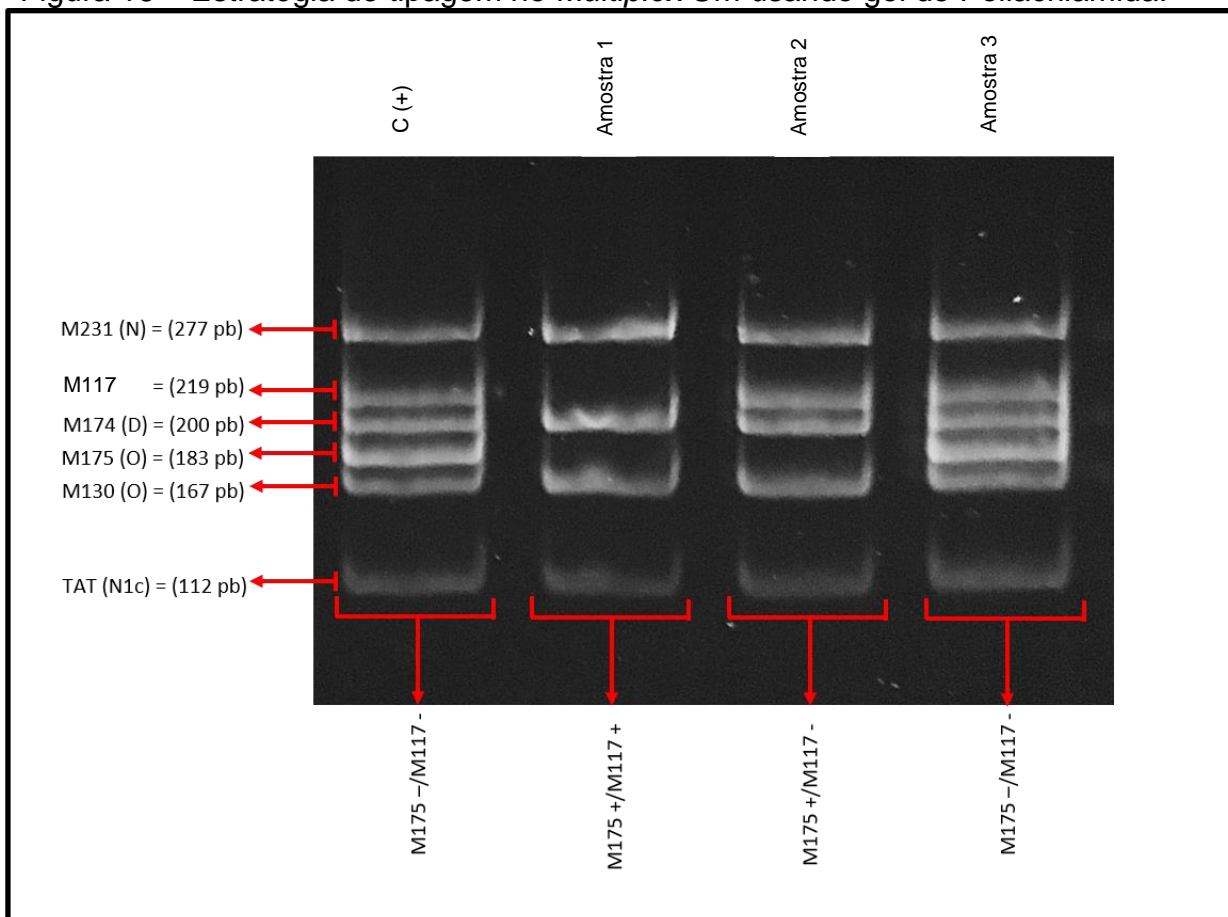
Quadro 3 – *Multiplex Ásia Um*.

<i>Multiplex Um</i>	
SNP	Haplogrupo
M175	O
M174	D
M130	C
M231	N
M117	O3A3C
Tat	N1c

Fonte: A Autora, 2019.

Nesse *multiplex*, os dois marcadores selecionados do haplogrupo O (M175 e M117; um mais basal e outro mais derivado, respectivamente) representam inserções/deleções. Sendo assim, os *primers* foram desenhados complementarmente à sequência deletada, ou seja, quando existe deleção, os *primers* não se anelam e não aparecem as bandas no gel. No controle positivo, C (+), podemos observar todos os marcadores (Figura 16).

Figura 16 – Estratégia de tipagem no *Multiplex* Um usando gel de Poliacrilamida.

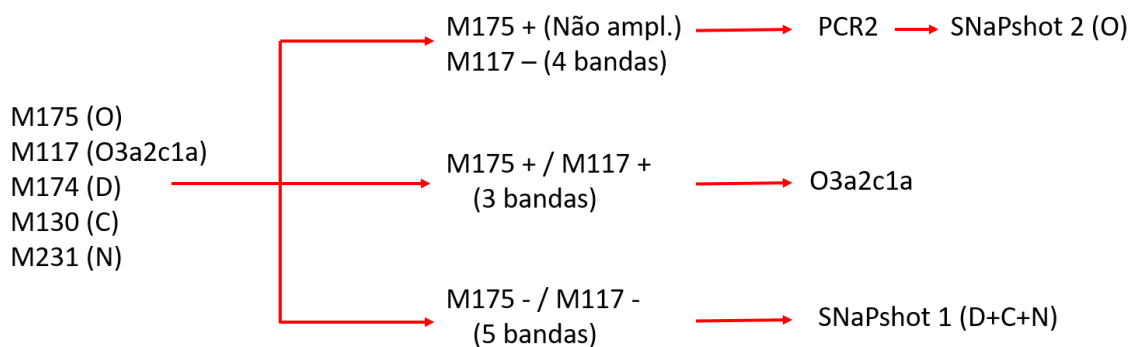


Legenda: A ausência da banda relacionada à posição do M175 indica que a Amostra 2 pertence ao haplogrupo O e a ausência das duas bandas na Amostra 1 demonstra que a mesma além de ser do Haplogrupo O, é uma O3A3C. Na amostra utilizada como controle positivo C (+) e na Amostra 3 aparecem todas as bandas, indicando ausência deleção, logo, estas não pertencem ao Haplogrupo O.

Fonte: A Autora, 2019.

No gel de poliacrilamida a 12%, a ausência da banda relacionada a deleção M175 indicou que essa amostra pertencia ao haplogrupo O, conseqüentemente, foi feito o SNaPshot com os marcadores mais derivados deste haplogrupo (Figura 17). Quando o gel indicou a ausência das bandas relacionadas ao M175 e M117, concluímos que a amostra em questão se tratou de uma O3A3C, pois o segundo marcador corresponde a um SNP mais derivado. E, por último, quando todas as bandas estavam presentes, indicou que essa amostra podia pertencer a qualquer um dos três haplogrupos, exceto O, e foi realizado o SNaPshot com os marcadores relacionados ao C, D e N.

Figura 17 – Descrição da estratégia utilizada na genotipagem dos quinze Y-STRs.

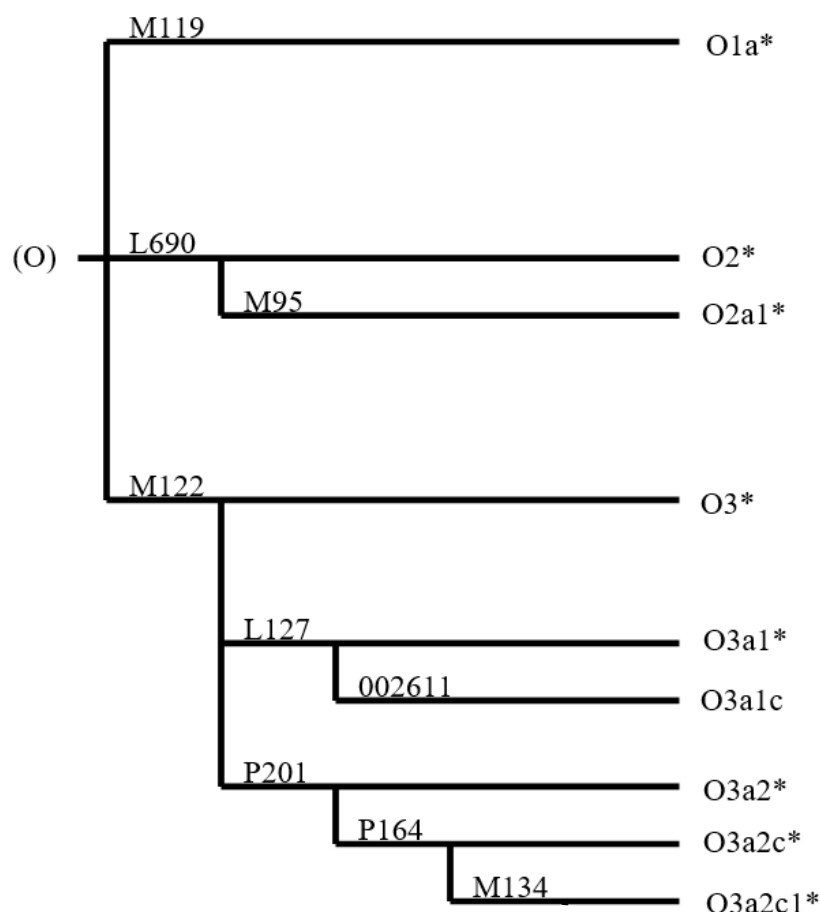


Legenda: Os marcadores M175 e M117 são inserções/deleções. Os *primers* para amplificação destes marcadores foram desenhados sobrepostos à sequência polimórfica. Logo, quando esta sequência está deletada os *primers* não se anelam e, conseqüentemente, não aparecem as bandas correspondentes no gel.

Fonte: A Autora, 2019.

No segundo *multiplex* (*multiplex* Ásia 2) foram selecionados vários marcadores do haplogrupo O, desde os mais basais aos mais derivados, tal como indicado na Figura 18.

Figura 18: Multiplex Ásia Dois.



Legenda: Marcadores do haplogrupo O utilizados neste trabalho.  
 Fonte: A Autora, 2019.

### 3.4.1 Protocolos utilizados com *Multiplex Um e Dois*

#### 3.4.1.1 Seleção das sequências alvo e desenho dos *primers*

As sequências contendo os marcadores selecionados foram retiradas do *site* (<https://www.ncbi.nlm.nih.gov/genbank/>) GeneBank com base nas referências (rs) indicadas nos trabalhos: Karafet e colaboradores (KARAFET *et al.*, 2008) e Yan e colaboradores (YAN *et al.*, 2011). Utilizando a informação contida no GeneBank, foi ainda investigada a possível presença de SNPs vizinhos, uma vez que estes podiam

interferir no anelamento dos *primers* desenhados levando ao aparecimento de alelos nulos.

Para a seleção dos *primers* de PCR e de sequenciamento foi utilizado o programa *Primer3* (<http://primer3.ut.ee/>). Os *primers* selecionados, são indicados nos Quadro 4 e Quadro 5.

Após seleção dos *primers* para construir os *Multiplex* 1 e 2, foi verificada a ausência de dímeros e *hairpin*, com o auxílio do programa *AutoDimer* (<https://strbase.nist.gov//AutoDimerHomepage/AutoDimerProgramHomepage.htm>).

Quadro 4 – Lista dos *primers* de PCR inclusos nos dois *Multiplex*. *Multiplex* Um representado em vermelho e o Dois em azul.

Nome da Seq	Nº Bases	Sequência 5'-3'	Tamanho Fragmento
M175_F	24	gtctgacaaACATGCCTTCTCACTTCTTTCTC	183
M175_R	22	TCTTGCAGCATTTTCAGTTAGC	
M117_F	22	ATTTGGGTAGAAAACTGCAAG	219
M117_R	23	ACCAAAGGAATGCACATCTATCT	
M174_F	20	TCTCCGTCACAGCAAAAATG	200
M174_R	20	GAAGGTCCTGGAGATGCAAA	
M130_F	20	TGAGGCATGTGTGTTTTGGT	167
M130_R	20	TGGCCAGCCTCTTATCTCTC	
M231_F	20	TTGCCTGTGCATGAAAACT	277
M231_R	23	TTTGACACCACAGAAATTACAGG	
002611_F	20	AGCCAACATACTCGCCAATC	257
002611_R	20	GGCTTGCCCTACTGAGAACA	
M119_F	20	TGGGTATTCCAATTCAGCA	95
M119_R	20	CAAACCGCAGTGCTATGTGT	
M95_F	20	CCTTCTTGGGATCAAATGGA	219
M95_R	20	GTTGTGAGGTCCTTCCAGA	
L690/F167_F	20	CTCTCTCATGGGCTGGACAT	207
L690/F167_R	20	TGGCAGATGCAAGCTACCTA	
M122_F	20	GGTATTCAGGCGATGCTGAT	222
M122_R	21	GTGACTGCAAATGGTATGCAA	
L127_F	20	CCCTAAAACAAGGGGTCGAT	259
L127_R	20	GGGGAAGGGATAGCATTAGG	
P201_F	20	TGTGCTGTGCAAGTTGTGTG	152
P201_R	20	AACCCCAAATCCCAAGGTAG	
P164_F+SBE	21	AGCATTTTGGTCCCATCTTTT	80
P164_R	20	CCCTCTTTTCTCCCATTC	
M134_F	21	CCCTTCTTTGGCTTCTCTTTG	260
M134_R	20	TTTCCCCACAACCAGACAAT	

Nota: Foi adicionada uma cauda no *primer* M175\_F para uma melhor separação do fragmento no gel.  
Fonte: A Autora, 2019. Letras minúsculas representam a cauda no *primer*.

Quadro 5 – Lista dos *primers* de extensão (SBE) selecionados. Representado em vermelho os *primers* do *Multiplex Um* e em azul os do *Multiplex Dois*.

Nome da Sequência	Nº Bases	Sequência 5'-3'
M174_SBE	19	CACCCCTCACTTCTGCACT
M130_SBE	21	GGGCAATAAACCTTGGATTC
M231_SBE	27	ACAACATTTACTGTTTCTACTGCTTTC
P201_SBE	29	gtctgacaaGTGAGAGCCAGTTAAAGCCC
M122_SBE	25	gacaaTCAGATTTTCCCCTGAGAGC
M95_SBE	27	GGATAAGGAAAGACTACCATATTAGTG
M134_SBE	32	agtctgacaaAAGAAAAGGCCCAGGAAAGTAT
L690/F167_SBE	32	ctgacaaAGGCAGGTATTCAGAGAAGAAGCAA
M119_SBE	38	gtgaaagtctgacaaTTATCCAATTCAGCATACAGGC
002611_SBE	42	ccacgtcgtgaaagtctgacaaCGAGGCCCTGTGCTTCCAGA
L127_SBE	46	cgctcgtgaaagtctgacaaTCAGAATAATAACAGAGTAATTGGCAG

Fonte: A Autora, 2019.

### 3.4.1.2 *Multiplex Um* e *Dois*

Os marcadores e *primers* inclusos em cada um dos *Multiplex* estão indicados no Quadro 5, sendo separados por cores (Vermelho: *Multiplex Um* e Azul: *Multiplex Dois*). Nos dois casos, as reações de PCR foram feitas num volume final de 5 µL, contendo 2,5 µL da *master mix* 2X incluída no kit QIAGEN Multiplex PCR, 0,5 µL de mistura de *primers* 10X (2 µM de concentração de cada um dos *primers forward* e *reverse*), 1 µL de água miliQ e 1 µL de DNA a 2 µM. As condições termocíclicas utilizadas nas reações de amplificação foram: desnaturação a 95°C por 15 minutos; anelamento com 35 ciclos de 94°C por 30 segundos, 60°C por 1 minuto e 30 segundos, e 72°C por 1 minuto; seguido com uma extensão final de 72°C por 10 minutos.

Neste trabalho foi utilizado o kit SNaPshot (*ThermoFisher Scientific*). O SNaPshot é uma técnica de minisequenciamento que se baseia na extensão do *primer* em apenas um nucleotídeo marcado por fluorescência.

Previamente ao sequenciamento por SNaPshot foi feita a purificação dos produtos de PCR pela adição de 0,5 µL da enzima ExoSAP\_it™ (*ThermoFisher Scientific*) a 1 µL de produto amplificado e incubação a 37°C por 15 minutos, seguida de desnaturação a 85°C por 15 minutos em termociclador.

Após purificação, foi feito o SNaPshot, adicionando 1 µL do SNaPshot mix e 2,5 µL de solução de *primers* de extensão 2X (0,4 µM dos *primers* M130 µM e Tat, 0,3 µM do M231 e 0,2 µM do M174). Utilizando um termociclador, as amostras foram deixadas a 96°C por 10 segundos, 50°C por 5 segundos e 60°C por 30 segundos, por 25 ciclos.

Após a reação de SNaPshot, foi feita a purificação adicionando 1 µL de SAP (*shrimp alkaline phosphatase*) ao produto final, incubando a 37°C por 1 hora e 85°C por 15 minutos.

Para a separação e detecção, a 1µL do produto amplificado foi adicionada 8,8 µL de formamida Hi-Di (*ThermoFisher Scientific*) e 0,2 µL do padrão interno *GeneScan™ LIZ500™* (*ThermoFisher Scientific*), os produtos foram submetidos ao termociclador 3500 *Genetic Analyser*, com posterior análise utilizando o software *GeneMapper v4.1*.

#### 3.4.1.3 Análise estatística dos resultados

As frequências dos haplogrupos foram estimadas por contagem direta. As frequências haplotípicas e de diversidade foram estimadas de acordo com Saltou e colaborador (SAITOU; NEI, 1987), usando o *software* ARLEQUIN (EXCOFFIER; LISCHER, 2010).

Os cálculos de distâncias genéticas entre pares de populações ( $F_{ST}$  e  $R_{ST}$ ) e AMOVA (Análise de Variância Molecular) foram efetuados também através do programa Arlequin (EXCOFFIER; LISCHER, 2010). Para a análise do escalonamento multidimensional (MDS) da distância genética ( $F_{ST}$ ) entre as populações foi utilizado o programa *SPSS Statistics v.17.0* (*SPSS Incorporation, USA*).

Ao se comparar várias populações pode ocorrer um erro nas análises em relação ao nível de significância. Assim, foi feito a correção de Bonferroni para múltiplos testes, com o intuito de ajustar o valor da significância e, com isso, diminuir a chance de erros.



## 4 RESULTADOS E DISCUSSÃO

Como mencionado anteriormente, esse estudo teve como foco principal aprofundar o conhecimento relativo à composição genética de populações do Sudeste Asiático através do estudo dos haplótipos e principais haplogrupos da região, e assim contribuir para o entendimento das teorias das rotas de migração. Para que esse trabalho pudesse ser concretizado, utilizamos amostras coletadas de duas populações do Sudeste Asiático, Vietnã e Filipinas, bem como marcadores moleculares localizados no cromossomo Y dos tipos STR e SNP.

### 4.1 Composição genética do Vietnã

Os resultados obtidos da genotipagem de Y-STRs (n=305) e Y-SNPs (n=314) das amostras do Vietnã, e as informações sobre o local da amostragem e grupo étnico, encontram-se descritos no Apêndice A.

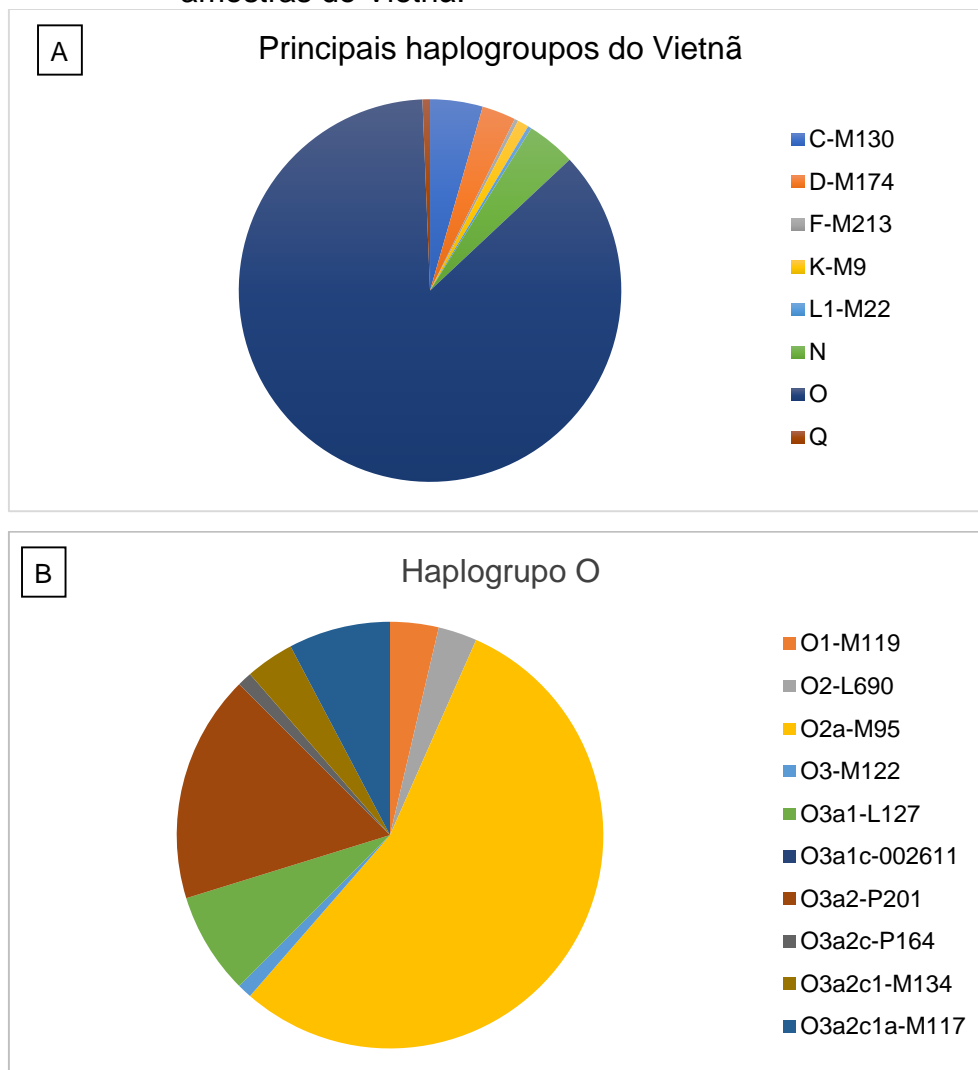
Dados preliminares para Y-STRs foram publicados (Anexo 1) e estão disponíveis na base de dados *online* YHRD (Y-Chromosome STR Haplotype Reference Database; <http://www.yhrd.org>), sob os seguintes números de acesso: Vietname [Kinh] YA003984, Vietname [Mong] YA003985, Vietname [Tay] YA003986, Vietname [Nung] YA003987, Vietname [Hoa] YA003988.

Em relação à composição de haplogrupos, com base na tipagem dos Y-SNPs, o haplogrupo O foi o mais frequente. Este foi encontrado em aproximadamente 88,6 % das amostras do Vietnã, o que está de acordo com dados dos haplogrupos do Sudeste Asiático apresentados na literatura (DELFIN *et al.*, 2011; TREJAUT *et al.*, 2014). Outros haplogrupos também foram encontrados nas nossas amostras, como o haplogrupo C encontrado em 14 % (segundo mais frequente) das amostras, o N em 13 % e também o D em 2,6 % (Figura 19).

Dentro do haplogrupo O, o sub-haplogrupo O2 é o mais frequente na Indochina, que inclui o Vietnã, Laos e Camboja (TREJAUT *et al.*, 2014) estando representado na maioria das nossas amostras (54,8 %) pelo sub-haplogrupo O2-M95 (Figura 19) e

existindo ainda outros oito sub-haplogrupos de menor frequência, dentro do haplogrupo O.

Figura 19 – Representação gráfica dos haplogrupos encontrados nas amostras do Vietnã.



Legenda: A) Principais haplogrupos no Vietnã. B) Sub-haplogrupos do Vietnã.  
Fonte: A Autora, 2019

#### 4.1.1 Análise da subestrutura genética das províncias

Com a finalidade de compreender como se distribui a diversidade genética ao longo do território do Vietnã, foram efetuadas comparações entre amostras de diferentes províncias agrupadas por regiões geográficas (Norte, Centro e Sul). Para

isso, foi feita uma Análise de Variância Molecular (AMOVA) com dados de Y-SNPs ( $F_{ST}$ ) e Y-STRs ( $F_{ST}$  e  $R_{ST}$ ) para essas diferentes províncias.

Os resultados mostraram que há uma diferença maior entre as populações (províncias) dentro das regiões que entre as regiões (Tabela 1). Isso condiz com a análise da frequência dos marcadores Y-SNPs, pois Lao Cai (província do Norte) mostrou uma frequência elevada do sub-haplogrupo O3\*-P201, em contraste com as províncias restantes (mesmo as outras do Norte), em que o sub-haplogrupo O2\*-M95 é o mais frequente. Portanto, há uma diferença maior entre duas províncias do Norte que uma do Norte e outra do Sul/Centro, o que indica que o critério geográfico não seria ideal para uma caracterização genética do Vietnã.

Tabela 1 – Resultado do teste estatístico AMOVA para diferentes províncias vietnamitas.

	Porcentagem de variação					
	Y-SNPs ( $F_{ST}$ )	Valor de P	Y-STRs ( $F_{ST}$ )	Valor de P	Y-STRs ( $R_{ST}$ )	Valor de P
Entre regiões	-4,41	0,534	-0,39	0,929	-3,76	0,750
Entre populações dentro das regiões	13,22	<0,000005	0,76	<0,000005	7,65	<0,000005
Dentro das populações	91,19	<0,000005	99,63	<0,000005	96,11	<0,000005

Nota: As análises foram realizadas com dados de Y-SNPs ( $F_{ST}$ ) e Y-STRs ( $F_{ST}$  e  $R_{ST}$ ) mostrando que há uma diferença maior entre populações dentro dos grupos (segunda linha) que entre os grupos (primeira linha). s.e.  $\leq 0,00494$ .

Fonte: A Autora, 2019.

Para melhor interpretação dos resultados, foram feitas análises de distâncias genéticas par a par tanto para Y-STRs quanto para Y-SNPs (Tabela 2). Após a aplicação de correção de Bonferroni para múltiplos testes (nível de significância de 0,00179), a província de Yen Bai apresentou uma distância genética significativa em relação à Cao Bang, nas análises de Y-STRs usando o  $F_{ST}$ . Porém, essa distância genética não foi significativa para análises de Y-STRs com  $R_{ST}$  e de Y-SNPs, o que sugere uma necessidade de aumentar o número de amostras de Yen Bai.

Tabela 2 – Resultado das análises de distância genética par a par para Y-SNPs e Y-STRs das províncias

Y-SNPs - Fsts	● Ho Chi Minh	● Da Nang	● Cao Bang	● Yen Bai	● Ha Noi	● Hai Phong	● Lao Cai	● Hoa Binh
● Ho Chi Minh	*	0,002	0,003	0,039	0,250	0,159	<b>0,00006</b>	0,004
● Da Nang	0,096	*	0,853	0,049	0,061	0,667	<b>&lt;0,000005</b>	0,273
● Cao Bang	0,080	-0,013	*	0,130	0,055	0,965	<b>&lt;0,000005</b>	0,261
● Yen Bai	0,031	0,027	0,012	*	0,292	0,957	<b>&lt;0,000005</b>	0,023
● Ha Noi	0,007	0,033	0,027	0,003	*	0,816	<b>&lt;0,000005</b>	0,025
● Hai Phong	0,021	-0,016	-0,028	-0,025	-0,020	*	<b>0,00001</b>	0,114
● Lao Cai	<b>0,109</b>	<b>0,297</b>	<b>0,274</b>	<b>0,168</b>	<b>0,157</b>	<b>0,198</b>	*	<b>&lt;0,000005</b>
● Hoa Binh	0,153	0,004	0,008	0,077	0,088	0,036	<b>0,368</b>	*
Y-STRs - Fsts								
● Ho Chi Minh	*	1,000	0,086	0,038	1,000	1,000	0,013	1,000
● Da Nang	-0,001	*	0,102	0,013	1,000	1,000	0,003	1,000
● Cao Bang	0,003	0,003	*	<b>0,00059</b>	0,080	0,304	<b>&lt;0,000005</b>	0,283
● Yen Bai	0,005	0,006	0,009	*	0,013	0,076	<b>&lt;0,000005</b>	0,148
● Ha Noi	0,000	-0,001	0,003	0,006	*	1,000	0,009	1,000
● Hai Phong	0,000	0,000	0,002	0,006	-0,002	*	0,067	1,000
● Lao Cai	0,008	0,008	0,011	0,013	0,007	0,008	*	0,124
● Hoa Binh	0,000	0,000	0,003	0,006	0,000	0,000	0,008	*
Y-STRs - Rsts								
● Ho Chi Minh	*	0,120	0,166	0,169	0,797	0,931	<b>&lt;0,000005</b>	0,005
● Da Nang	0,017	*	0,259	0,030	0,534	0,960	<b>&lt;0,000005</b>	0,026
● Cao Bang	0,011	0,005	*	0,152	0,144	0,781	<b>&lt;0,000005</b>	0,003
● Yen Bai	0,010	0,025	0,009	*	0,181	0,715	<b>&lt;0,000005</b>	0,005
● Ha Noi	-0,011	-0,003	0,011	0,009	*	0,997	<b>&lt;0,000005</b>	0,004
● Hai Phong	-0,026	-0,028	-0,015	-0,013	-0,035	*	0,003	0,089
● Lao Cai	0,084	<b>0,154</b>	<b>0,178</b>	<b>0,101</b>	0,087	0,089	*	<b>0,00003</b>
● Hoa Binh	0,095	0,066	0,100	0,094	0,095	0,049	<b>0,226</b>	*

Nota: Acima da diagonal (\*) estão os valores de probabilidade de não diferenciação e abaixo os valores de  $F_{ST}$  e  $R_{ST}$ . Com um nível de significância a 0,00179, as amostras que apresentaram uma distância genética significativa encontram-se destacadas em vermelho e, em laranja, estão as amostras que apresentaram um alto valor de distância genética. A província localizada ao Sul (Ho Chi Minh) está representada por uma marcação em azul, a província central (Da Nang) em amarelo e as demais localizadas ao Norte foram representadas com a marcação em vermelho.  $s.e \leq 0,0015$ .

Fonte: A Autora, 2019.

Lao Cai apresentou uma distância genética elevada, na ordem de 10 a 37 % (Tabela 2), em relação às outras províncias, sendo essa distância significativa nas análises para Y-STRs, com  $R_{ST}$  (exceto Hai Phong). Já nas análises de  $F_{ST}$ , algumas dessas amostras não tiveram significância. Isso se deve ao fato dos Y-SNPs, por possuírem menor taxa de mutação, serem mais sensíveis a diferenças populacionais que os Y-STRs. Esse resultado corrobora os resultados das frequências dos haplogrupos, logo existe uma diferença genética entre Lao Cai e as outras províncias.

As amostras da província de Lao Cai deste trabalho pertencem exclusivamente à etnia Mong. Existe uma relação étnica e topográfica nessa província, nas quais as pessoas do grupo étnico Kinh residem em locais de baixa altitude e os Mongs em altas altitudes, em montanhas isoladas e pouco povoadas (ROCHE; MICHAUD, 2000). Desse modo, esse isolamento geográfico dos Mongs contribuiu para a distância genética elevada entre eles e as outras etnias estudadas. Ao correlacionar os marcadores com as etnias, o sub-haplogrupo O2-M95 se manteve como o mais frequente em todas as etnias, excetuando a etnia Mong, onde o sub-haplogrupo mais frequente foi o O3-P201, o que confirma mais uma vez a diferença entre os Mong e as outras etnias.

As amostras do presente estudo também foram comparadas com outras na literatura (PURPS *et al.*, 2014; TREJAUT *et al.*, 2014) e não foi observado uma distância genética significativa entre elas.

#### 4.1.2 Análise da diversidade das províncias

Como discutido anteriormente, Lao Cai apresentou uma distância genética significativa em relação a todas as outras províncias e isso pode ser explicado, devido ao seu isolamento geográfico por se tratar de uma região montanhosa. Todavia, esse isolamento também pode indicar a ocorrência de deriva genética. Para o esclarecimento, foi feita uma análise da diversidade haplotípica das províncias vietnamitas (Tabela 3).

Tabela 3 – Diversidade haplotípica.

	Diversidade Haplotípica
Ho Chi Minh	1,0000 ± 0,0082
Da Nang	1,0000 ± 0,0063
Cao Bang	0,9935 ± 0,0072
Yen Bai	0,9884 ± 0,0072
Ha Noi	1,0000 ± 0,0063
Hai Phong	1,0000 ± 0,0202
Lao Cai	0,9852 ± 0,0059
Hoa Binh	1,0000 ± 0,0340

Nota: Valores de diversidade haplotípica estimados para os oito grupos populacionais estudados.

Fonte: A Autora, 2019.

Devido aos altos valores de diversidade haplotípica encontrados foi descartada a hipótese de deriva genética na maioria dos grupos estudados. Contudo, a diversidade haplotípica mostrou-se um pouco mais baixa nas amostras de Lao Cai e Yen Bai. O número de haplótipos compartilhados nas populações dessas províncias também foi alto, sendo 35,5 % em Yen Bai, 45,1 % em Lao Cai contra apenas 14 % em Cao Bang e zero nas demais.

Os valores baixos da diversidade dessas duas províncias e o compartilhamento dos haplótipos sugerem um isolamento dessa população, pois estes valores podem estar relacionados com as suas respectivas etnias, Mong e Dao; sendo que a segunda está associada com uma migração vinda da China e um posterior isolamento geográfico e, assim como os Mong, vivem em altas altitudes nas regiões montanhosas (DANG, 2010). Esses resultados sugerem mais uma vez, que o critério geográfico não seria o ideal para caracterização da população do Vietnã. Portanto, as próximas análises foram baseadas nos grupos étnicos.

#### 4.1.3 Análise da subestrutura genética por etnias

O Vietnã possui 54 comunidades étnicas (*Ethnic groups in Viet Nam : an analysis of key indicators from the 2009 Viet Nam population and housing census / UNFPA, 2011*) e diversas classificações para estas, porém, a classificação por agrupamento linguístico mostra-se extremamente usada e citada na literatura. Em

vista disso, foi feito uma análise para conferir, através das distâncias genéticas, se as etnias poderiam ser agrupadas geneticamente de acordo com a classificação dos grupos linguísticos (Quadro 6).

Quadro 6 – Classificação das etnias, segundo a literatura.

Línguas	Grupos Étnicos desse trabalho
Viet_Muong	Kinh e Muong
Tay e Thai	Nung, Cao Lan, Tay e Thai
Mong-Dao	Dao, Mong
Chinese	Hoa

Fonte: A Autora, 2019.

A etnia Mong apresentou uma distância genética elevada (na ordem de 10-37 %), a um nível de significância de 0,00139 (após a aplicação de correção de Bonferroni para múltiplos testes), com todas as etnias nas análises de  $F_{ST}$ , exceto com Thai (Tabela 4). Já para  $R_{ST}$  e SNPs, o grupo étnico Mong apresentou uma distância significativa entre Hoa, porém pode ser um erro, devido ao fato desta etnia ter uma amostra pequena. Baseado na classificação linguística e os resultados de distância genética, Kinh e Muong poderiam ser agrupados com Thai, Tay, Nung e Cao Lan por não apresentarem diferenças significativas entre eles. Já Dao deveria ser separado de Mong. Portanto, alguns grupos separados por classificação linguística poderiam ser unidos geneticamente e outros que foram classificados como mesmo grupo, seriam separados. Logo, o critério linguístico não seria uma boa classificação para se trabalhar com análises genéticas. (Tabela 4).

Tabela 4 – Resultados das análises de distância genética das etnias do Vietnã

Y-SNPs	Cao Lan	Dao	Hoa	kinh	Mong	Muong	Nung	Tay	Thai
Cao Lan	*	0,040	0,456	0,028	<0,000005	0,005	0,034	0,027	0,285
Dao	0,054	*	0,242	0,600	<0,000005	0,137	0,438	0,507	0,880
Hoa	-0,005	0,034	*	0,395	0,003	0,125	0,285	0,087	0,377
kinh	0,035	-0,007	-0,005	*	<0,000005	0,060	0,238	0,163	0,841
Mong	0,127	0,210	0,169	0,195	*	<0,000005	<0,000005	<0,000005	<0,000005
Muong	0,145	0,030	0,130	0,039	0,351	*	1,000	0,353	0,134
Nung	0,107	-0,003	0,053	0,014	0,318	-0,062	*	0,731	0,279
Tay	0,070	-0,008	0,090	0,012	0,282	-0,002	-0,024	*	0,452
Thai	0,015	-0,042	0,013	-0,025	0,160	0,060	0,033	-0,005	*
Y-STRs - FSTs									
Cao Lan	*	0,007	0,155	0,008	<0,000005	0,070	0,119	0,030	0,097
Dao	0,057	*	0,074	0,000	<0,000005	0,007	0,089	0,009	0,011
Hoa	0,032	0,059	*	0,343	0,001	0,145	0,271	0,205	0,020
kinh	0,022	0,055	0,005	*	<0,000005	0,166	0,520	0,178	0,003
Mong	0,078	0,146	0,141	0,109	*	<0,000005	<0,000005	<0,000005	0,080
Muong	0,031	0,078	0,032	0,010	0,171	*	0,664	0,359	0,003
Nung	0,032	0,043	0,019	-0,003	0,188	-0,013	*	0,828	0,004
Tay	0,039	0,053	0,022	0,006	0,172	0,002	-0,023	*	0,006
Thai	0,050	0,134	0,126	0,100	0,048	0,175	0,188	0,157	*
Y-STRs - RSTs									
Cao Lan	*	0,371	0,179	0,070	0,000	0,002	0,525	0,189	0,109
Dao	0,002	*	0,090	0,008	<0,000005	0,011	0,543	0,226	0,108
Hoa	0,037	0,074	*	0,034	0,004	0,001	0,214	0,098	0,016
kinh	0,018	0,040	0,073	*	<0,000005	0,013	0,611	0,045	0,082
Mong	0,093	0,148	0,177	0,113	*	<0,000005	<0,000005	<0,000005	0,135
Muong	0,120	0,104	0,261	0,051	0,211	*	0,05558	0,003	0,004
Nung	-0,010	-0,016	0,041	-0,011	0,178	0,073	*	0,926	0,008
Tay	0,016	0,013	0,058	0,023	0,199	0,115	-0,041	*	0,024
Thai	0,070	0,075	0,199	0,059	0,049	0,240	0,193	0,121	*

Nota: Acima da diagonal (\*) estão os valores de probabilidade de não diferenciação e abaixo os valores de  $F_{ST}$  e  $R_{ST}$ . Com um nível de significância a 0,00139, as amostras que apresentaram uma distância genética significativa encontram-se destacadas em vermelho e em laranja as amostras que apresentaram um alto valor de distância genética.  $s.e \leq 0,0056$ .

Fonte: A Autora, 2019.

## 4.2 Composição genética das Filipinas

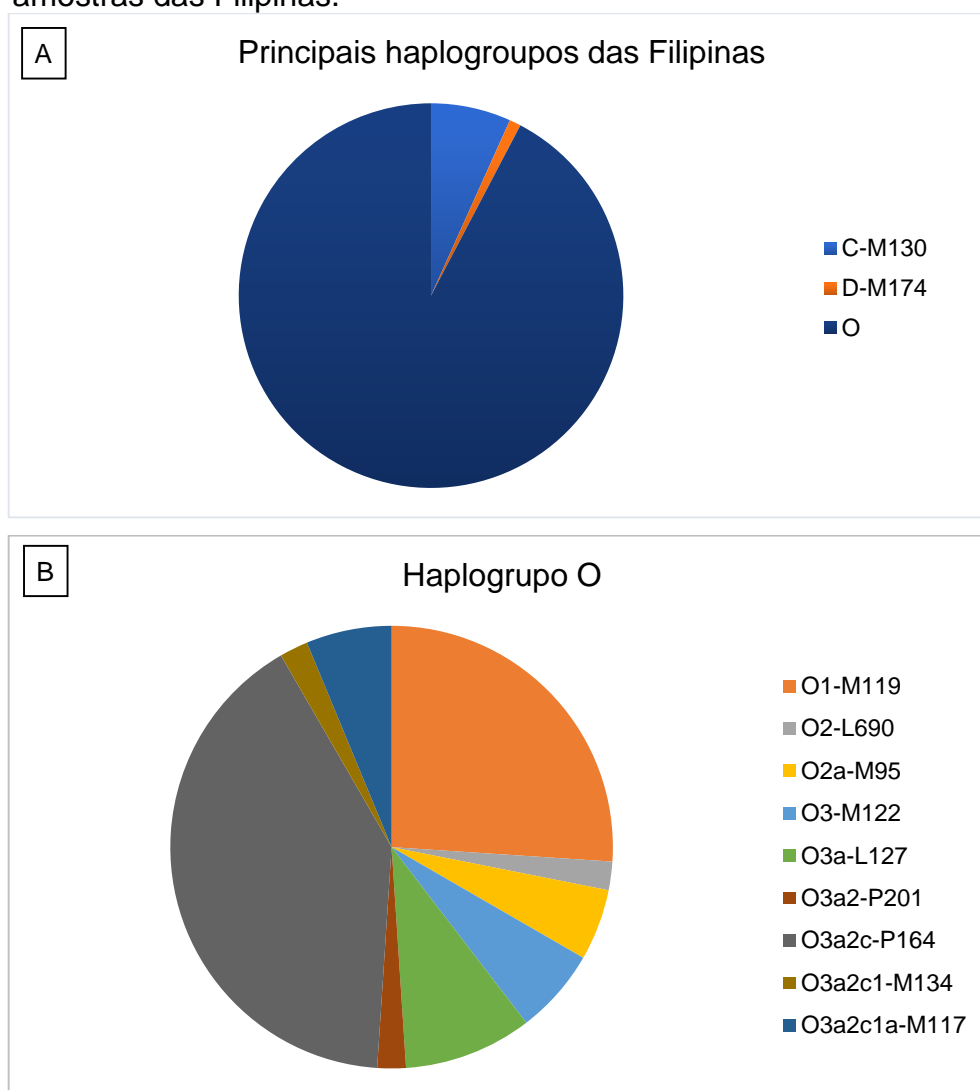
Os resultados obtidos da genotipagem de Y-STRs (n=119) e Y-SNPs (n=104) das amostras das Filipinas encontram-se descritos em Apêndice B, bem como as informações sobre o local da amostragem. Os dados publicados podem ser encontrados no Anexo 2.

Em relação à composição de haplogrupos, o haplogrupo O, no Vietnã e como na maioria de outros países do Sudeste Asiático (DELFIN *et al.*, 2011; TREJAUT *et*



*al.*, 2014) foi o mais frequente estando presente em 93,2 % das amostras das Filipinas (Figura 20). Outro dado observado nas análises foi a presença, nas amostras de Luzon, do haplogrupo C em cinco dos seis cromossomos C-M130. Este haplogrupo não foi observado nas 103 amostras anteriormente estudadas por Trejaut e colaboradores (TREJAUT *et al.*, 2014), apesar de ter sido observado antes por Delfin e colaboradores (DELFIN *et al.*, 2011) em ambos os grupos, Negrito e Não Negrito, com frequências de 0,089 e 0,071, respectivamente.

Figura 20 – Representação gráfica dos haplogrupos encontrados nas amostras das Filipinas.



Legenda: A) Principais haplogrupos das Filipinas. B) Sub-haplogrupos das Filipinas.  
Fonte: A Autora, 2019.

#### 4.2.1 Análise da subestrutura genética

Para compreender como se distribui a diversidade genética entre os três grupos de ilhas das Filipinas, as amostras dos mesmos foram comparadas entre si, através da AMOVA, com dados de Y-SNPs ( $F_{ST}$ ) e Y-STRs ( $F_{ST}$  e  $R_{ST}$ ), para esses diferentes grupos (Tabela 5). Os resultados mostraram valores altos de diversidade dentro dos grupos, indicando uma grande diversidade genética na população. Já os valores de diferenciação entre as populações (grupos das Filipinas) foram muito baixos, mostrando uma distância genética não significativa entre eles. Para confirmar essa possibilidade, foi feito a análise par a par entre as amostras.

Tabela 5 – Resultado do teste estatístico AMOVA entre os grupos das Filipinas

	Porcentagem de variação					
	Y-SNPs ( $F_{ST}$ )	Valor de $P$	Y-STRs ( $F_{ST}$ )	Valor de $P$	Y-STRs ( $R_{ST}$ )	Valor de $P$
Entre populações	-1,12	0,681	0.19	0,317	0.78	0,216
Dentro das populações	101,12	—	99.81	—	99.22	—

Nota: As análises foram realizadas com dados de Y-SNPs ( $F_{ST}$ ) e Y-STRs ( $F_{ST}$  e  $R_{ST}$ ). s.e.  $\leq 0,0048$   
 Fonte: A Autora, 2019.

As análises de distâncias genéticas  $F_{ST}$  e  $R_{ST}$  par a par foram feitas tanto para Y-STRs quanto para Y-SNPs (Tabela 6). O resultado mostrou em todas as análises que não há uma distância genética significativa ( $p < 0,01666$ ) para os grupos de ilhas das Filipinas nas amostras desse trabalho, corroborando o resultado anterior. Com isso, é possível unificá-las para as análises de comparação com outras amostras das Filipinas e até mesmo com outros países do Sudeste Asiático

Tabela 6 – Resultado das análises de distância genética do grupo de ilhas das Filipinas.

Y-SNPs	Luzon	Mindanao	Visayas
Luzon	*	0,649	0,600
Mindanao	-0,039	*	0,387
Visayas	-0,007	-0,001	*
Y-STRs - FSTs			
Luzon	*	0,292	0,402
Mindanao	0,009	*	0,308
Visayas	0,001	0,008	*
Y-STRs - RSTs			
Luzon	*	0,119	0,419
Mindanao	0,052	*	0,300
Visayas	0,000	0,018	*

Nota: Análises de par a par para Y-SNP ( $F_{ST}$ ) e Y-STR ( $F_{ST}$  e  $R_{ST}$ ).  
s.e $\leq$ 0,0049

Fonte: A Autora, 2019.

Para uma análise da subestrutura genética dos haplogrupos, unimos as amostras de Mindanao e Visayas, pois não possuem uma distância genética significativa, além do número de amostras de Mindanao ser muito pequeno podendo induzir a um erro.

Dentro do haplogrupo O, o sub-haplogrupo dominante nas amostras de Luzon (34,6 %) e Mindanao/Visayas (29,9 %) foi o O3a2c\*\_P164. Estudos realizados em cromossomo Y, mostraram que o sub-haplogrupo dominante na Polinésia é o C2\_208, porém existe uma frequência alta, cerca de 53 %, do sub-haplogrupo O3a2c\*\_P164 na Ilha Polinésia, Tonga (OXENHAM; BUCKLEY, 2015). Isso pode ser explicado por pesquisas que já demonstraram que uma população indígena do norte das Filipinas é a mais próxima das primeiras populações da Oceania Remota (incluindo Tonga). A segunda população mais próxima seria a dos indígenas de Taiwan. Isso condiz com o modelo das rotas de migração *Out of Taiwan* (SKOGLUND *et al.*, 2016).

O segundo sub-haplogrupo mais frequente nas amostras foi o O1a\*-M119, com uma frequência de 21 % em Luzon e 28,9 % em Mindanao/Visayas. Já nas amostras das Filipinas estudadas por Trejaut e colaboradores (TREJAUT *et al.*, 2014) este foi o sub-haplogrupo mais frequente. Neste trabalho foi encontrado uma frequência mais baixa deste sub-haplogrupo em Luzon que a descrita por Trejaut e colaboradores (TREJAUT *et al.*, 2014).

O sub-haplogrupo O1a\*-M119 apresenta frequências altas em todas as populações do Sudeste Asiático (com exceção da Indochina) e tem sido muito citado em estudos da expansão Austronésia, antes da ingressão na Polinésia, pois nesta região a frequência desse haplogrupo tem se mostrado baixa (MIRABAL *et al.*, 2012).

#### 4.2.2 Comparação com outras amostras das Filipinas na literatura

##### 4.2.2.1 Comparação com 21 Y-STRs

Primeiramente, os dados haplotípicos deste trabalho foram comparados com outros das Filipinas já existentes na literatura (PURPS *et al.*, 2014) para o 21 Y-STRs inclusos em ambos os kits 23-Plex e YFiler Plus. A comparação par a par das distâncias genéticas  $F_{ST}$  e  $R_{ST}$  não revelaram diferenças significativas entre quaisquer pares de amostras estudadas (Tabela 7 e Tabela 8).

Tabela 7 – Comparação par a par das distâncias genéticas  $F_{ST}$  entre amostras populacionais das Filipinas.

	Luzon (a)	Filipinas (a)	Luzon (b)	Mindanao + Visayas (b)
Luzon (a)	*	0,36822	0,01182	0,86583
Filipinas (a)	0,00017	*	0,0117	0,74635
Luzon (b)	0,00463	0,00612	*	0,86794
Mindanao + Visayas (b)	-0,00327	-0,00261	-0,00462	*

Legenda: Comparação par a par das distâncias genéticas  $F_{ST}$  entre amostras populacionais das Filipinas (abaixo da diagonal “\*\*”) e o valor de  $p$  correspondente a não diferenciação (acima da diagonal (“\*\*”). (a) Purps e colaboradores (PURPS *et al.*, 2014) (b) Presente estudo; Número de permutações: 50175; s.e.  $\leq$  0.0021.

Fonte: A Autora, 2019.

Tabela 8 – Comparação par a par das distâncias genéticas  $R_{ST}$  entre amostras populacionais das Filipinas.

	Luzon (a)	Filipinas (a)	Luzon (b)	Mindanao + Visayas (b)
Luzon (a)	*	0,11167	0,0595	0,87635
Filipinas (a)	0,00215	*	0,24394	0,73428
Luzon (b)	0,00586	0,00235	*	0,30938
Mindanao + Visayas (b)	-0,0067	-0,00507	0,00261	*

Legenda: Comparação par a par das distâncias genéticas  $R_{ST}$  entre amostras populacionais das Filipinas (abaixo da diagonal) e o valor de  $p$  correspondente a não diferenciação (acima da diagonal). (a) Purps e colaboradores (PURPS *et al.*, 2014); (b) Presente estudo; Número de permutações: 50175; s.e.  $\leq$  0.0049.

Fonte: A Autora, 2019.

#### 4.2.2.2 Comparação com 17 Y-STRs

As amostras desse estudo foram comparadas com dados de outras amostras populacionais das Filipinas que foram previamente publicados por Purps e colaboradores (PURPS *et al.*, 2014) e Trejaut e colaboradores (TREJAUT *et al.*, 2014), para um conjunto comum de 17 *loci* Y-STR(DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 e GATA H4).

Na análise de distâncias genéticas  $F_{ST}$ , os valores de probabilidade de não-diferenciação foram estatisticamente significantes apenas entre as amostras de Luzon estudadas por Trejaut e colaboradores (TREJAUT *et al.*, 2014) e as amostras restantes das Filipinas. No entanto, estas diferenças não foram significativas com base nas distâncias genéticas  $R_{ST}$ . Foram detectadas distâncias genéticas  $F_{ST}$  e  $R_{ST}$  pequenas, e estatisticamente não significativas, entre os pares de amostras restantes (Tabela 9 e Tabela 10).

Tabela 9 – Comparação par a par de distâncias genéticas  $F_{ST}$  entre amostras populacionais das Filipinas.

	Luzon (a)	Filipinas (a)	Luzon (b)	Mindanao + Visayas (b)	Batan (c)	Luzon (c)	Mindanao + Visayas (c)
Luzon (a)	*	0,61486	0,08466	0,86354	0,21749	0,00074	0,6419
Filipinas (a)	-0,00039	*	0,12349	0,9084	0,27712	0,00291	0,58779
Luzon (b)	0,00257	0,00275	*	0,91474	0,02117	0,00018	0,11426
Mindanao + Visayas (b)	-0,00366	-0,00500	-0,00602	*	0,11344	0,03193	0,46411
Batan (c)	0,00496	0,00364	0,02153	0,01221	*	0,65452	0,35385
Luzon (c)	0,01031	0,01113	0,02195	0,01227	-0,00489	*	0,03585
Mindanao + Visayas (c)	-0,00156	-0,00138	0,00595	-0,00029	0,00243	0,01163	*

Legenda: Distâncias genéticas  $F_{ST}$  entre amostras populacionais das Filipinas (abaixo da diagonal) e o valor correspondente de probabilidade de não diferenciação (acima da diagonal). (a) Purps e colaboradores (PURPS *et al.*, 2014); (b) Presente estudo; (c) Trejaut e colaboradores (TREJAUT *et al.*, 2014). Número de permutações: 50175; s.e. ≤ 0.0023

Fonte: A Autora, 2019.

Tabela 10 – Comparação par a par das distâncias genéticas  $R_{ST}$  entre amostras populacionais das Filipinas.

	Luzon (a)	Filipinas (a)	Luzon (b)	Mindanao + Visayas (b)	Batan (c)	Luzon (c)	Mindanao + Visayas (c)
Luzon (a)	*	0,22087	0,25146	0,64716	0,25582	0,01505	0,27027
Filipinas (a)	0,0011	*	0,16642	0,61024	0,68706	0,13642	0,70419
Luzon (b)	0,0017	0,00394	*	0,53232	0,13999	0,00762	0,24532
Mindanao + Visayas (b)	-0,00344	-0,00330	-0,00222	*	0,20523	0,08375	0,3659
Batan (c)	0,00605	-0,00816	0,0155	0,0135	*	0,62459	0,81626
Luzon (c)	0,01235	0,00547	0,02297	0,01607	-0,00883	*	0,72963
Mindanao + Visayas (c)	0,00278	-0,00458	0,00486	0,00148	-0,01632	-0,00719	*

Legenda: Comparação par a par das distâncias genéticas  $R_{ST}$  entre amostras populacionais das Filipinas (abaixo da diagonal) e o valor de  $p$  correspondente a não diferenciação (acima da diagonal) (a) Purps e colaboradores (PURPS *et al.*, 2014); (b) Presente estudo; (c) Trejaut e colaboradores (TREJAUT *et al.*, 2014).; Número de permutações: 50175; s.e.  $\leq 0.0053$ .

Fonte: A Autora, 2019.

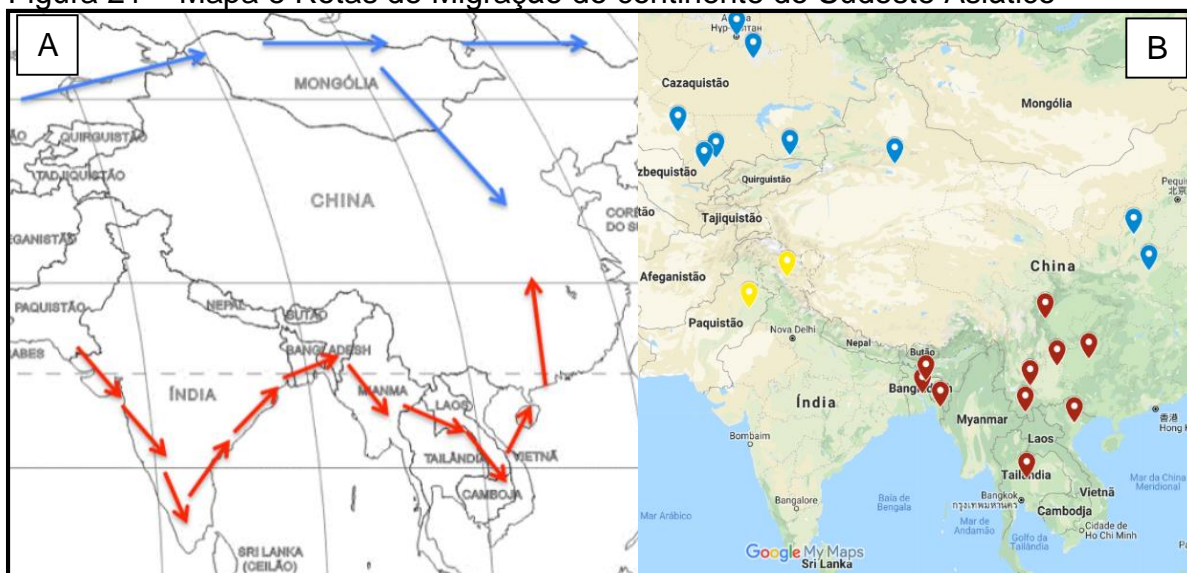
#### 4.2.3 Comparação entre as populações do Sudeste da Ásia.

Na comparação de populações do Sudeste Asiático, foi feita uma seleção dos países/regiões do Sudeste Asiático que se localizam próximos as rotas de migração do continente e ilhas do Sudeste Asiático, discutidas anteriormente. Os dados de Y-STR dessas localidades foram retirados da literatura (Apêndice C) e comparados com as amostras do presente trabalho. Foi feita uma primeira análise de diferenciação populacional entre todas as amostras selecionadas (para os 17 Y-STRs incluídos em todos os estudos), e as amostras que não apresentaram uma distância genética significativa foram unificadas.

##### 4.2.3.1 Análise das rotas de migração continentais

Primeiramente, foi feita uma análise envolvendo as populações selecionadas ao longo das rotas de migração continentais (Figura 21), com base em distâncias genéticas ( $F_{ST}$ ) calculadas entre pares de populações (Apêndice D).

Figura 21 – Mapa e Rotas de Migração do continente do Sudeste Asiático



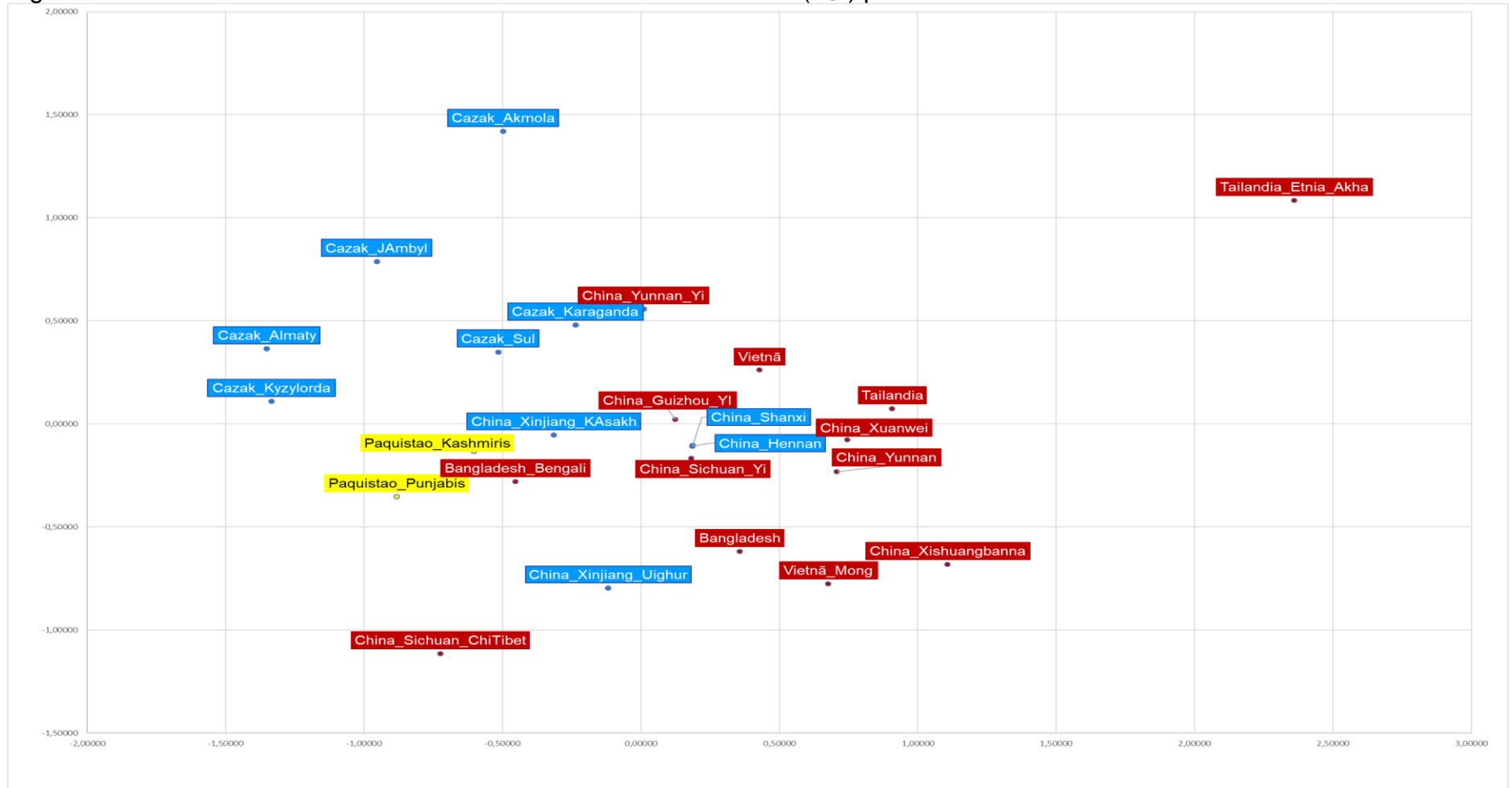
Legenda: A) Hipóteses das rotas de migração no continente do Sudeste Asiático. B) Mapa mostrando as localizações das amostras analisadas. Amostras escolhidas na rota norte (azul), na rota sul (vermelho) e ponto de intermédia essas rotas (amarelo).

Fonte: A Autora, 2019.

Foi construído um gráfico MDS representando as distâncias genéticas,  $F_{ST}$  (Figura 22), no qual se observa aglomerados entre as amostras das populações da rota do norte e da rota do sul.

O Cazaquistão localiza-se no início da rota do norte, enquanto o Yunnan (China) próximo à rota do sul. No entanto, essas amostras estão lado a lado no MDS, indicando uma aproximação genética entre elas. A hipótese de apenas uma rota sugere uma dispersão para o interior do continente. Todavia, se essa hipótese fosse condizente, observaríamos uma aproximação maior entre Yunnan, Vietnã e Tailândia que entre Yunnan e as amostras de Cazaquistão. Já o Paquistão se localiza entre as duas rotas e no gráfico encontra-se muito próximo a Bangladesh (Bengali), na rota Sul e ao grupo Kasak de Xinjiang (China), na rota Norte, indicando que pode ter ocorrido uma bifurcação de uma rota próxima ao Paquistão, originando uma rota ao norte e outra ao sul, no Sudeste Asiático. Portanto, esses resultados estão de acordo com a teoria da existência de duas rotas de migração no Sudeste Asiático.

Figura 22 – Escalonamento Multidimensional das análises de Y-STR ( $F_{ST}$ ) para as amostras do continente do Sudeste Asiático.



Legenda: Amostras em azul representam a rota do norte e amostras em vermelho a rota do sul. As amostras em amarelo destacam o país que fica entre essas rotas.

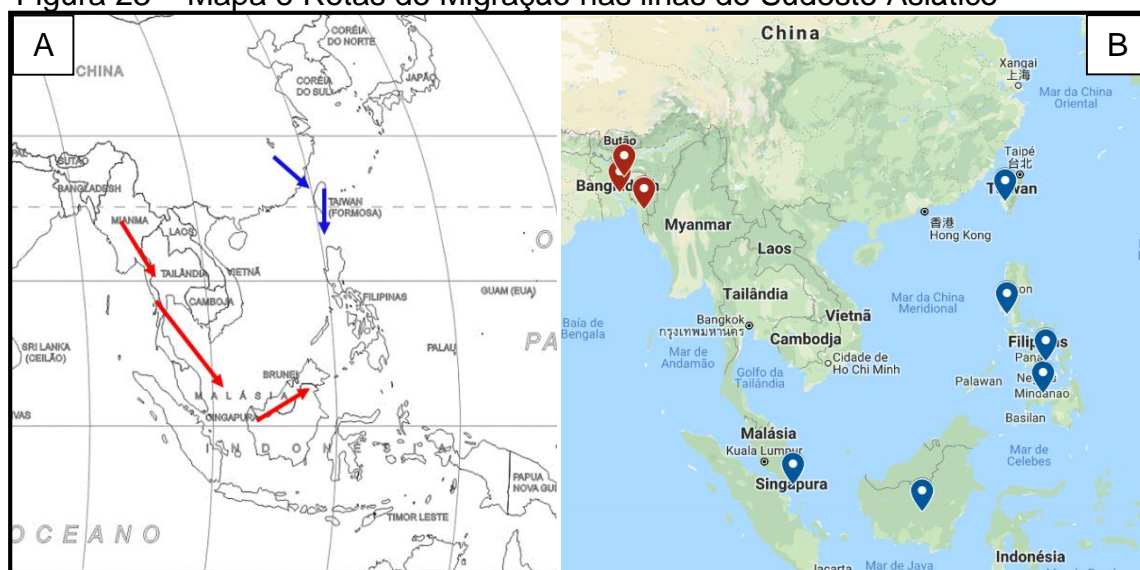
Fonte: A Autora, 2019.



#### 4.2.3.2 Escalonamento multidimensional entre as amostras populacionais das ilhas do Sudeste Asiático.

A próxima análise foi realizada com amostras de Y-STR tanto das populações das ilhas do Sudeste Asiático, quanto de populações continentais que poderiam estar envolvidas nas rotas das ilhas (Figura 23). Para as populações selecionadas, foram calculadas as distâncias genéticas ( $F_{ST}$ ) par a par (APÊNDICE E).

Figura 23 – Mapa e Rotas de Migração nas ilhas do Sudeste Asiático

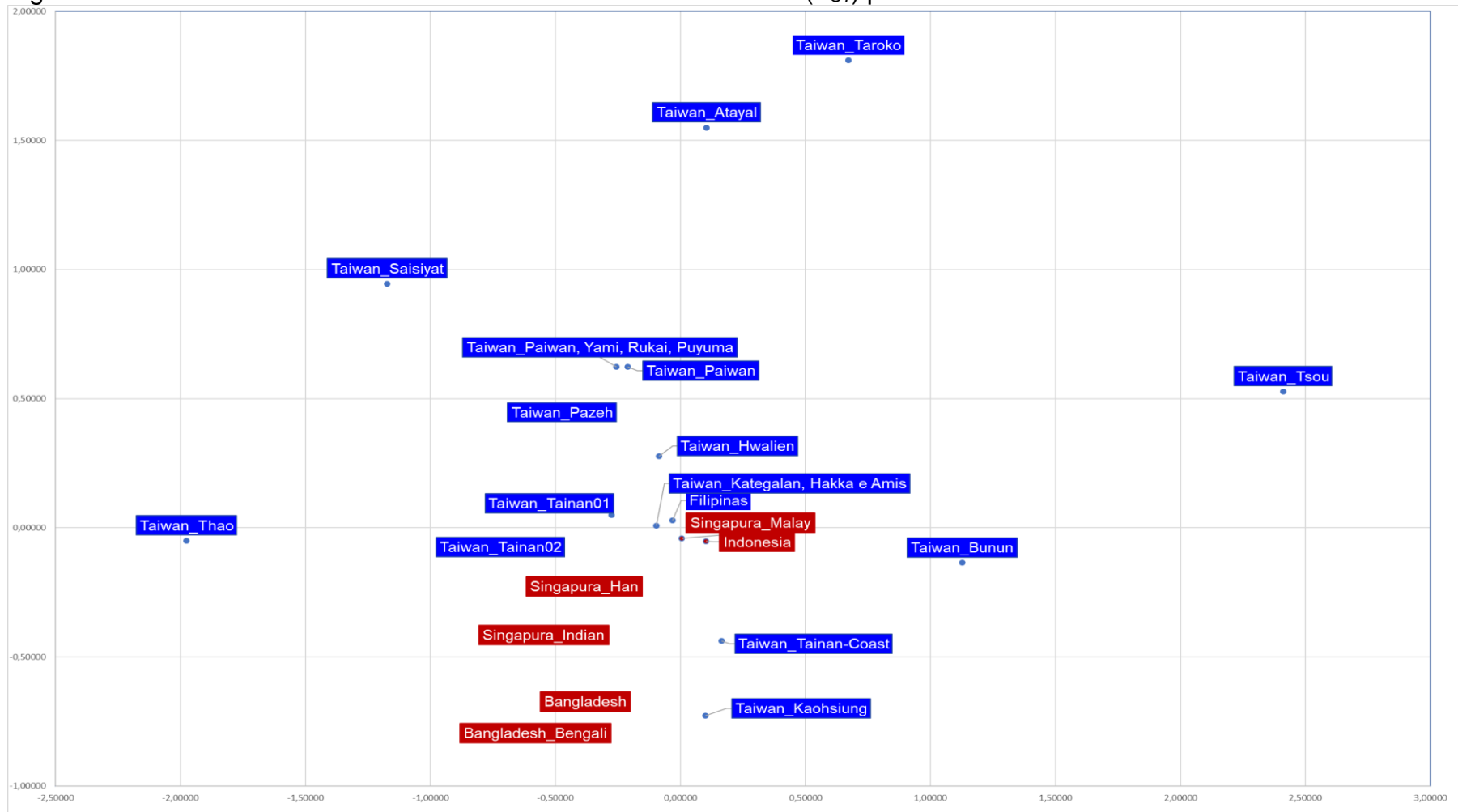


Legenda: A) Hipóteses das rotas de migração nas ilhas do Sudeste Asiático. B) Mapa mostrando as localizações das amostras analisadas. Azul escuro representando as amostras escolhidas nas ilhas, e vermelho representando as amostras no continente.

Fonte: A Autora, 2019.

A representação das distâncias genéticas,  $F_{ST}$  (Figura 24), por gráfico MDS, mostra aglomerados entre as amostras das populações da ilha que podem ser interpretadas em sequências tanto como uma rota que sairia por Taiwan, passando por Filipinas, Singapura (Malay) e Indonésia; quanto outra que viria de Bangladesh, passaria por Singapura (Han e Indian) em direção a Taiwan. A interpretação desse gráfico sugere a existência de duas rotas simultâneas que migraram em sentidos opostos pelas ilhas.

Figura 24 – Escalonamento multidimensional das análises de Y-STR ( $F_{ST}$ ) para as amostras das ilhas do Sudeste Asiático.



Legenda: Amostras em azul representam a rota do norte e amostras em vermelho a rota do sul.

Fonte: A Autora, 2019.

## CONCLUSÕES

O presente estudo permitiu alcançar as seguintes conclusões:

1. Com base nos haplótipos (Y-STR) encontrados e suas frequências, pode-se concluir que o Vietnã se trata de um país com baixa heterogeneidade genética entre as províncias. Sendo assim, o critério geográfico não seria ideal para uma caracterização do país. No entanto, ao utilizar critérios de etnicidade, verificou-se uma diferenciação genética, indicando ser esse o melhor parâmetro para agrupamento das subpopulações do Vietnã. Nas Filipinas, existe uma baixa distância genética entre os três principais grupos de ilhas (Luzon, Mindanao e Visayas), não se observando uma estratificação populacional.
2. Os sistemas *Multiplex* de marcadores Y-SNP mostraram-se eficazes na determinação dos principais haplogrupos asiáticos. Esta técnica utilizada baseou-se numa simples separação em gel de poliacrilamida, sendo isso uma vantagem pois se trata de uma técnica com baixo custo.
3. O Vietnã e as Filipinas apresentaram um perfil tipicamente do Sudeste Asiático pois o haplogrupo O foi o mais frequente em ambos. Nessas duas populações também foram detectados os haplogrupos C e D. No entanto, nas amostras Vietnamitas encontrou-se uma maior variedade de haplogrupos estando também presentes os F, K, L1, N.
4. Com base nos haplogrupos encontrados nas amostras vietnamitas, detectou-se a presença de uma possível estratificação por etnia neste país. O sub-haplogrupo O2\_M95 mostrou-se o mais frequente em todas as etnias, exceto na Mong onde o O3-P201 foi o mais frequente. Já nas amostras Filipinas, observou-se uma ausência de estratificação, pois o mesmo sub-haplogrupo (O3a2c\*\_P164) foi o mais frequente em todos os grupos de ilhas.
5. A diversidade genética intrapopulacional no Vietnã é maior tanto à nível dos haplótipos quanto dos haplogrupos em relação as Filipinas. Na diversidade interpopulacional também foi verificado uma menor diferenciação entre as populações das Filipinas que entre as Vietnamitas. As amostras do Vietnã deste trabalho foram comparadas com outras vietnamitas da literatura e não apresentaram diferença

significativa entre elas. Os resultados das análises feitas com amostras das Filipinas (em que se utilizou 17 marcadores) também não apresentou diferença significativa com outras populações das Filipinas da literatura.

6. Ao correlacionar a distância genética das populações do presente trabalho com a distância genética de outras do Sudeste Asiático na literatura, corroborou-se tanto com a hipótese de duas rotas de migração nos continentes sendo uma ao norte e outra ao sul, como também, a hipótese de duas rotas simultâneas que migram em sentido contrário nas ilhas do Pacífico.

## REFERÊNCIAS

- ADNAN, A. et al. Population data of 17 Y-STRs (Yfiler) from Punjabis and Kashmiris of Pakistan. *Int J Legal Med*, v.132, n. 1, p. 137-138, Jan 2018.
- AloTrip International Limited. *Origin of Vietnamese people – AloTrip*. Hanoi, Vietnam: Copyright ©. Disponível em: < <https://www.alotrip.com/about-vietnam-people/origin-vietnamese-people>>. Acesso em: 20 set. 2019
- AMORIM, A.; FERNANDES, T.; TAVEIRA, N. *Mitochondrial DNA in human identification: a review*. PeerJ, 7, p. e7314, 2019.
- BAI, R. et al. Analysis of 27 Y-chromosomal STR haplotypes in a Han population of Henan province, Central China. *Int J Legal Med*, 130, n. 5, p. 1191-1194, Sep 2016.
- BAI, R. et al. *Haplotype diversity of 17 Y-STR loci in a Chinese Han population sample from Shanxi Province, Northern China*. Forensic Sci Int Genet, 7, n. 1, p. 214-216, Jan 2013.
- BANISTER, J.; CENSUS, U. S. B. O. T. *The Population of Vietnam*. U.S. Department of Commerce, Bureau of the Census, 1985.
- BARRIOS, J. *Tagalog for Beginners: An Introduction to Filipino, the National Language of the Philippines (Downloadable MP3 Audio Included)*. Tuttle Publishing, 2014. 9781462910397.
- BAYE, T. M. Inter-chromosomal variation in the pattern of human population genetic structure. *Human Genomics*, 5, n. 4, p. 220, 2011/05/01 2011.
- BELEZA, S. et al. Extending STR markers in Y chromosome haplotypes. *Int J Legal Med*, 117, n. 1, p. 27-33, Feb 2003.
- BELLWOOD, P.; FOX, J. J.; TRYON, D. *The Austronesians: Historical and Comparative Perspectives*. ANU E Press, 2006. 9781920942854.
- BELLWOOD, P.; NESS, I. *The Global Prehistory of Human Migration*. Wiley, 2014. 9781118970591.
- BUCKLETON, J. S.; KRAWCZAK, M.; WEIR, B. *The Interpretation of Lineage Markers in Forensic DNA Testing*. Forensic science international. Genetics, 5, p. 78-83, 03/01 2011.
- BUGOYE, F. C.; MULIMA, E.; MISINZO, G. *Analysis of Mutation Rate of 17 Y-Chromosome Short Tandem Repeats Loci Using Tanzanian Father-Son Paired Samples*. Genet Res Int, 2018, p. 8090469, 2018.
- BUTLER, J. M. *Advanced Topics in Forensic DNA Typing: Interpretation*. Elsevier Science, 2014. 9780124058545.

- CALAFELL, F.; LARMUSEAU, M. H. D. *The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research*. Hum Genet, 136, n. 5, p. 559-573, May 2017.
- CAMPBELL, M. C.; TISHKOFF, S. A. *African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping*. Annu Rev Genomics Hum Genet, 9, p. 403-433, 2008.
- CAVALLI-SFORZA, L. L.; FELDMAN, M. W. *The application of molecular genetic approaches to the study of human evolution*. Nat Genet, 33 Suppl, p. 266-275, Mar 2003.
- CENSUS, T. V. P. A. H. *The 2009 Vietnam population and housing Census*. 2009. Hanoi, Vietnam. Disponível em: <<http://www.gso.gov.vn>>. Acesso em: 10 jan. 2016.
- COMAS, D.; LUISELLI, D.; RICKARDS, O. *Human population genetics of the Mediterranean*. Annals of Human Biology, 45, n. 1, p. 1-4, 2018/01/02 2018.
- CONSORTIUM, T. H. P.-A. S. *Mapping Human Genetic Diversity in Asia*. Science, New Series,, 326, n. 5959, p. 4, 2009.
- CONSORTIUM, Y. C. *A nomenclature system for the tree of human Y-chromosomal binary haplogroups*. Genome Res, 12, n. 2, p. 339-348, Feb 2002.
- DANG, A. D. *Strategy and Action Plan for the Greater Mekong Subregion North-South Economic Corridor*. Asian Development Bank, 2010. 9789292547738.
- DANG, H.-A. H. *Vietnam: A Widening Poverty Gap for Ethnic Minorities Indigenous Peoples*, Poverty and Development, 2012.
- DELFIN, F. et al. *The Y-chromosome landscape of the Philippines: extensive heterogeneity and varying genetic affinities of Negrito and non-Negrito groups*. Eur J Hum Genet, 19, n. 2, p. 224-230, Feb 2011.
- DETROIT, F. et al. *"Small size" in the Philippine human fossil record: is it meaningful for a better understanding of the evolutionary history of the negritos?* Hum Biol, 85, n. 1-3, p. 45-65, Feb-Jun 2013.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the World*. Twenty-second edition. Dallas, Texas: SIL International. Disponível em: <<http://www.ethnologue.com>>. Acesso em: 15 ago. de 2019.
- ECKERT, K. A.; HILE, S. E. *Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome*. Mol Carcinog, 48, n. 4, p. 379-388, Apr 2009.

*Ethnic groups in Viet Nam : an analysis of key indicators from the 2009 Viet Nam population and housing census / UNFPA.* Ha Noi, Viet Nam: UNFPA, 2011. v. Accessed from <https://nla.gov.au/nla.cat-vn5917110>).

EXCOFFIER, L.; LISCHER, H. E. *Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.* Mol Ecol Resour, 10, n. 3, p. 564-567, May 2010.

FAN, G. Y. et al. *Forensic and phylogenetic analyses among three Yi populations in Southwest China with 27 Y chromosomal STR loci.* Int J Legal Med, 133, n. 3, p. 795-797, May 2019.

FORSTER, P. M., S. *Did Early Humans Go North or South?* Science, New Series,, 308, n. 5724, p. 1, 2005.

GeneBank. Disponível em: <<https://www.ncbi.nlm.nih.gov/genbank/>>. Acesso em: 25 jan. 2015.

GUSMÃO, L.; BRIÓN, M.; GOMES, I. *Chapter 30 The human Y chromosome male-specific polymorphisms and forensic genetics.* Handbook of Analytical Separations, 6, 01/01 2007.

GYMREK, M. et al. *Interpreting short tandem repeat variations in humans using mutational constraint.* Nat Genet, 49, n. 10, p. 1495-1501, Oct 2017.

HASAN, M. et al. *Phylogenetic analysis and forensic evaluation among Rakhine, Marma, Hajong, and Manipuri tribes from four culturally defined regions of Bangladesh using 17 Y-chromosomal STRs.* Int J Legal Med, 132, n. 6, p. 1641-1644, Nov 2018.

Instituto Médico Howard Hughes e Instituto Nacional da Saúde. *Primer3.* Versão 4.1.0. Disponível em: <<http://primer3.ut.ee/>>. Acesso em: 12 mar. 2015.

Instituto Nacional de Padrões e Tecnologia. *AutoDimer.* Maryland, USA. Disponível em: <<https://strbase.nist.gov//AutoDimerHomepage/AutoDimerProgramHomepage.htm>>. Acesso em: 12 mar. 2015.

JINAM, T. A. et al. *Evolutionary history of continental southeast Asians: "early train" hypothesis based on genetic analysis of mitochondrial and autosomal DNA data.* Mol Biol Evol, 29, n. 11, p. 3513-3527, Nov 2012.

JOBLING, M.; HURLES, M.; TYLER-SMITH, C. *Human Evolutionary Genetics: Origins, Peoples & Disease.* Garland Science, 2013. 9781136844140.

JOBLING, M. A. *Y-chromosomal SNP haplotype diversity in forensic analysis.* Forensic Sci Int, 118, n. 2-3, p. 158-162, May 15 2001.

JOBLING, M. A.; TYLER-SMITH, C. *The human Y chromosome: an evolutionary marker comes of age.* Nat Rev Genet, 4, n. 8, p. 598-612, Aug 2003.

JONSSON, H. *Does the house hold? History and the shape of Mien (Yao) society.* Ethnohistory, 48, n. 4, p. 613-654, 2001.

JORDE, L. B. et al. *The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data.* Am J Hum Genet, 66, n. 3, p. 979-988, Mar 2000.

KARAFET, T. M. et al. *New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree.* Genome Res, 18, n. 5, p. 830-838, May 2008.

KARKI, R. et al. *Defining "mutation" and "polymorphism" in the era of personal genomics.* BMC Med Genomics, 8, p. 37, Jul 15 2015.

KAYSER, M. et al. *Melanesian origin of Polynesian Y chromosomes.* Curr Biol, 10, n. 20, p. 1237-1246, Oct 19 2000.

KEYES, C. *Presidential Address: "The Peoples of Asia"-Science and Politics in the Classification of Ethnic Groups in Thailand, China, and Vietnam.* The Journal of Asian Studies, 61, n. 4, p. 1163-1203, 2002.

KLAMER, M. *The dispersal of Austronesian languages in Island South East Asia: Current findings and debates.* Wiley, p. 26, 2019.

KLEIN, H. L. et al. *Guidelines for DNA recombination and repair studies: Cellular assays of DNA repair pathways.* Microb Cell, 6, n. 1, p. 1-64, Jan 7 2019.

LANDER, E. S. et al. *Initial sequencing and analysis of the human genome.* Nature, 409, n. 6822, p. 860-921, Feb 15 2001.

LAPPALAINEN, T. et al. *Genomic Analysis in the Age of Human Genome Sequencing.* Cell, 177, n. 1, p. 70-84, Mar 21 2019.

LAREU, M. V. et al. *Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations.* Forensic Sci Int, 66, n. 1, p. 41-52, May 25 1994.

LESADA, J. D. *TAGLISH IN METRO MANILA: AN ANALYSIS OF TAGALOG-ENGLISH CODE-SWITCHING.* University of Michigan Department of Linguistics, p. 88, 2017.

LESSIG, R. et al. *Asian online Y-STR Haplotype Reference Database.* Leg Med (Tokyo), 5 Suppl 1, p. S160-163, Mar 2003.

LI, R. *Forensic Biology.* CRC Press, 2015. 9781439889725.

LIU, H. et al. *Chromatin structure is distinct between coding and non-coding single nucleotide polymorphisms.* BMC Mol Biol, 15, p. 22, Oct 5 2014.



LIU, Y. *Genetic Diversity and Disease Susceptibility*. IntechOpen, 2018. 9781789842012.

LOPEZ, S.; VAN DORP, L.; HELLENTHAL, G. *Human Dispersal Out of Africa: A Lasting Debate*. *Evol Bioinform Online*, 11, n. Suppl 2, p. 57-68, 2015.

MATSUMURA, H. et al. *Craniometrics Reveal "Two Layers" of Prehistoric Human Dispersal in Eastern Eurasia*. *Sci Rep*, 9, n. 1, p. 1451, Feb 5 2019.

MELLARS, P. *Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia*. *Science*, 313, n. 5788, p. 796-800, Aug 11 2006.

MICHAUD, J.; SWAIN, M. B.; BARKATAKI-RUSCHEWEYH, M. *Historical Dictionary of the Peoples of the Southeast Asian Massif*. Rowman & Littlefield Publishers, 2016. 9781442272798.

MILLS, R. E. et al. *An initial map of insertion and deletion (INDEL) variation in the human genome*. *Genome Res*, 16, n. 9, p. 1182-1190, Sep 2006.

MIRABAL, S. et al. *Increased Y-chromosome resolution of haplogroup O suggests genetic ties between the Ami aborigines of Taiwan and the Polynesian Islands of Samoa and Tonga*. *Gene*, 492, n. 2, p. 339-348, Jan 25 2012.

MORTON, N. E. *Parameters of the human genome*. *Proc Natl Acad Sci U S A*, 88, n. 17, p. 7474-7476, Sep 1 1991.

NORMILE, D. *SNP Study Supports Southern Migration Route to Asia*. *Science, New Series*, 326, p. 1, 2009.

OXENHAM, M.; BUCKLEY, H. *The Routledge Handbook of Bioarchaeology in Southeast Asia and the Pacific Islands*. Taylor & Francis, 2015. 9781317534013.

PARKER, E. *Vietnam*. Evans, 2005. 9780237527549.

PAWLEY, A. *Explaining the aberrant Austronesian languages of Southeast Melanesia: 150 years of debate*. *Journal of the Polynesian Society*, 2006.

PISCHEDDA, S. et al. *Phylogeographic and genome-wide investigations of Vietnam ethnic groups reveal signatures of complex historical demographic movements*. *Scientific Reports*, 7, n. 1, p. 12630, 2017/10/03 2017.

PURPS, J. et al. *A global analysis of Y-chromosomal haplotype diversity for 23 STR loci*. *Forensic Sci Int Genet*, 12, p. 12-23, Sep 2014.

PYERITZ, R. E.; KORF, B. R.; GRODY, W. W. *Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics: Cardiovascular, Respiratory, and Gastrointestinal Disorders*. Elsevier Science, 2019. 9780128126806.

RAMACHANDRAN, S. T., H.; GUTENKUNST, R.; BUSTAMANTE, C.;. *Genetics and Genomics of Human Population Structure*. Vogel and Motulsky's Human Genetics: Problems and Approaches,, p. 104, 2010. book.

RAYCHAUDHURI, S. *Mapping rare and common causal alleles for complex human diseases*. Cell, 147, n. 1, p. 57-69, Sep 30 2011.

REICH, D. et al. *Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania*. Am J Hum Genet, 89, n. 4, p. 516-528, Oct 7 2011.

REYES-CENTENO, H. et al. *Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia*. Proc Natl Acad Sci U S A, 111, n. 20, p. 7248-7253, May 20 2014.

ROBERT, F.; PELLETIER, J. *Exploring the Impact of Single-Nucleotide Polymorphisms on Translation*. Front Genet, 9, p. 507, 2018.

ROCHE, Y.; MICHAUD, J. *Mapping ethnic groups in Lao Cai Province, Vietnam*. Asia Pacific Viewpoint, 41, p. 101-110, 04/01 2000.

ROOTSI, S. et al. *A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe*. Eur J Hum Genet, 15, n. 2, p. 204-211, Feb 2007.

Sascha Willuweit & Lutz Roewer. Y\_STR Haplotype Reference Database. Disponível em: < <https://yhrd.org/>>. Acesso em: 20 ago. 2019.

SAITOU, N.; NEI, M. *The neighbor-joining method: a new method for reconstructing phylogenetic trees*. Mol Biol Evol, 4, n. 4, p. 406-425, Jul 1987.

SCHLIESINGER, J. *Ethnic Groups of Laos Vol 1: Introduction and Overview*. BooksMango, 2015. 9781633232310.

SCHUMACHER, B.; VIJG, J. *Age is in the nucleus*. Nature Metabolism, 1, n. 10, p. 931-932, 2019/10/01 2019.

SHAN, W. et al. *Genetic polymorphism of 17 Y chromosomal STRs in Kazakh and Uighur populations from Xinjiang, China*. Int J Legal Med, 128, n. 5, p. 743-744, Sep 2014.

SKALETSKY, H. et al. *The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes*. Nature, 423, n. 6942, p. 825-837, Jun 19 2003.

SKOGLUND, P. et al. *Genomic insights into the peopling of the Southwest Pacific*. Nature, 538, n. 7626, p. 510-513, Oct 27 2016.

TAYLOR, K. W. *A History of the Vietnamese*. Cambridge University Press, 2013. 9780521875868.

TIMMERMANN, A.; FRIEDRICH, T. *Late Pleistocene climate drivers of early human migration*. Nature, 538, n. 7623, p. 92-95, Oct 6 2016.

TIMSUKSAI, P.; RAMBO, A. T. *The Influence of Culture on Agroecosystem Structure: A Comparison of the Spatial Patterns of Homegardens of Different Ethnic Groups in Thailand and Vietnam*. PLoS One, 11, n. 1, p. e0146118, 2016.

TREJAUT, J. A. et al. *Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia*. BMC Genet, 15, p. 77, 2014.

UNDERHILL, P. A. et al. *The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations*. Ann Hum Genet, 65, n. Pt 1, p. 43-62, Jan 2001.

United Kingdom: Home Office, *Country Policy and Information Note Vietnam: Ethnic and Religious Groups*. 2018. Disponível em: <<https://www.refworld.org/docid/5bb5db7b7.html>>. Acesso em: 10 ago. 2019

VAN OVEN, M. et al. *Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome*. Hum Mutat, 35, n. 2, p. 187-191, Feb 2014.

VÂNIA PEREIRA, L. G. *Análise de Marcadores STRs de Cromossoma Y*. Book, Capítulo 9, p. 517, 2019.

Vietnam. *Vietnam Government Portal*. Hanoi, Vietnam: Copyright ©. Disponível em: <<http://chinhphu.vn/portal/page/portal/English/TheSocialistRepublicOfVietnam/AboutVietnam>>. Acesso em: 4 jun. 2019.

WAHAB, M. R. A. et al. *Selected Topics on Archaeology, History and Culture in the Malay World*. Springer Singapore, 2018. 9789811056697.

WANG, C. C.; LI, H. *Inferring human history in East Asia from Y chromosomes*. Investig Genet, 4, n. 1, p. 11, Jun 3 2013.

WEI, L. H. et al. *Phylogeography of Y-chromosome haplogroup O3a2b2-N6 reveals patrilineal traces of Austronesian populations on the eastern coastal regions of Asia*. PLoS One, 12, n. 4, p. e0175080, 2017.

WU, F. C. et al. *Y-chromosomal STRs haplotypes in the Taiwanese Paiwan population*. Int J Legal Med, 125, n. 1, p. 39-43, Jan 2011.

XIE, T. et al. *A set of autosomal multiple InDel markers for forensic application and population genetic analysis in the Chinese Xinjiang Hui group*. Forensic Sci Int Genet, 35, p. 1-8, Jul 2018.

YAN, S. et al. *An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4*. Eur J Hum Genet, 19, n. 9, p. 1013-1015, Sep 2011.

ZHABAGIN, M. et al. *Development of the Kazakhstan Y-chromosome haplotype reference database: analysis of 27 Y-STR in Kazakh population*. Int J Legal Med, 133, n. 4, p. 1029-1032, Jul 2019.

ZHAO, Z. et al. *Mutation rate estimation for 15 autosomal STR loci in a large population from Mainland China*. Meta Gene, 5, p. 150-156, Sep 2015.

ZHONG, H. et al. *Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia*. J Hum Genet, 55, n. 7, p. 428-435, Jul 2010.

ZHONG, H. S., H.; QI, X.; DUAN, Z.; TAN, P.; JIN, L.; SU, B.; MA, R.;. *Extended Y Chromosome Investigation Suggests Postglacial Migrations of Modern Humans into East Asia via the Northern Route*. Molecular Biology and Evolution., 28, n. 1, p. 10, 2011.

## APÊNDICE A - Tabela das amostras do Vietnã

Amostras	Etnia	Provincia	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Haplogrupo e marcador final
V003	Cao Lan	Yen Bai	16	14	24	31	16	16	16-18	14	10	11	20	13	10	14	11	18	O2a1*_M95
V004	Cao Lan	Yen Bai	14	12	24	28	18	14	13-18	12	10	13	20	14	12	15	11	20	O3a2c1a_M117
V005	Cao Lan	Yen Bai	15	13	24	29	17	16	13-18	13	11	12	23	13	11	14	10	18	O2a1*_M95
V006	Cao Lan	Yen Bai	14	12	25	28	17	16	12-12	12	10	12	22	13	13	14	10	19,2	O3a1*_L127
V007	Cao Lan	Yen Bai	15	12	24	28	19	16	13-22	13	11	11	23	13	12	15	10	20	O3a2*_P201
V008	Cao Lan	Yen Bai	15	13	24	30	19	15	13-17	15	11	12	23	13	11	14	10	18	O2a1*_M95
V009	Cao Lan	Yen Bai	15	14	23	31	16	15	11-12	13	11	12	24	14	11	14	11	19	N_M231
V010	Cao Lan	Yen Bai	15	13	25	29	17	15	15-21	14	11	12	22	13	11	14	10	18	O2a1*_M95
V011	Cao Lan	Yen Bai	14	12	25	28	17	17	12-12	12	10	12	22	13	13	14	10	19,2	O3a1*_L127
V012	Cao Lan	Yen Bai	15	14	23	31	15	14	11-12	13	11	11	24	14	12	14	11	19	N_M231
V013	Cao Lan	Yen Bai	14	11	24	28	17	16	14-18	14	11	15	21	13	10	14	10	18	O2a1*_M95
V014	Cao Lan	Yen Bai	15	12	23	28	17	16	15-18	12	10	11	21	13	12	15	10	20	O3a2*_P201
V015	Cao Lan	Yen Bai	15	13	24	30	18	17	13-17	14	11	12	22	13	11	14	10	18	O2a1*_M95
V016	Cao Lan	Yen Bai	14	12	24	30	16	14	14-15	13	10	11	19	11	12	15	11	19	D_M174
V020	Cao Lan	Yen Bai	14	12	25	27	18	16	12-19	12	10	12	22	13	12	14	10	19	O3a1*_L127
V021	Cao Lan	Yen Bai	15	14	24	30	19	16	12-18	14	11	12	23	13	11	14	10	18	O2a1*_M95
V024	Cao Lan	Yen Bai	16	14	24	31	16	16	16-18	14	10	11	20	13	10	14	11	18	O2a1*_M95
V025	Cao Lan	Yen Bai	15	13	24	29	19	15	14-21	12	10	12	21	13	11	15	10	19	O3a2*_P201
V031	Cao Lan	Yen Bai	15	13	23	30	14	14	11-12	13	10	12	24	14	12	14	10	19	N_M231
V032	Cao Lan	Yen Bai	15	13	24	29	16	15	12-18	14	11	12	22	13	12	14	10	18	O2*_L690
V033	Cao Lan	Yen Bai	15	13	24	29	17	16	13-19	14	11	12	21	13	11	14	10	17	O2a1*_M95
V034	Cao Lan	Yen Bai	15	13	23	30	14	14	11-12	13	10	9	24	14	12	14	10	19	N_M231
V035	Cao Lan	Yen Bai	14	12	24	30	16	14	14-15	13	10	11	19	11	12	15	11	19	D_M174
V036	Cao Lan	Yen Bai	15	13	23	30	14	14	11-12	13	10	12	24	14	12	14	10	19	N_M231
V037	Cao Lan	Yen Bai	15	12	24	30	19	15	14-16	?	10	11	23	13	10	14	10	18	O2a1*_M95
V038	Cao Lan	Yen Bai	15	13	23	30	14	14	11-12	13	10	12	24	14	12	14	10	19	N_M231
V053	Dao	Yen Bai	15	12	24	29	18	15	14-16	15	10	11	23	13	10	14	10	18	O2a1*_M95
V054	Dao	Yen Bai	15	13	24	28	17	16	13-19	14	11	12	22	14	11	14	10	18	O2a1*_M95
V055	Dao	Yen Bai	15	14	24	32	15	15	14-18	14	9	11	20	13	10	14	11	18	K_M9(xM22,Tat,M242)
V057	Dao	Yen Bai	15	13	24	29	20	15	13-18	15	11	12	22	13	11	14	10	18	O2a1*_M95
V058	Dao	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	K_M9(xM22,Tat,M242)
V059	Dao	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O3a2c1a_M117
V060	Dao	Yen Bai	15	12	24	28	15	15	14-17	14	10	11	24	13	10	14	11	17	O2a1*_M95
V061	Dao	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O1a*_M119
V062	Dao	Yen Bai	15	12	22	27	17	15	12-19	12	10	12	22	13	12	15	10	20	O3a2c*_P164
V063	Dao	Yen Bai	15	12	21	28	19	15	11-17	12	10	12	23	11	11	14	10	20	nd
V064	Dao	Yen Bai	15	13	24	29	18	16	13-18	14	11	13	24	13	11	14	10	18	O2a1*_M95
V065	Dao	Yen Bai	15	12	24	29	17	15	14-16	15	10	11	23	13	10	14	10	18	O2a1*_M95
V066	Dao	Yen Bai	16	12	25	28	17	16	13-22	10	11	14	24	13	11	14	10	18	O3a2*_P201
V067	Dao	Yen Bai	15	13	24	31	16	14	16-19	13	10	11	20	11	12	15	9	20	O3a2*_P201
V068	Dao	Yen Bai	15	14	24	32	15	15	14-18	14	9	11	20	13	10	14	11	18	nd

Amostras	Etnia	Provincia	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Haplogrupo e marcador final
V070	Dao	Yen Bai	15	13	24	31	16	14	16-19	13	10	11	20	11	12	15	9	20	nd
V071	Dao	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O2a1*_M95
V072	Dao	Yen Bai	14	11	25	29	18	16	13-18	14	11	11	22	13	10	14	10	18	O2a1*_M95
V073	Dao	Yen Bai	15	14	24	32	15	15	14-18	14	9	11	20	13	10	14	11	18	O2a1*_M95
V074	Dao	Yen Bai	15	12	23	28	17	14	14-16	12	10	12	20	14	12	15	11	20	O3a2c1a_M117
V075	Dao	Yen Bai	15	13	25	29	17	15	17-20	14	11	13	25	13	11	14	10	18	O2a1*_M95
V076	Dao	Yen Bai	15	14	24	32	15	15	14-18	14	9	11	20	13	10	14	11	18	O2a1*_M95
V077	Dao	Yen Bai	14	11	25	28	18	16	13-18	14	11	11	22	13	10	14	10	18	O2a1*_M95
V078	Dao	Yen Bai	15	14	24	32	15	15	14-18	14	9	11	20	13	10	14	11	18	O2a1*_M95
V079	Dao	Yen Bai	15	12	21	28	19	15	11-17	12	10	12	23	11	11	14	10	20	C_M130
V080	Dao	Yen Bai	15	14	24	32	15	15	14-18	14	9	11	20	13	10	14	11	18	O2a1*_M95
HCM09	Hoa	Ho Chi Minh	15	12	25	27	19	16	12-20	12	10	11	21	15	12	14	10	19	O3a1*_L127
HCM11	Hoa	Ho Chi Minh	16	12	23	27	14	15	13-13	13	11	12	23	14	13	14	10	18	O1a*_M119
HCM24	Hoa	Ho Chi Minh	16	14	24	30	15	16	12-19	14	10	12	21	13	12	14	10	18	O2*_L690
HCM29	Hoa	Ho Chi Minh	14	12	25	28	16	16	12-20	12	10	11	20	13	12	14	10	19	O3a1*_L127
HCM30	Hoa	Ho Chi Minh	16	14	25	29	15	15	15-19	14	10	11	21	13	11	14	10	18	O2a1*_M95
HCM36	Hoa	Ho Chi Minh	15	13	25	29	15	15	16-21	14	11	12	21	13	11	14	10	18	O2a1*_M95
HCM68	Hoa	Ho Chi Minh	15	12	24	28	18	15	15-18	13	11	11	23	13	10	13	10	18	O2a1*_M95
CB07	Kinh	Cao Bang	15	13	23	29	16	15	11-18	14	12	10	21	11	12	14	10	19	C_M130
CB08	Kinh	Cao Bang	14	13	24	30	18	14	13-19	14	11	13	21	13	11	14	10	18	O2a1*_M95
CB22	Kinh	Cao Bang	15	13	23	29	18	16	13-18	14	10	13	21	13	11	14	10	18	O2a1*_M95
CB27	Kinh	Cao Bang	14	14	24	29	16	15	12-17	13	10	11	21	13	11	14	10	18	O2*_L690
CB35	Kinh	Cao Bang	15	13	25	31	17	16	14-19	14	10	13	22	13	10	14	10	18	O2a1*_M95
CB37	Kinh	Cao Bang	15	13	24	30	16	15	14-18	14	10	11	20	13	10	14	11	18	O2a1*_M95
CB59	Kinh	Cao Bang	15	13	24	29	17	15	13-19	14	11	12	23	13	11	14	10	18	O2a1*_M95
CB60	Kinh	Cao Bang	15	14	24	30	18	17	13-19	13	10	12	21	13	11	14	10	18	O2a1*_M95
CB67	Kinh	Cao Bang	15	13	25	28	17	16	13-18	12	10	12	23	13	12	14	10	20	O3a2*_P201
CB72	Kinh	Cao Bang	15	13	24	30	18	14	15-15	14	11	12	21	13	11	14	10	18	O2a1*_M95
CB73	Kinh	Cao Bang	15	13	24	29	17	14	12-13-14	13	10	13	21	12	11	15	10	19	O2a1*_M95
CB76	Kinh	Cao Bang	15	13	24	30	18	17	13-19	13	11	11	22	13	11	14	10	18	O2a1*_M95
CB80	Kinh	Cao Bang	15	12	25	28	18	14	13-16	12	10	14	20	14	11	15	11	20	O3a2c1a_M117
CB85	Kinh	Cao Bang	15	13	25	28	17	16	13-18	12	10	12	23	13	12	15	10	21	O3a2*_P201
CB96	Kinh	Cao Bang	13	12	25	29	18	15	12-20	12	10	13	23	13	12	14	10	19	O3a1*_L127
ĐN02	Kinh	Da Nang	14	14	24	29	16	15	12-17	14	10	11	21	13	12	14	10	18	O2a1*_M95
ĐN04	Kinh	Da Nang	16	14	25	29	16	14	13-18	12	10	12	21	13	11	15	10	20	O3*_M122
ĐN05	Kinh	Da Nang	15	13	24	30	18	16	13-19	14	11	13	21	13	11	14	10	18	O2a1*_M95
ĐN06	Kinh	Da Nang	15	13	25	30	17	16	13-18	13	10	12	22	13	11	14	10	18	O2a1*_M95
ĐN07	Kinh	Da Nang	15	12	25	28	20	14	13-17	12	10	12	19	14	12	15	11	20	O3a2c1a_M117
ĐN14	Kinh	Da Nang	14	13	23	29	18	15	12-18	14	11	13	21	13	12	14	11	18	O2a1*_M95
ĐN21	Kinh	Da Nang	14	13	24	29	17	15	13-17	14	10	13	20	13	11	14	10	18	O2a1*_M95
ĐN23	Kinh	Da Nang	15	14	24	30	17	15	13-18	14	11	11	22	13	11	14	10	18	O2a1*_M95
ĐN24	Kinh	Da Nang	14	11	25	27	18	15	13-18	14	11	12	21	13	10	14	10	18	O2a1*_M95
ĐN32	Kinh	Da Nang	15	12	24	29	18	15	14-16	14	10	11	23	13	10	15	10	18	O2a1*_M95
ĐN34	Kinh	Da Nang	15	13	25	29	16	15	15-19	14	10	13	21	14	12	14	10	18	O2a1*_M95

Amostras	Etnia	Provincia	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Haplogrupo e marcador final
ĐN36	Kinh	Da Nang	14	11	25	29	20	16	14-18	14	11	12	21	13	10	14	10	19	O2a1*_M95
ĐN37	Kinh	Da Nang	17	13	25	29	17	15	15-19	14	10	14	21	13	11	14	10	18	O2a1*_M95
ĐN38	Kinh	Da Nang	16	13	26	29	17	16	16-19	14	10	11	23	13	11	14	10	17	O2a1*_M95
ĐN42	Kinh	Da Nang	14	11	25	27	17	16	13-18	14	11	12	21	13	10	14	10	18	O2a1*_M95
ĐN43	Kinh	Da Nang	14	11	25	28	18	16	13-18	14	11	12	21	13	10	14	10	18	O2a1*_M95
ĐN46	Kinh	Da Nang	15	13	25	28	17	16	13-19	12	10	12	23	13	12	15	10	21	O3a2*_P201
ĐN51	Kinh	Da Nang	14	12	25	28	18	16	8-18	13	11	12	21	13	10	14	10	18	O2a1*_M95
ĐN55	Kinh	Da Nang	15	12	24	28	17	16	12-20	12	10	11	20	13	12	14	10	19	O3a1*_L127
ĐN59	Kinh	Da Nang	14	11	25	28	18	15	13-18	14	11	12	21	13	10	14	10	18	O2a1*_M95
ĐN67	Kinh	Da Nang	15	13	24	30	19	16	13-19	14	10	13	22	13	11	14	11	18	O2a1*_M95
ĐN68	Kinh	Da Nang	15	12	23	29	15	15	13-15	12	10	11	21	12	12	14	10	19	O3a2c1*_M134
ĐN75	Kinh	Da Nang	14	11	26	28	18	15	13-18	14	11	12	20	13	10	14	10	18	O2a1*_M95
ĐN81	Kinh	Da Nang	15	14	24	30	18	16	12-20	14	11	12	24	13	11	14	10	18	O2a1*_M95
ĐN87	Kinh	Da Nang	15	14	24	30	16	15	13-14-18	14	11	13	22	13	11	14	10	18	O2a1*_M95
ĐN94	Kinh	Da Nang	15	13	24	29	17	14	14-18	12	11	12	20	15	12	16	10	20	O3a2c1a_M117
ĐN95	Kinh	Da Nang	16	13	24	32	18	15	12-20	14	10	13	23	13	11	14	10	18	O2*_L690
ĐN101	Kinh	Da Nang	15	14	23	30	17	16	13-18	14	11	12	22	13	11	14	10	18	O2a1*_M95
ĐN113	Kinh	Da Nang	14	14	24	29	18	17	12-19	12	10	12	23	13	13	14	10	19	O3a1*_L127
ĐN116	Kinh	Da Nang	16	13	24	30	15	15	16-19	14	11	12	21	13	11	14	10	18	O2a1*_M95
ĐN129	Kinh	Da Nang	15	12	24	28	17	15	13-19	12	10	11	21	14	12	15	11	20	O3a2c1a_M117
ĐN130	Kinh	Da Nang	15	13	26	28	18	15	17-21	14	10	12	22	13	11	14	10	18	O2a1*_M95
ĐN131	Kinh	Da Nang	15	13	24	29	16	16	13-19	14	11	12	22	13	11	14	10	17	O2a1*_M95
HN01	Kinh	Ha Noi	15	14	24	30	19	16	12-13-18	14	11	12	21	13	11	14	10	18	O2a1*_M95
HN02	Kinh	Ha Noi	15	13	23	30	16	15	11-18	14	10	10	20	11	12	14	10	20	C_M130
HN03	Kinh	Ha Noi	16	12	23	29	16	15	12-16	12	10	13	19	12	12	14	10	19	O3a2c1*_M134
HN04	Kinh	Ha Noi	14	11	25	28	18	16	14-17	14	11	12	21	13	10	14	10	18	O2a1*_M95
HN05	Kinh	Ha Noi	13	12	25	28	15	15	12-20	12	10	12	23	13	12	14	10	19	O3a1*_L127
HN06	Kinh	Ha Noi	14	13	25	30	19	17	14-19	12	10	11	23	13	12	14	10	20	O3a1*_L127
HN07	Kinh	Ha Noi	14	11	25	28	19	16	13-18	14	11	11	21	13	10	14	10	19	O1a*_M119
HN08	Kinh	Ha Noi	15	13	24	30	18	15	13-18	14	10	11	20	13	11	14	10	17	O2a1*_M95
HN10	Kinh	Ha Noi	15	12	24	27	17	16	13-18	14	10	12	21	13	11	14	10	18	O2a1*_M95
HN11	Kinh	Ha Noi	15	12	24	28	18	14	12-20	11	9	14	21	14	12	15	11	20	O3a2c1a_M117
HN12	Kinh	Ha Noi	15	13	24	29	18	16	11-18	15	10	12	21	11	12	14	10	21	C_M130
HN13	Kinh	Ha Noi	14	14	24	29	18	17	12-20	12	10	12	23	13	13	14	10	19	O3a1*_L127
HN14	Kinh	Ha Noi	15	13	22	29	14	14	13-13	13	10	12	21	14	12	15	10	18	O1a*_M119
HN15	Kinh	Ha Noi	15	12	25	29	18	17	13-18	14	11	12	22	13	11	14	10	18	O2a1*_M95
HN18	Kinh	Ha Noi	15	13	24	28	17	15	13-13	12	10	12	23	13	12	14	10	21	O3a2*_P201
HN20	Kinh	Ha Noi	15	14	25	30	16	15	15-19	14	10	11	21	13	11	14	10	18	O2a1*_M95
HN21	Kinh	Ha Noi	15	13	24	29	20	15	15-17	12	10	11	20	10	11	14	10	19	D_M174
HN22	Kinh	Ha Noi	17	14	24	30	16	13	16-20	14	6	11	22	14	10	14	12	19	Q_P36.2
HN23	Kinh	Ha Noi	15	12	23	27	17	17	13-20	13	10	13	23	13	11	15	10	22	O3*_M122
HN24	Kinh	Ha Noi	16	14	25	30	15	15	16-19	14	10	11	21	13	11	14	10	18	O2a1*_M95
HN25	Kinh	Ha Noi	16	13	25	29	15	15	16-20	15	10	11	22	13	11	14	10	18	O2a1*_M95
HN26	Kinh	Ha Noi	15	13	24	29	18	16	13-16	14	11	13	22	13	11	14	10	18	O2a1*_M95

Amostras	Etnia	Provincia	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Haplogrupo e marcador final
HN27	Kinh	Ha Noi	14	13	23	29	17	15	12-17	14	11	13	21	13	12	14	11	18	O2a1*_M95
HN28	Kinh	Ha Noi	15	13	25	28	18	15	17-20	14	10	12	22	14	11	14	10	18	O2a1*_M95
HN29	Kinh	Ha Noi	15	13	24	30	17	16	13-19	14	11	12	22	13	11	14	10	18	O3*_M122
HN30	Kinh	Ha Noi	14	13	23	29	18	17	13-18	14	10	14	21	13	11	14	10	18	O2a1*_M95
HN31	Kinh	Ha Noi	14	11	25	28	18	17	13-18	14	11	12	21	13	10	14	11	18	O2a1*_M95
HN32	Kinh	Ha Noi	15	13	25	29	16	16	15-19	14	11	12	22	13	11	14	11	18	O2a1*_M95
HN43	Kinh	Ha Noi	15	14	24	31	17	16	13-20	14	10	13	21	13	11	14	10	19	O2a1*_M95
HN44	Kinh	Ha Noi	15	13	24	28	18	16	13-18	12	10	12	23	13	12	15	10	21	O3a2*_P201
HN45	Kinh	Ha Noi	15	13	24	30	17	16	19-20	14	11	13	22	13	11	14	10	18	O2a1*_M95
HN46	Kinh	Ha Noi	13	12	25	28	18	15	12-21	12	10	12	23	13	12	14	10	19	O3a1*_L127
HN47	Kinh	Ha Noi	15	12	22	29	17	15	12-17	12	11	12	19	12	12	15	10	19	O3a2c1*_M134
HN49	Kinh	Ha Noi	15	12	25	28	17	15	13-17	12	10	11	21	14	12	15	11	20	O3a2c1a_M117
HN50	Kinh	Ha Noi	17	12	23	29	15	15	13-13	13	11	11	19	14	12	14	10	18	O1a*_M119
HP02	Kinh	Hai Phong	15	13	23	29	16	14	12-13	13	10	11	21	15	12	14	10	19	N_M231
HP14	Kinh	Hai Phong	15	14	23	30	16	15	11-19	15	10	11	21	11	11	14	10	21	C_M130
HP16	Kinh	Hai Phong	15	14	24	30	15	15	15-20	14	10	11	21	13	11	14	10	18	O2a1*_M95
HP18	Kinh	Hai Phong	15	12	24	28	18	16	13-18	14	11	12	21	13	11	14	10	18	O2a1*_M95
HP80	Kinh	Hai Phong	15	15	25	33	20	16	11-19	13	10	11	23	13	11	14	10	18	O2a1*_M95
HP96	Kinh	Hai Phong	15	13	24	28	16	17	13-18	14	12	13	22	13	11	14	11	18	O2a1*_M95
HP105	Kinh	Hai Phong	13	12	25	28	18	15	12-20	12	10	12	23	13	12	14	10	19	O3a1*_L127
HP107	Kinh	Hai Phong	14	12	24	29	16	15	13-17	12	10	13	20	14	12	15	11	20	O3a2c1*_M134
HP114	Kinh	Hai Phong	14	11	26	27	18	16	13-18	13	10	12	21	13	10	14	10	18	O2a1*_M95
HP123	Kinh	Hai Phong	15	13	24	30	19	15	13-18	15	12	12	23	13	11	14	10	18	O2a1*_M95
HP126	Kinh	Hai Phong	15	12	24	28	20	15	15-17	12	10	11	20	10	11	14	10	19	D_M174
HP127	Kinh	Hai Phong	15	13	24	30	18	16	13-18	14	11	12	21	13	11	14	10	18	O2a1*_M95
HP129	Kinh	Hai Phong	15	12	24	30	18	15	14-18	15	11	11	24	13	10	14	10	18	O2a1*_M95
HCM04	Kinh	Ho Chi Minh	14	11	25	27	18	16	13-18	14	11	12	21	13	10	14	10	18	O2a1*_M95
HCM14	Kinh	Ho Chi Minh	15	13	24	29	16	16	12-17	14	10	11	21	13	12	14	10	18	O2*_L690
HCM15	Kinh	Ho Chi Minh	15	14	25	29	18	16	13-18	12	10	12	23	13	12	15	10	21	O3a2*_P201
HCM18	Kinh	Ho Chi Minh	15	12	23	30	19	15	12-12	12	10	12	24	12	14	14	11	21	O3a2c1*_M134
HCM27	Kinh	Ho Chi Minh	14	12	25	28	17	16	12-19	12	10	12	23	13	11	14	10	19	O3a1*_L127
HCM40	Kinh	Ho Chi Minh	15	13	24	28	18	15	13-22	12	11	11	21	14	11	15	10	19	O3a2*_P201
HCM41	Kinh	Ho Chi Minh	15	13	25	29	20	15	19-19	14	11	12	23	13	11	14	10	18	O2a1*_M95
HCM43	Kinh	Ho Chi Minh	15	13	23	28	19	16	12-18	12	10	11	21	13	12	15	10	20	O3a2*_P201
HCM45	Kinh	Ho Chi Minh	14	13	23	29	18	15	12-17	14	11	13	21	13	12	14	11	18	O2*_L690
HCM56	Kinh	Ho Chi Minh	17	13	24	28	17	15	12-18	14	10	13	21	13	12	14	10	18	O2*_L690
HCM58	Kinh	Ho Chi Minh	15	13	23	29	17	15	12-21	14	10	13	23	13	11	14	10	18	O2*_L690
HCM62	Kinh	Ho Chi Minh	14	12	23	28	17	17	12-19	12	11	11	21	13	12	14	10	19	O3a1*_L127
HCM64	Kinh	Ho Chi Minh	15	14	23	30	17	15	11-18	14	11	11	21	11	11	14	10	21	C_M130
HCM67	Kinh	Ho Chi Minh	15	12	24	29	18	15	14-16	15	10	11	23	13	10	14	10	18	O2a1*_M95
HCM71	Kinh	Ho Chi Minh	15	13	24	31	17	13	14-15	13	10	11	22	12	10	13	11	20	Q_M346
HCM72	Kinh	Ho Chi Minh	16	14	24	30	19	15	13-18	14	10	12	22	13	11	14	10	18	O2a1*_M95
HCM77	Kinh	Ho Chi Minh	15	12	22	28	15	14	13-17	11	10	14	23	14	12	15	10	19	L_M22
HCM92	Kinh	Ho Chi Minh	14	11	25	27	18	16	13-18	14	11	12	21	13	10	14	10	17	O2a1*_M95



Amostras	Etnia	Provincia	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Haplogrupo e marcador final
HCM93	Kinh	Ho Chi Minh	16	13	24	29	15	15	14.2-20	14	10	11	21	13	11	14	10	17	O2a1*_M95
HCM96	Kinh	Ho Chi Minh	15	13	25	30	17	16	13-19	14	12	11	21	13	11	14	10	18	O2a1*_M95
HCM99	Kinh	Ho Chi Minh	16	14	23	29	18	15	12-19	12	10	12	21	12	12	15	11	19	O3a2c1*_M134
HCM100	Kinh	Ho Chi Minh	15	13	23	29	17	16	15-17	12	10	11	21	13	12	15	10	20	O3a2*_P201
HCM104	Kinh	Ho Chi Minh	15	14	25	30	16	15	13-13	13	11	12	22	11	12	14	10	21	C_M130
HCM105	Kinh	Ho Chi Minh	15	14	24	31	16	16	13-20	14	11	12	23	13	11	14	10	18	O2a1*_M95
V160	Kinh	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O3a2*_P201
V161	Kinh	Yen Bai	14	14	25	32	16	17	13-19	13	10	12	22	13	11	14	10	18	O2a1*_M95
V164	Kinh	Yen Bai	15	13	25	30	17	15	15-21	14	11	12	22	13	11	14	10	18	O2a1*_M95
V169	Kinh	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O2a1*_M95
V171	Kinh	Yen Bai	15	14	22	32	18	15	11-13	13	10	11	22	14	12	14	10	19	N_M231
ĐN17	Kinh	Da Nang	16	14	24	30	18	16	13-13	14	10	11	22	13	11	14	10	18	O2a1*_M95
ĐN65	Kinh	Da Nang	15	12	24	29	18	15	14-17	15	10	13	23	13	10	14	10	18	O2a1*_M95
ĐN102	Kinh	Da Nang	15	13	24	30	17	15	12-17	12	10	12	20	13	12	15	10	20	O3a2c*_P164
ĐN134	Kinh	Da Nang	12	12	23	28	20	14	13-18	12	10	10	20	14	10	15	11	20	O3a2c1a_M117
HP08	Kinh	Hai Phong	15	12	24	28	17	14	14-18	12	10	12	20	14	12	15	11	20	O3a2c1a_M117
HP11	Kinh	Hai Phong	15	13	23	29	19	16	13-18	14	10	12	21	13	11	14	10	18	O2a1*_M95
HP40	Kinh	Hai Phong	15	13	25	28	17	16	13-20	12	10	14	22	13	12	15	10	21	O3a2*_P201
HP47	Kinh	Hai Phong	16	14	22	30	15	15	15-19	14	10	11	21	13	11	14	10	18	O2a1*_M95
V085	Mong	Yen Bai	14	12	24	29	17	16	13-21	12	10	10	23	13	12	14	10	20	O3a2*_P201
V090	Mong	Yen Bai	14	12	24	29	16	14	15-15	13	11	11	19	11	12	15	11	19	D_M174
LC04	Mong	Lao Cai	15	12	23	27	16	14	13-19	12	10	12	20	14	12	15	11	20	O3a2c1a_M117
LC06	Mong	Lao Cai	15	12	23	27	16	14	13-19	12	10	12	20	14	12	15	11	20	O3a2c1a_M117
LC09	Mong	Lao Cai	14	12	24	28	19	17	13-22	12	10	11	23	13	12	15	10	20	O3a2*_P201
LC10	Mong	Lao Cai	14	12	24	30	16	14	15-15	13	10	12	19	11	12	15	11	19	D_M174
LC13	Mong	Lao Cai	15	12	24	28	19	17	13-22	12	10	11	23	13	12	15	10	20	K_M9(xM22,Tat,M242)
LC15	Mong	Lao Cai	15	12	24	27	19	16	13-23	12	10	12	23	13	12	15	10	20	O3a2*_P201
LC18	Mong	Lao Cai	15	12	24	29	17	17	13-21	12	10	10	23	13	12	15	10	19	O3a2*_P201
LC19	Mong	Lao Cai	15	12	23	27	16	14	13-19	12	10	12	20	14	12	15	11	20	O3a2c1a_M117
LC20	Mong	Lao Cai	15	12	24	28	18	17	13-22	12	10	11	23	13	12	14	10	20	O3a2*_P201
LC21	Mong	Lao Cai	15	12	24	28	19	18	13-22	12	10	11	23	13	12	15	10	20	O3a2*_P201
LC22	Mong	Lao Cai	15	13	24	29	15	15	11-12	15	10	11	23	11	11	14	10	21	C_M130
LC23	Mong	Lao Cai	15	12	24	28	18	17	13-22	12	10	11	22	13	12	15	10	19	O3a2*_P201
LC25	Mong	Lao Cai	15	12	24	28	18	17	13-22	12	10	11	23	13	12	15	10	20	O3a2*_P201
LC27	Mong	Lao Cai	15	12	24	29	17	17	13-20	12	10	10	23	13	12	15	10	19	O3a2*_P201
LC32	Mong	Lao Cai	14	13	24	30	19	14	13-19	12	10	12	20	14	12	14	11	20	O3a2c1a_M117
LC33	Mong	Lao Cai	14	13	24	30	19	14	13-19	12	10	12	20	14	12	14	11	20	O3a2c1a_M117
LC35	Mong	Lao Cai	14	12	24	29	17	14	15-15	13	10	11	19	11	12	15	11	19	D_M174
LC37	Mong	Lao Cai	15	12	24	29	17	17	13-20	12	10	10	23	13	11	15	10	19	O3a2c1*_M134
LC39	Mong	Lao Cai	14	12	24	29	17	17	13-21	12	11	10	24	13	12	15	10	20	O3a2*_P201
LC40	Mong	Lao Cai	15	14	24	31	16	15	15-17	14	10	11	20	12	10	14	10	18	O2a1*_M95
LC44	Mong	Lao Cai	15	12	24	29	17	17	13-20	12	10	10	23	13	11	15	10	19	O1a*_M119
LC45	Mong	Lao Cai	14	13	22	29	15	15	13-13	13	10	12	23	14	11	15	10	18	O1a*_M119
LC46	Mong	Lao Cai	14	12	24	29	16	14	15-15	13	10	11	19	11	12	15	11	19	O2a1*_M95

Amostras	Etnia	Provincia	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Haplogrupo e marcador final
LC49	Mong	Lao Cai	15	13	24	29	15	15	11-11	15	10	11	22	11	12	14	10	21	C_M130
LC50	Mong	Lao Cai	15	12	23	28	18	15	12-17	12	10	13	19	12	11	14	10	20	O3a2c1*_M134
LC51	Mong	Lao Cai	15	13	23	29	15	15	11-11	14	11	11	21	11	11	15	10	21	C_M130
LC53	Mong	Lao Cai	14	12	24	29	17	15	12-18	13	10	10	21	13	11	15	10	19	O3a2c*_P164
LC54	Mong	Lao Cai	15	12	24	29	17	17	13-20	12	10	10	23	13	11	15	10	19	O3a2*_P201
LC56	Mong	Lao Cai	15	13	23	30	18	14	12-15	13	11	11	21	15	12	14	10	19	N_M231
LC58	Mong	Lao Cai	14	12	24	29	17	17	12-21	12	10	10	22	14	12	15	10	20	O3a2*_P201
LC60	Mong	Lao Cai	15	12	24	28	18	16	13-23	12	10	11	22	13	12	15	10	20	O3a2*_P201
LC65	Mong	Lao Cai	15	13	23	29	15	15	11-11	14	11	11	21	11	11	15	10	21	C_M130
LC66	Mong	Lao Cai	15	13	24	29	18	16	13-18	14	11	13	22	13	11	14	10	18	nd
LC67	Mong	Lao Cai	14	12	24	30	17	17	13-20	12	10	11	23	13	12	15	10	20	O3a2*_P201
LC69	Mong	Lao Cai	15	12	24	27	18	17	13-22	12	10	11	23	13	12	15	10	20	O3a2*_P201
LC71	Mong	Lao Cai	15	13	24	29	19	16	13-22	12	10	11	22	13	12	15	10	20	O3a2*_P201
LC72	Mong	Lao Cai	15	13	24	29	20	17	13-22	12	10	11	22	13	12	15	10	20	O3a2*_P201
LC73	Mong	Lao Cai	15	12	24	29	19	17	13-22	12	10	11	24	13	12	15	10	20	O3a2*_P201
LC78	Mong	Lao Cai	14	12	24	29	17	17	13-21	12	10	10	23	13	11	15	10	20	O3a2*_P201
LC80	Mong	Lao Cai	15	13	24	29	18	17	13-22	12	10	11	22	13	11	15	10	20	O3a2*_P201
LC81	Mong	Lao Cai	16	12	25	28	18	17	13-23	12	10	11	24	13	12	15	10	20	O3a2*_P201
LC82	Mong	Lao Cai	15	13	24	29	18	17	13-22	12	10	11	22	13	11	15	10	20	O3a2*_P201
LC83	Mong	Lao Cai	15	13	24	29	18	17	13-22	12	10	11	22	13	11	15	10	20	O3a2*_P201
LC84	Mong	Lao Cai	15	13	24	29	18	17	13-22	12	10	11	22	13	11	15	10	20	O3a2*_P201
LC85	Mong	Lao Cai	15	13	23	29	15	15	11-11	14	11	11	21	11	11	15	10	21	C_M130
LC86	Mong	Lao Cai	15	13	24	29	19	17	13-22	12	10	11	22	13	12	15	10	20	O3a2*_P201
LC88	Mong	Lao Cai	15	13	24	30	18	16	13-18	14	11	13	22	13	11	14	10	18	O2a1*_M95
LC89	Mong	Lao Cai	15	13	23	30	18	14	12-15	13	11	11	21	15	12	14	10	19	N_M231
LC94	Mong	Lao Cai	14	12	25	27	17	15	15-18	12	10	12	23	13	12	14	10	18	O3a1*_L127
LC96	Mong	Lao Cai	15	13	24	29	15	15	11-11	15	11	11	21	11	11	14	10	21	C_M130
LC97	Mong	Lao Cai	15	12	24	29	17	17	13-21	12	10	10	23	13	12	15	10	19	O3a2*_P201
LC99	Mong	Lao Cai	15	12	24	29	17	17	13-21	12	10	10	23	13	12	15	10	19	O2a1*_M95
LC102	Mong	Lao Cai	15	12	24	29	17	14	13-19	12	10	14	20	14	13	16	11	20	O3a2c1*_M134
LC103	Mong	Lao Cai	17	14	23	32	16	14	11-12	13	10	11	24	14	12	14	11	19	N_M231
LC105	Mong	Lao Cai	15	12	24	28	19	17	12-22	12	10	11	23	13	12	15	10	20	O3a2*_P201
LC111	Mong	Lao Cai	17	12	24	28	16	15	13-14	13	10	10	19	14	12	14	10	18	O1a*_M119
LC112	Mong	Lao Cai	17	12	24	28	16	15	13-14	13	10	10	19	14	12	14	10	18	O1a*_M119
LC114	Mong	Lao Cai	15	11	24	27	18	16	13-20	12	10	11	23	13	12	15	10	20	O3a2*_P201
LC115	Mong	Lao Cai	15	11	24	27	18	16	13-20	12	10	11	23	13	12	15	10	20	O3a2*_P201
LC116	Mong	Lao Cai	15	13	24	29	18	16	13-18	14	11	13	22	13	11	14	10	18	O2a1*_M95
LC118	Mong	Lao Cai	14	12	25	27	18	16	12-19	12	10	12	21	12	12	14	10	19	O3a1*_L127
LC119	Mong	Lao Cai	14	12	25	27	18	16	12-19	12	10	12	21	12	12	14	10	19	O3a1*_L127
V098	Muong	Hoa Binh	15	12	24	29	18	15	14-16	15	10	12	23	13	10	14	10	18	O2a1*_M95
V100	Muong	Hoa Binh	14	12	24	28	20	16	13-15	12	10	11	21	13	12	14	10	18	O3a1*_L127
V103	Muong	Hoa Binh	15	14	23	30	19	17	13-18	13	10	12	21	13	11	14	10	18	O2a1*_M95
V104	Muong	Hoa Binh	15	13	24	29	19	16	13-17	14	11	12	22	13	11	14	10	18	O2a1*_M95
V105	Muong	Hoa Binh	15	13	25	30	18	17	13-18	14	11	11	22	13	11	14	10	17	O2a1*_M95

Amostras	Etnia	Provincia	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Haplogrupo e marcador final
V106	Muong	Hoa Binh	16	13	24	30	18	16	13-19	14	11	12	22	14	11	14	10	19	O2a1*_M95
V107	Muong	Hoa Binh	16	13	24	29	18	16	13-19	14	11	11	22	13	11	14	10	18	O2a1*_M95
V114	Muong	Hoa Binh	15	12	24	29	18	15	14-16	14	10	11	23	13	10	14	10	18	O2a1*_M95
V115	Muong	Hoa Binh	15	13	24	30	19	16	13-19	14	11	11	23	13	11	14	10	18	O2a1*_M95
V119	Muong	Hoa Binh	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O2a1*_M95
V120	Muong	Hoa Binh	15	14	25	31	18	16	12-18	14	11	12	23	13	11	14	10	18	O2a1*_M95
V123	Muong	Hoa Binh	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O3a1*_L127
V127	Muong	Hoa Binh	16	13	25	29	17	16	15-20	14	11	12	22	13	11	14	10	18	O2a1*_M95
V136	Muong	Hoa Binh	16	13	25	29	16	15	15-19	14	11	12	21	13	11	14	11	19	O2a1*_M95
V095	Muong	Yen Bai	15	12	24	28	18	14	13-17	12	10	15	20	14	12	15	11	20	O3a2c1a_M117
V096	Muong	Yen Bai	14	11	25	28	19	16	12-17	13	11	12	21	13	10	14	10	18	O2a1*_M95
V097	Muong	Yen Bai	15	13	23	29	20	16	13-19	14	10	12	21	13	11	14	10	18	O2a1*_M95
CB05	Nung	Cao Bang	15	12	23	28	17	15	13-14	13	10	11	20	14	12	14	10	18	O1a*_M119
CB12	Nung	Cao Bang	15	13	23	29	18	17	13-18	14	11	14	23	13	11	14	10	18	O2a1*_M95
CB29	Nung	Cao Bang	15	13	25	31	16	15	13-19	14	11	12	22	13	11	14	10	18	O2a1*_M95
CB31	Nung	Cao Bang	15	14	24	31	16	16	14-18	14	10	11	20	13	10	14	11	18	O2a1*_M95
CB74	Nung	Cao Bang	15	13	25	29	16	15	13-17	13	11	12	22	13	11	14	10	18	O2a1*_M95
CB86	Nung	Cao Bang	16	13	22	29	18	16	11-11	12	10	12	21	13	12	15	10	19	O3a1*_L127
CB94	Nung	Cao Bang	15	13	24	30	18	16	13-19	14	11	12	24	13	11	14	10	19	O2a1*_M95
CB112	Nung	Cao Bang	16	14	25	30	16	15	15-19	14	10	11	21	13	10	14	11	18	O2a1*_M95
HN09	Nung	Ha Noi	15	13	25	29	17	15	15-21	14	10	12	22	13	11	14	10	18	O2a1*_M95
V139	Nung	Yen Bai	15	14	25	30	17	15	15-20	14	11	12	22	13	11	14	10	18	O2a1*_M95
CB04	Tay	Cao Bang	18	12	23	28	18	16	11-18	12	10	11	19	12	12	14	10	19	O3a2c1*_M134
CB10	Tay	Cao Bang	15	13	25	29	18	15	15-22	14	11	12	23	13	11	14	10	18	O2a1*_M95
CB18	Tay	Cao Bang	15	13	24	29	19	16	14-18	14	11	12	21	13	11	14	11	18	O2a1*_M95
CB21	Tay	Cao Bang	15	13	23	29	17	14	12-13	14	10	11	21	15	12	14	10	19	N_M231
CB40	Tay	Cao Bang	14	12	24	30	16	14	15-15	12	10	11	20	11	12	15	11	19	D_M174
CB41	Tay	Cao Bang	15	14	25	31	15	15	15-20	14	11	12	21	13	11	14	10	18	O2a1*_M95
CB43	Tay	Cao Bang	15	14	25	31	15	15	15-19	14	11	12	21	13	11	14	10	18	O2a1*_M95
CB45	Tay	Cao Bang	17	13	24	28	18	14	13-21	12	10	12	20	14	15	16	11	20	O3a2c1a_M117
CB46	Tay	Cao Bang	15	12	24	30	18	15	15-16	15	10	11	22	13	10	14	10	18	O2a1*_M95
CB47	Tay	Cao Bang	15	13	25	28	18	16	13-19	12	10	13	23	13	12	15	10	21	O3a2*_P201
CB49	Tay	Cao Bang	15	13	24	30	16	15	13-17	14	11	12	22	13	11	14	10	17	O2a1*_M95
CB51	Tay	Cao Bang	15	12	24	30	18	15	15-16	15	10	11	22	13	10	14	10	18	O2a1*_M95
CB55	Tay	Cao Bang	15	13	24	30	18	16	13-19	14	11	12	21	13	11	14	10	18	O2a1*_M95
CB58	Tay	Cao Bang	14	13	23	29	17	12	12-17	13	10	13	21	11	10	16	10	19	F_M213(xK)
CB63	Tay	Cao Bang	15	15	25	31	15	15	15-19	14	11	12	21	13	11	14	10	18	O2a1*_M95
CB64	Tay	Cao Bang	15	12	24	28	17	16	12-20	15	11	12	23	13	11	14	10	18	O2a1*_M95
CB71	Tay	Cao Bang	15	13	25	29	17	15	15-21	14	11	12	22	13	11	14	10	18	O2a1*_M95
CB75	Tay	Cao Bang	15	12	23	29	18	15	14-16	14	10	11	22	13	10	14	10	18	O2a1*_M95
CB82	Tay	Cao Bang	15	14	23	31	15	14	11-12	13	11	12	24	14	12	14	11	19	N_M231
CB92	Tay	Cao Bang	15	13	25	30	18	15	15-21	14	11	12	21	13	11	14	10	18	O2a1*_M95
CB98	Tay	Cao Bang	15	13	25	29	17	15	15-21	14	11	13	22	13	11	14	10	18	O2a1*_M95
CB99	Tay	Cao Bang	15	15	25	32	16	15	15-20	14	11	12	21	13	11	14	10	18	O2a1*_M95

Amostras	Etnia	Provincia	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Haplogrupo e marcador final
CB115	Tay	Cao Bang	15	14	25	31	15	15	14-20	14	11	12	21	13	11	14	10	18	O2a1*_M95
CB118	Tay	Cao Bang	15	14	25	31	15	15	14-14	14	11	12	21	13	11	14	10	18	O2a1*_M95
HN48	Thai	Ha Noi	15	12	24	28	19	14	13-16	12	10	12	21	14	12	15	11	20	O3a2c1a_M117
V141	Thai	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O2a1*_M95
V145	Thai	Yen Bai	14	12	24	30	17	14	15-15	14	10	11	19	11	12	15	11	19	D_M174
V146	Thai	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O2a1*_M95
V149	Thai	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O2a1*_M95
V150	Thai	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O2a1*_M95
V151	Thai	Yen Bai	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	O2a1*_M95
V153	Thai	Yen Bai	15	12	25	28	17	14	13-17	12	10	12	21	14	12	15	11	20	O3a2c1a_M117
V156	Thai	Yen Bai	16	12	23	29	19	16	13-19	14	11	13	21	13	11	14	10	18	O2a1*_M95
V174	Thai	Yen Bai	15	13	23	30	16	16	11-20	14	10	11	22	11	12	14	11	20	C_M130
V176	Thai	Yen Bai	15	12	24	28	17	16	13-20	12	10	13	21	13	12	15	10	21	O3a2*_P201

Fonte: A Autora, 2019.

## APÊNDICE B - Tabela das amostras das Filipinas

Amostras	Provincia	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	GATA-H4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	Haplogrupo e marcador final	
PHI002	Visayas	17	12	20	28	19	9	18	15	11	20	10	11	25	14	14	34	18	14	14	34	12	12	23	36	37	2	O3a2c1a_M117
PHI003	Luzón	62	12	18	18	21	10	15	15	21	80	17	40	74	04	66	94	11	13	13	31	13	14	24	36	41	1	O1a*_M119
PHI007	Luzón	19	22	29	29	19	10	16	13	11	20	12	45	03	77	64	13	13	22	32	13	14	24	39	40	1	O3*_M122	
PHI008	Luzón	19	22	18	18	24	9	16	15	21	11	15	40	54	03	77	64	11	16	30	13	12	20	40	41	4	O3a2c*_P164	
	Luzón	18	21	17	17	23	11	15	15	11	17	10	73	15	65	74	12	17	30	13	10	30	34	39	40	2	O1a*_M119	
PHI021	Mindanao	18	21	18	18	25	9	16	15	21	91	15	40	39	03	99	64	13	16	31	12	35	40	39	40	3	O3a2c*_P164	
PHI025	Luzón	19	31	29	19	21	10	18	15	33	90	15	40	54	03	53	45	16	31	13	33	42	39	39	2	O3a2c*_P164		
PHI026	Mindanao	17	23	38	25	10	16	15	15	21	91	15	40	36	64	12	31	12	31	23	54	39	40	2	O3a2c*_P164			
PHI033	Visayas	84	13	33	24	10	17	12	12	11	20	10	54	03	95	518	29	13	18	29	13	13	36	38	1	O3*_M122		
PHI034	Luzón	20	31	10	21	9	20	15	15	21	80	15	20	52	03	88	85	11	32	12	21	44	36	40	1	O3a1*_L127		
PHI037	Visayas	84	10	0	22	10	16	14	14	11	00	64	13	48	419	31	13	19	31	23	33	39	1	O3*_M122				
PHI041	Mindanao	82	08	21	21	10	16	15	15	21	90	45	03	77	74	19	33	12	33	22	16	37	2	O3a1*_L127				
PHI048	Luzón	17	31	0	22	9	16	15	15	21	91	15	40	37	64	15	31	12	31	23	54	39	41	3	O3a2c*_P164			
PHI052	Luzón	19	23	38	23	9	17	15	15	33	90	55	03	77	54	17	32	13	22	32	22	39	40	2	O3a2c*_P164			
PHI053	Luzón	18	24	8	20	10	18	14	14	21	11	15	10	10	10	74	11	16	30	12	37	40	1	C_M130				
PHI057	Visayas	82	48	23	10	17	15	15	15	21	90	54	03	96	516	32	13	16	32	13	24	37	38	2	O3a2c*_P164			
PHI058	Luzón	18	40	22	21	11	19	14	14	21	90	51	01	37	413	29	13	13	29	14	23	37	37	2	nd			
PHI059	Visayas	19	30	29	20	10	18	13	13	21	90	55	58	22	99	415.2	23	15	23	22	22	37	40	1	D_M174			
PHI060	Visayas	17	21	19	19	10	15	15	15	21	80	73	04	73	94	13	30	13	33	33	33	38	1	O1a*_M119				
PHI064	Visayas	18	22	28	23	9	16	16	16	33	90	54	40	38	54	15	29	12	29	13	24	40	2	O3a2c*_P164				
PHI065	Visayas	17	21	19	22	9	16	15	15	21	90	54	40	36	44	16	32	13	32	44	39	39	2	O3a2c*_P164				
PHI069	Visayas	19	43	32	19	12	16	15	15	21	20	05	51	13	13	94	14	14	31	13	03	37	40	2	nd			
PHI070	Luzón	16	21	19	21	10	14	15	15	21	20	95	49	14	91	17	32	14	32	12	22	37	38	2	nd			
PHI074	Luzón	16	21	18	21	10	17	17	17	21	20	04	40	37	84	17	27	14	27	15	38	1	O3a1*_L127					
PHI076	Luzón	16	31	0	20	9	17	15	15	21	90	75	13	86	521	32	14	21	32	12	33	38	1	O3a1*_L127				
PHI077	Luzón	19	31	12	21	9	18	15	15	21	91	43	03	46	420	30	12	13	21	33	40	1	O3a1*_L127					
PHI078	Luzón	19	31	12	21	9	18	15	15	21	91	43	03	46	420	30	12	13	21	33	40	1	O3a1*_L127					
PHI082	Luzón	15	22	29	19	10	15	15	15	21	47	95	10	48	64	14	29	14	29	13	41	36	41	1	O1a*_M119			
PHI083	Luzón	18	30	00	17	10	16	16	16	21	29	17	30	49	94	15	30	13	30	22	35	38	39	1	O1a*_M119			
PHI085	Visayas	20	31	0	23	10	16	17	17	21	29	05	30	53	95	16	29	13	29	13	22	36	38	0	nd			
PHI088	Luzón	19	22	27	22	9	16	15	15	21	29	05	40	37	54	12	32	12	32	32	44	39	40	2	O3a2c*_P164			
PHI093	Visayas	18	22	28	22	9	15	15	15	21	19	15	40	36	64	12	31	12	31	22	54	38	41	3	O3a2c*_P164			

Amostras	Provincia	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	GATA-H4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	Haplogrupo e marcador final
PHI096	Luzón	18	14	20	31	21	10	18	12	20	10	10	15	20	13	38	13	16	15	13	31	13	12	23	35	1	O3*_M122
PHI098	Visayas	17	12	19	28	20	10	15	15	12	18	10	19	30	48	8	8	8	4	13	30	13	10	44	39	1	O1a*_M119
PHI099	Visayas	18	12	29	20	20	10	18	13	10	10	14	50	37	6	5	21	13	21	31	13	11	52	38	1	O3*_M122	
PHI100	Luzón	19	13	21	21	21	9	16	15	12	19	10	64	35	3	5	5	5	12	29	29	33	13	23	38	1	O3a2c*_P164
PHI101	Luzón	17	12	10	3	21	10	18	15	12	16	10	73	30	4	6	8	4	13	29	29	13	17	7	38	1	O1a*_M119
PHI106	Luzón	17	12	28	24	24	9	17	15	11	19	10	54	40	3	6	5	4	12	30	30	34	44	40	40	2	O3a2c*_P164
PHI110	Visayas	17	12	18	20	20	10	14	16	12	18	10	53	02	8	0	4	15	28	28	42	55	40	1	37	1	O1a*_M119
PHI113	Visayas	5	3	2	9	20	10	16	14	11	16	10	54	9	1	4	0	1	21	29	4	1	6	37	1	nd	
PHI120	Luzón	18	12	08	19	19	9	18	15	12	20	10	54	14	4	4	9	5	14	35	1	2	3	36	1	O3a2c1a_M117	
PHI125	Visayas	8	1	2	23	10	16	15	15	11	19	10	55	03	6	4	5	12	31	31	32	2	4	39	1	O3a2c*_P164	
PHI131	Visayas	9	2	2	22	20	16	15	15	11	19	10	54	03	6	5	5	13	30	30	32	4	4	38	1	O3a2c*_P164	
PHI134	Luzón	18	12	08	19	19	9	18	14	12	20	10	52	14	6	7	5	19	32	32	2	2	3	35	1	O3a2c1a_M117	
PHI135	Mindanao	17	13	10	21	21	11	17	17	11	16	10	52	04	6	9	4	13	29	29	32	8	8	37	1	O1a*_M119	
PHI136	Luzón	18	12	9	20	20	10	15	15	12	17	10	73	03	6	8	4	13	31	31	31	4	4	36	1	O1a*_M119	
PHI137	Luzón	16	12	18	22	22	10	17	15	11	18	10	73	03	4	8	1	12	29	29	1	1	6	37	1	O1a*_M119	
PHI139	Visayas	4	2	5	20	20	16	14	14	12	11	15	10	28	8	8	4	16	31	31	2	1	8	33	1	C_M130	
PHI140	Visayas	7	2	2	23	11	17	15	15	11	19	10	54	03	4	5	5	15	31	31	32	2	4	39	1	O3a2c*_P164	
PHI141	Luzón	18	12	18	23	23	9	17	15	11	19	10	54	03	6	6	4	17	30	30	33	3	3	39	1	O3a2c*_P164	
PHI142	Luzón	18	13	19	17	17	11	18	16	12	11	15	30	01	8	7	4	16	30	30	4	1	3	37	1	C_M130	
PHI143	Luzón	de	1	2	2	de	1	de	15	1	1	1	2	1	1	3	de	1	12	de	1	1	de	34	1	O1a*_M119	
PHI144	Mindanao	18	13	4	23	23	10	18	16	11	18	10	45	03	6	8	4	19	28	28	1	1	2	36	1	O3a1*_L127	
PHI147	Luzón	18	12	3	22	22	10	16	15	12	19	10	54	04	7	4	5	16	31	31	4	2	5	39	1	O3a2c*_P164	
PHI148	Luzón	18	12	08	19	19	10	20	15	12	11	10	54	14	5	8	5	18	33	33	2	3	3	36	1	O3a2c1a_M117	
PHI151	Luzón	20	12	2	21	21	9	20	15	11	19	10	53	02	8	7	5	16	36	36	1	3	4	35	1	O3a2c1*_M134	
PHI153	Luzón	16	12	2	21	21	10	16	15	11	19	10	54	03	7	5	5	16	29	29	3	3	4	37	1	O3a2c*_P164	
PHI155	Luzón	17	12	9	20	20	10	15	15	12	18	10	93	04	8	8	4	14	30	30	3	1	2	38	1	O1a*_M119	
PHI157	Luzón	19	14	3	22	22	11	16	14	12	19	10	54	23	8	9	5	13	29	29	3	2	2	35	1	nd	
PHI158	Luzón	18	13	1	21	21	11	15	14	11	18	10	55	03	2	7	4	19	29	29	5	2	3	36	1	O2a1*_M95	
PHI160	Visayas	5	2	1	22	22	10	16	15	12	16	10	73	04	6	7	4	14	28	28	3	1	7	39	1	O1a*_M119	
PHI165	Luzón	20	13	2	22	22	9	16	15	11	19	10	54	03	6	5	4	16	30	30	3	2	4	40	1	O3a2c*_P164	
PHI166	Luzón	17	12	18	21	21	10	18	17	11	16	10	53	04	6	7	4	14	30	30	3	1	7	38	1	O1a*_M119	
PHI169	Visayas	9	3	2	22	22	10	17	14	11	19	10	55	03	1	7	4	19	27	27	3	1	4	38	1	O2a1*_M95	
PHI170	Luzón	15	12	0	21	21	10	13	15	12	18	10	73	04	8	9	4	15	30	30	3	1	5	37	1	O1a*_M119	
PHI173	Visayas	8	2	1	22	22	10	18	15	12	10	14	30	03	9	0	4	20	32	32	1	1	6	37	1	O3a2*_P201	

Amostras	Provincia	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	GATA-H4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	Haplogrupo e marcador final
PHI176	Visayas	18	12	20	28	23	9	17	17	21	19	10	15	23	11	12	39	17	5	16	31	12	13	23	34	1	O3a2c1*_M134
PHI178	Luzón	18	13	21	29	23	10	16	17	22	19	11	15	23	11	13	32	22	6	18	24	14	10	20	37	1	nd
PHI179	Luzón	18	14	21	31	21	10	18	12	10	0	10	5	21	11	4	0	5	5	17	30	13	23	3	36	1	O3*_M122
PHI184	Luzón	16	22	22	28	20	11	18	15	20	11	10	7	24	11	3	6	7	4	14	29	33	17	2	38	1	O1a*_M119
PHI189	Luzón	20	31	23	30	21	9	20	15	21	8	0	5	21	11	3	8	8	5	11	32	12	2	4	40	1	O3a1*_L127
PHI190	Luzón	19	31	23	11	22	11	16	15	11	2	0	5	3	0	1	3	6	4	17	32	4	2	4	38	2	C_M130
PHI192	Visayas	15	22	22	29	20	10	13	15	13	7	9	5	10	2	9	3	7	4	14	28	13	3	3	36	1	O1a*_M119
PHI201	Luzón	16	21	22	19	20	10	17	15	11	8	0	7	21	11	4	8	7	4	13	32	13	2	4	37	1	O1a*_M119
PHI207	Luzón	19	23	22	39	22	9	16	15	12	9	0	5	3	1	3	7	4	5	16	30	13	2	4	39	1	O3a2c*_P164
PHI208	Luzón	18	22	22	28	22	11	16	15	11	9	0	5	4	0	3	6	6	5	15	31	13	2	4	39	1	O3a2c*_P164
PHI209	Luzón	18	31	22	29	23	11	17	15	11	9	0	6	4	0	3	3	7	4	20	30	4	2	2	37	1	O2a1*_M95
PHI210	Luzón	18	22	22	28	22	9	16	15	22	9	0	5	4	0	3	6	5	4	16	31	13	2	4	40	1	O3a2c*_P164
PHI211	Luzón	18	23	22	38	21	10	18	14	11	2	1	0	5	1	0	1	4	9	16	29	2	2	8	33	2	C_M130
PHI212	Luzón	17	22	22	18	23	9	15	15	11	9	1	5	5	0	3	8	6	4	16	31	12	2	5	39	1	O3a2c*_P164
PHI213	Luzón	18	22	22	08	20	9	19	14	22	2	1	0	5	3	1	4	7	5	18	34	12	2	3	35	1	O3a2c1a_M117
PHI214	Luzón	19	22	22	18	22	11	17	15	11	1	0	5	1	0	1	7	1	4	16	26	12	0	6	38	1	C_M130
PHI215	Luzón	18	43	23	0	23	11	17	14	21	9	1	4	3	2	3	8	7	5	13	28	13	2	3	33	1	nd
PHI216	Luzón	19	22	22	18	22	11	17	15	11	1	0	5	1	0	1	7	2	1	16	26	2	0	6	38	1	C_M130
PHI217	Luzón	19	31	22	28	20	10	19	15	22	6	0	6	4	0	3	6	8	4	15	30	12	1	6	37	1	O1a*_M119
PHI218	Visayas	18	22	22	18	21	9	16	15	21	9	0	5	4	0	4	8	5	4	15	32	13	2	4	40	1	O3a2c*_P164
PHI219	Luzón	16	22	22	28	22	10	17	17	22	6	0	7	3	0	4	8	1	4	13	30	13	1	7	37	1	O1a*_M119
PHI220	Luzón	18	34	23	1	17	12	16	16	11	2	1	1	6	5	1	1	1	9	14	32	14	0	3	37	1	nd
PHI221	Luzón	18	22	22	28	22	9	16	15	22	9	0	5	4	0	3	5	5	4	16	30	13	2	3	40	1	O3a2c*_P164
PHI222	Visayas	19	31	23	30	21	11	16	17	21	9	0	7	3	0	4	9	9	4	14	32	14	2	7	38	0	nd
PHI223	Luzón	16	33	23	9	22	11	16	15	11	3	8	1	6	3	2	4	7	0	13	30	3	2	2	35	1	nd
PHI224	Visayas	18	22	22	8	20	9	16	15	22	1	0	7	4	0	3	7	4	4	16	29	13	3	5	39	1	O3a2c*_P164
PHI225	Luzón	18	31	22	29	20	10	16	15	22	1	1	1	6	5	0	3	0	7	19	30	3	2	5	36	0	O2a1*_M95
PHI226	Luzón	18	22	22	28	22	9	13	16	22	1	9	0	5	4	0	2	8	5	15	30	12	3	4	37	1	O3a2c*_P164
PHI227	Luzón	19	31	22	29	22	10	18	16	11	1	9	9	1	4	5	0	3	6	20	32	12	2	4	36	1	O3a1*_L127
PHI228	Visayas	17	22	22	77	21	11	17	15	13	8	0	6	3	1	4	7	0	5	14	31	13	2	3	37	1	nd
PHI229	Luzón	18	22	22	18	22	9	14	15	22	9	0	5	4	0	5	7	4	6	16	31	13	2	4	39	1	O3a2c*_P164
PHI231	Visayas	17	41	23	10	22	11	15	16	11	2	1	1	2	1	1	3	2	1	12	29	13	2	3	39	1	O2*_L690
PHI232	Luzón	16	41	23	10	22	11	15	16	22	8	0	5	4	0	3	9	1	4	18	30	14	2	3	41	1	O2*_L690
PHI233	Luzón	20	20	22	08	18	9	17	14	22	2	0	0	5	5	1	4	6	7	17	31	2	4	3	34	1	O3a2c1a_M117

Amostras	Provincia	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	GATA-H4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	Haplogrupo e marcador final
PHI234	Luzón	18	12	21	29	21	9	15	15	13	19	9	14	25	9	13	37	15	15	12-19	30	12	10	25	40-41	12	O3a2c*_P164
PHI235	Luzón	19	12	21	28	22	9	16	15	12	19	0	15	4	0	3	36	14	15	10-16	31	13	16	26	39-40	13	O3a2c*_P164
PHI236	Luzón	18	14	23	30	23	10	18	14	11	19	1	15	2	0	3	8	7	5	11-14	30	13	12	23	35-36	2	nd
PHI237	Luzón	16	12	20	28	20	10	17	16	12	17	0	16	3	0	4	7	3	4	13-14	32	33	33	23	37-38	1	O1a*_M119
PHI238	Visayas	15	12	22	22	21	12	18	17	12	16	0	15	3	0	4	6	4	4	13-14	30	13	17	7	38-39	2	O1a*_M119
PHI239	Luzón	19	12	23	28	21	9	19	15	12	19	0	16	4	0	3	5	4	5	12-17	31	33	23	2	40-40	2	O3a2c*_P164
PHI240	Luzón	19	12	22	28	24	10	16	16	12	19	0	15	4	0	3	6	5	5	12-18	31	13	12	4	38-40	3	O3a2c*_P164
PHI241	Luzón	18	13	19	20	20	10	16	15	11	18	0	16	5	0	1	4	1	5	15-19	30	33	25	5	36-39	0	O2a1*_M95
PHI242	Visayas	18	12	21	28	23	9	19	15	14	0	0	15	5	0	3	0	0	4	13-21	29	13	15	2	37-37	2	O3a2*_P201
PHI243	Visayas	20	14	19	23	23	10	17	16	11	19	0	15	5	0	4	7	9	4	13-16	32	44	22	2	36-37	2	nd
PHI244	Luzón	18	12	22	28	20	9	16	15	11	19	0	17	4	0	3	6	4	5	12-16	29	33	25	5	40-40	2	O3a2c*_P164
PHI245	Luzón	19	12	21	28	20	9	13	15	11	19	0	14	4	0	3	7	6	6	12-15	30	33	15	2	39-39	1	O3a2c*_P164
PHI246	Luzón	18	13	19	23	23	9	17	15	12	19	0	15	4	0	3	8	6	4	12-16	30	33	13	3	40-41	3	O3a2c*_P164
PHI247	Luzón	14	12	23	28	21	11	16	15	12	19	0	14	4	0	3	8	7	4	15-16	27	15	05	2	37-37	1	nd
PHI248	Luzón	19	12	22	28	22	10	16	15	12	19	1	15	2	0	3	8	1	4	12-17	31	12	23	3	40-40	2	O3a2c*_P164
PHI249	Luzón	18	12	21	28	22	9	16	15	12	19	1	15	4	0	3	7	6	4	12-17	31	12	23	3	40-40	3	O3a2c*_P164
PHI250	Luzón	17	12	23	30	22	9	16	14-16	11	19	0	15	4	0	3	5	5	4	12-16	32	13	24	4	38-40	3	O3a2c*_P164
PHI251	Visayas	15	12	23	29	20	10	14	15	13	17	9	15	1	0	4	9	6	4	12-14	28	44	13	3	36-40	1	O1a*_M119
PHI252	Luzón	18	14	20	23	19	11	18	14	12	19	0	16	1	0	1	2	7	4	13-13	29	44	23	2	37-37	2	nd
PHI253	Luzón	15	12	20	28	19	9	15	16	12	17	0	17	3	0	4	6	0	4	14-14	32	13	21	1	36-37	1	O1a*_M119

Fonte: A Autora, 2019.



### APÊNDICE C - Referências das amostras do Sudeste Asiático

Localidade das amostras	Referência
Vietnã	Presente trabalho
Vietnã [Mong]	Presente trabalho
Filipinas + Literatura	Presente trabalho, (TREJAUT <i>et al.</i> , 2014); (PURPS <i>et al.</i> , 2014)
Cazaquistao_Akmola [Argyn]	(ZHABAGIN <i>et al.</i> , 2019)
Cazaquistao_Almaty: [Alban, Shaprashty, Suan, Zhalayr]	(ZHABAGIN <i>et al.</i> , 2019)
Cazaquistao_JAmbyl [Dulat]	(ZHABAGIN <i>et al.</i> , 2019)
Cazaquistao_Karaganda	(ZHABAGIN <i>et al.</i> , 2019)
Cazaquistao_Kyzylorda [Kyrgyz]	(ZHABAGIN <i>et al.</i> , 2019)
Cazaquistao_Regiao Sul [Katagan, Konyrat, Sunaki]	(ZHABAGIN <i>et al.</i> , 2019)
Paquistao_Azad Jammu e Kashmiris [Kashmiris]	(ADNAN <i>et al.</i> , 2018)
Paquistao_Punjabi [Punjabis]	(ADNAN <i>et al.</i> , 2018)
China_Guizou [Maioria YI]	(FAN <i>et al.</i> , 2019)
China_Hennan [Han]	(BAI <i>et al.</i> , 2016)
China_Shanxi [Han]	(BAI <i>et al.</i> , 2013)
China_Sichuan [maioria Yi]	(FAN <i>et al.</i> , 2019)
China_Sichuan [Chinese tibetano]	(SHAN <i>et al.</i> , 2014)
China_Xishuangbanna [Dai]	(PURPS <i>et al.</i> , 2014)
China_Xuanwei [Han]	(PURPS <i>et al.</i> , 2014)
China_Yunnan [Bai]	(PURPS <i>et al.</i> , 2014)
China_Xinjiang [Uighur]	(SHAN <i>et al.</i> , 2014)
China_Xinjiang [KAsakh]	(SHAN <i>et al.</i> , 2014)

Fonte: A Autora, 2019.

China_Yunnan [Maioria Yi]	(FAN <i>et al.</i> , 2019)
Bangladesh [Bengali]	(HASAN <i>et al.</i> , 2018)
Bangladesh_Chittagong [Rakhine, Marma e Hajong]	(HASAN <i>et al.</i> , 2018)
Tailandia	(TREJAUT <i>et al.</i> , 2014)
Tailandia [Akha]	(TREJAUT <i>et al.</i> , 2014)
Taiwan [Han Atayal]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo de montanha [Amis, Hakka e Kategalan]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo de montanha [Paiwan, Yami, Rukai]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo de montanha [Bunun]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo de montanha [Saisiyat]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo de montanha [Taroko]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo de montanha [Thao]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo de montanha [Tsou]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo simples [Pazeh]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo simples [Siraya Hwalien]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo simples [Siraya Kaohsiung]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo simples [Siraya Tainan01]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo simples [Siraya Tainan02]	(TREJAUT <i>et al.</i> , 2014)
Taiwan_Tribo simples [Siraya Tainan-Coast]	(TREJAUT <i>et al.</i> , 2014)
Taiwan [Paiwan]	(WU <i>et al.</i> , 2011)
Indonésia	(PURPS <i>et al.</i> , 2014)
Singapura [Han]	(PURPS <i>et al.</i> , 2014)
Singapura [Indian]	(PURPS <i>et al.</i> , 2014)
Singapura [Malay]	(PURPS <i>et al.</i> , 2014)

## APÊNDICE D - Análise das amostras do Continente Asiático

Y-STRs - FSTs		Vietnã_Mong	Vietnã	Cazak_Akmola	Cazak_Almaty	Cazak_JAmbyl	Cazak_Karaganda	Cazak_Kyzylorda	Cazak_Regiao Sul	Paquistao_Azad Jammu e Kashmiris_Etnia_Kashmiris	Paquistao_Punjabi_Punjabis	China_Etnia_Maioria_Yi	China_Hennan_Etnia_Han	China_Shanxi_Etnia_Han	China_Sichuan_Etnia_maioriaYi	China_Sichuan_Etnia_Chinese tibetano	China_Xishuangbanna_Etnia_Dai	China_Xuanwei_Etnia_Han	China_Yunnan_Etnia_Bai	China_Xinjiang_Etnia_Uighur	China_Xinjiang_Etnia_KAsakh	China_Yunnan_Etnia_Maioria Yi	Bangladesh_Etnia_Bengali	Bangladesh	Tailandia	Tailandia_Etnia_Akha
Vietnã_Mong	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Vietnã	0,0894	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Cazak_Akmola	0,2353	0,1763	*	0,0001	0,0003	0,0055	0,001	0,0002	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Cazak_Almaty	0,2202	0,1545	0,1814	*	<5E-6	<5E-6	0,029	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Cazak_JAmbyl	0,1839	0,144	0,1595	0,0906	*	<5E-6	0,0088	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Cazak_Karaganda	0,1147	0,1038	0,0943	0,1281	0,12	*	0,0022	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Cazak_Kyzylorda	0,2038	0,1625	0,1582	0,0451	0,0994	0,1069	*	0,0005	<5E-6	0,0008	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Cazak_Regiao Sul	0,1574	0,1028	0,0878	0,102	0,113	0,0707	0,0908	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Paquistao_Azad Jammu e Kashmiris_Etnia_Kashmiris	0,1479	0,1312	0,1269	0,1401	0,1513	0,0902	0,0984	0,0629	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Paquistao_Punjabi_Punjabis	0,1715	0,1476	0,1872	0,1253	0,1591	0,1001	0,0842	0,0743	0,039	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Etnia_Maioria_Yi	0,0839	0,0287	0,1315	0,1233	0,101	0,0768	0,1328	0,0643	0,0939	0,1032	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Hennan_Etnia_Han	0,0561	0,0461	0,1296	0,1344	0,1062	0,0774	0,1292	0,0807	0,0883	0,1089	0,0152	*	0,0003	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Shanxi_Etnia_Han	0,0518	0,0531	0,12	0,1386	0,1081	0,0662	0,1252	0,0773	0,0796	0,1051	0,0233	0,0038	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Sichuan_Etnia_maioriaYi	0,0562	0,0421	0,1613	0,1453	0,1119	0,0794	0,1361	0,0937	0,1003	0,1054	0,014	0,0123	0,0144	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Sichuan_Etnia_Chinese tibetano	0,2053	0,14	0,2109	0,156	0,1861	0,1686	0,1684	0,126	0,1396	0,1448	0,0988	0,0974	0,1065	0,1072	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Xishuangbanna_Etnia_Dai	0,1097	0,1267	0,2552	0,2671	0,2294	0,1635	0,26	0,1918	0,1673	0,1876	0,0927	0,0633	0,0654	0,0589	0,2091	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Xuanwei_Etnia_Han	0,0751	0,0674	0,175	0,1905	0,1672	0,1128	0,1915	0,133	0,1309	0,1643	0,0451	0,0243	0,0238	0,0383	0,1624	0,0557	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Yunnan_Etnia_Bai	0,1013	0,0901	0,1633	0,2005	0,1714	0,1229	0,1977	0,1342	0,1208	0,1552	0,0592	0,0353	0,0352	0,0509	0,1543	0,042	0,0221	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Xinjiang_Etnia_Uighur	0,1035	0,0945	0,1904	0,132	0,1544	0,1189	0,1606	0,1102	0,1372	0,1502	0,0709	0,0555	0,0615	0,0759	0,1427	0,161	0,0944	0,1169	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Xinjiang_Etnia_KAsakh	0,1006	0,0879	0,1144	0,0813	0,0921	0,0498	0,0543	0,0394	0,0346	0,0202	0,0477	0,0498	0,0448	0,0487	0,0995	0,1277	0,0939	0,0935	0,0856	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
China_Yunnan_Etnia_Maioria Yi	0,1429	0,061	0,1395	0,1246	0,1039	0,1026	0,1338	0,0717	0,1195	0,1202	0,0253	0,0589	0,0657	0,0519	0,1319	0,1742	0,1079	0,1171	0,113	0,0649	*	<5E-6	<5E-6	<5E-6	<5E-6	<5E-6
Bangladesh_Etnia_Bengali	0,1311	0,1142	0,1565	0,1112	0,1314	0,0901	0,0817	0,0587	0,0311	0,0157	0,0727	0,0708	0,0707	0,0747	0,0953	0,1436	0,1244	0,112	0,1034	0,0154	0,09585	*	<5E-6	<5E-6	<5E-6	<5E-6
Bangladesh	0,085	0,1142	0,2028	0,1952	0,169	0,1095	0,1691	0,1265	0,0984	0,1043	0,0729	0,0482	0,0441	0,0412	0,1399	0,0304	0,0747	0,0554	0,1232	0,068	0,13028	0,0732	*	<5E-6	<5E-6	
Tailandia	0,094	0,0343	0,2067	0,2055	0,1958	0,1249	0,2116	0,1346	0,1479	0,1729	0,0532	0,0556	0,0548	0,0561	0,1845	0,0865	0,0318	0,0541	0,1162	0,1057	0,10903	0,1371	0,1017	*	<5E-6	
Tailandia_Etnia_Akha	0,2792	0,2123	0,3393	0,3333	0,368	0,2424	0,3687	0,2305	0,2613	0,2795	0,2023	0,1984	0,1928	0,2195	0,3063	0,2651	0,1789	0,194	0,2895	0,2168	0,26433	0,242	0,2435	0,193	*	



## ANEXO – Artigos publicados

Forensic Science International: Genetics 22 (2016) e18–e20



Contents lists available at ScienceDirect

## Forensic Science International: Genetics

journal homepage: [www.elsevier.com/locate/fsig](http://www.elsevier.com/locate/fsig)

## Correspondence

## Y Chromosome STR haplotypes in different ethnic groups of Vietnam



Dear Editor,

Short tandem repeats (STRs) located at the male specific region of the Y chromosome are widely used in studies of population genetics and human identification. Because of the non-recombining nature of these markers, they can be combined in haplotypes with a high discrimination capacity between individuals inside populations. Its application to the forensic field requires the previous knowledge on population profile concerning haplotype frequencies and possible geographic and/or ethnic substructure.

In this letter we report the results of a study of 17 Y-STRs in a sample of unrelated males living in different districts of Vietnam, most belonging to the main ethnic group (Kinh) and the remaining are from other minority ethnic groups (Tay, Mong, Nung, Hoa and Thai).

The Socialist Republic of Vietnam has about 89 million inhabitants distributed in 54 official ethnic communities. The most important language of the country is the Vietnamese, which is the native language of the Kinh. The Kinh is the main ethnic community and the Tay, Thai, Hoa, Muong, Khmer and Mong (Hmong) are those with larger population sizes among the minority groups, being represented by at least 1 million people [1].

The Viet or Kinh people originated in a geographic area that is now part of northern Vietnam and southern China. According to the *Vietnam Population and Housing Census, 2009* [2] this ethnic group has a population of 73,594,427 people, accounting for 85.7% of the total population of Vietnam. The Kinh people are spread across the whole Vietnam, and in some other countries, but they are more concentrated in plains and urban areas.

The Tay is the second largest ethnic group in Vietnam, with a population of 1,626,392 people (1.895%) [2]. Most Tay people live in the Northeast provinces, along the valleys and the lower slopes of the mountains.

According to the census from 2009 [2] the Mong is the sixth largest group in Vietnam with a population of 1,068,189 people (1.244%), most living in mountainous regions, at altitudes from 800 to 1500 m above the sea level. In Vietnam, they occupy a relatively large geographical area, including most of the northern mountainous provinces bordering China and Laos, mainly in the northwestern provinces of the Northeast region.

The Nung have a close relationship with the Tay and the Choang (Zhuang). They have a population of 968,800 people (1.123%) [2], which are mainly concentrated in the northeastern province of the Tonkin region, along the border with China.

The Hoa are a group of people coming from China, which according to the classification of the People's Republic of China

belong to the Han ethnic group. In Vietnam they represent a population of 823,071 people (0.96%), which is concentrated in Ho Chi Minh City (414,045 people, accounting for 50.3% of the total Hoa population in Vietnam).

In this study, a total of 218 samples were collected from different cities, namely Hai Phong (all from Kinh ethnic group), Da Nang (Kinh group), Hanoi (most from the Kinh group) and Ho Chi Minh (Kinh and Hoa groups), as well as from rural areas in Lao Cai (Mong group) and Cao Bang (Tay, Nung and Kinh groups) provinces (see Supplementary material Fig. S1A). This study followed the ethical principles of the 2000 Helsinki Declaration of the World Medical Association (<http://www.uma.net/e/policy/b3.htm>) and written informed consent was obtained from participants for cooperation under strictly confidential conditions.

DNA was extracted by an organic method.

All samples were genotyped for the AmpFLSTR<sup>®</sup>Yfiler<sup>®</sup> PCR Amplification kit (Applied Biosystems, Foster City, CA) markers, following manufacturer's instructions. The Y-STR alleles were designated according to the ISFG recommendations [3].

The results are listed in Table S1, along with information on sampling place and ethnic group for each sample. Haplotype data are included in the YHRD (<http://www.yhrd.org>) under the following accession numbers: Vietnam [Kinh] YA003984, Vietnam [Mong] YA003985, Vietnam [Tay] YA003986, Vietnam [Nung] YA003987, Vietnam [Hoa] YA003988.

The haplotype diversity, calculated according to Nei [4], was high in all samples, except in Lao Cai district group (Table 1). In the sample of 59 unrelated individuals from Lao Cai, just 36 have unique haplotypes; the remaining haplotypes are shared among 2 to 4 individuals. In the other district groups, a single case of shared haplotypes was found for two samples from Cao Bang ( $n = 41$ ). No shared haplotypes were detected among samples from different districts. In the overall Vietnamese sample, haplotype diversity was  $0.9991 \pm 0.0006$ , with an average value of gene diversity over loci of  $0.633 \pm 0.160$ .

Based on the pairwise  $F_{ST}$  genetic distances, calculated using the software Arlequin 3.5.1.2 [5], significant differences were observed between Lao Cai and other regions, except Hai Phong, at a significance level of 0.003 (after applying the Bonferroni's correction for multiple tests) (see Fig. 1). No statistically significant differences were detected between the remaining samples. The comparison of our samples from Hanoi and Ho Chi Minh with two other samples from the same regions [6,7] did not reveal statistically significant differences as well ( $F_{ST} = 0.01093$ ;  $p = 0.09791 \pm 0.0028$  and  $F_{ST} = -0.01369$ ;  $p = 0.92832 \pm 0.0025$ , respectively).

These results could be explained by the different ethnic composition of the samples from each region. All samples from Lao Cai are from individuals that belong to the Mong ethnic group; the samples from Cao Bang is represented by Kinh, Tay and Nung

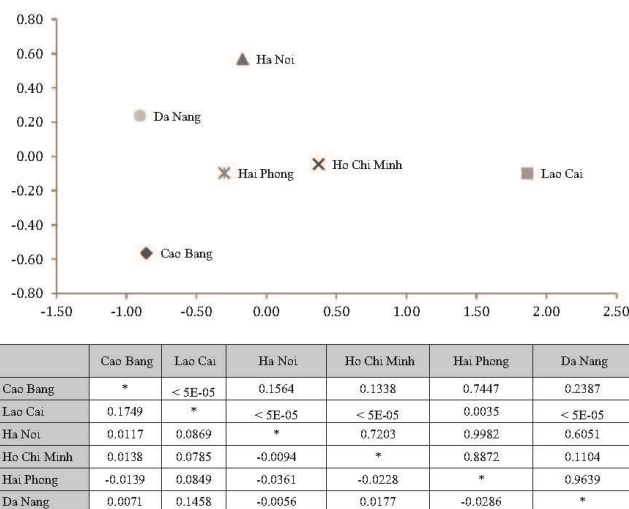
**Table 1**  
Haplotype diversity estimates for the six studied population groups and in the overall sample from Vietnam.

	Haplotype diversity
Cao Bang	0.9988 ± 0.0057
Lao Cai	0.9877 ± 0.0058
Hanoi	1.0000 ± 0.0068
Ho Chi Minh	1.0000 ± 0.0086
Hai Phong	1.0000 ± 0.0202
Da Nang	1.0000 ± 0.0065
Total	0.9991 ± 0.0006

populations, with strong signals of genetic drift and isolation from the remaining populations from South East Asia.

In summary, the results of the present study showed low heterogeneity among distant districts in Vietnam, demonstrating that a geographic criterion is not the most adequate for the structure definition of a forensic database. Meanwhile, it is important to take into account that many different ethnic groups exist in the country, which can lead to a significant differentiation of some neighbor districts.

This paper strictly followed the requirements of the guidelines for publication of population data [9,10].



**Fig. 1.** Multidimensional scaling plot of the pairwise  $F_{ST}$  genetic distances calculated between the samples from Vietnam (Stress = 0.006).

individuals; and the remaining population samples comprise mainly Kinh people (see Supplementary material Table S1). Therefore, a new analysis was performed between the three main ethnic groups represented in our sample from Vietnam: Kinh ( $n = 123$ ), Tay ( $n = 21$ ) and Mong ( $n = 59$ ). Because of the low sample size, samples from Nung ( $n = 7$ ), Hoa ( $n = 7$ ) and Thai ( $n = 1$ ) ethnic groups were not considered for this analysis. The results showed significant differences between the Kinh and Mong ( $F_{ST} = 0.10476$ ;  $p < 0.000005$ ) and between the Tay and the Mong ( $F_{ST} = 0.13912$ ;  $p < 0.000005$ ); no significant difference was found between the Kinh and the Tay ( $F_{ST} = -0.00206$ ;  $p = 0.55004 \pm 0.0054$ ).

Comparisons were also performed between Vietnam and other populations from South East and East Asian countries including China, Japan, Taiwan, Thailand, Vietnam, Malaysia and Indonesia [6–8]. Samples from the same country and ethnic group were pooled whenever non-differentiation  $p$ -values were obtained for pairwise  $F_{ST}$ s (see Supplementary material Fig. S1B for details and references on samples used for comparisons). Significant differences were found between the Mong ethnic group and all other samples (Supplementary material Table S2); and no differences were detected among the other two Vietnamese samples (Kinh and Tay) and Thailand. In the MDS plot of the pairwise genetic distances (Supplementary material Fig. S2), the Kinh and Tay group together, near to Thailand, also part of the Indochina territory. These are closer to Indonesia and the Malay from Singapore than to the East Asian populations. The Mong keep far apart from all other

#### Acknowledgement

IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT, the Portuguese Foundation for Science and Technology.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fsigen.2016.02.007>.

#### References

- [1] M.P. Lewis, F. Gary, Simons C.D. Fennig. Ethnologue: Languages of the World. State Committee for Ethnic Minority Affairs—Socialist Republic of Vietnam. (2014).
- [2] The 2009 Vietnam population and housing Census. [www.gso.gov.vn](http://www.gso.gov.vn) (2009).
- [3] L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith, P.M. Schneider. DNA commission of the international society of forensic genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int. J. Legal Med.* 120 (2006) 191–200.
- [4] N. Nei. *Molecular Evolutionary Genetics*, Columbia University Press, New York, 1987.
- [5] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10 (2010) 564–567.

- [6] J.A. Trejaut, E.S. Poloni, J.C. Yen, Y. Lai, J.H. Loo, C.L. Lee, C.L. He, M. Lin, Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia, *BMC Genet.* 77 (2014) 1–23.
- [7] J. Purps, S. Siegert, L. Roewer, A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, *Forensic Sci. Int. Genet.* 100 (2014) 12–23.
- [8] N. Mizuno, H. Nakahara, K. Sekiguchi, K. Yoshida, M. Nakano, K. Kasai, 16 Y chromosomal STR haplotypes in Japanese, *Forensic Sci. Int.* 174 (2008) 70–75.
- [9] A. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, New guidelines for the publication of genetic population data, *Forensic Sci. Int. Genet.* 7 (2013) 217–220.
- [10] A. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, Update of the guidelines for the publication of genetic population data, *Forensic Sci. Int. Genet.* 10 (2014) A1–A2.

**Fernanda Miranda-Barros**  
DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

**Carola Romanini**  
DNA Forensic Laboratory, Argentinean Forensic Anthropology Team (EAAF), Córdoba, Argentina

**Luz Adriana Pérez**  
Laboratorio de Genética Humana-ADN Antiguo, Universidad de los Andes, Bogotá, Colombia

**Nguyen Duc Nhu**  
National Institute of Forensic Medicine, Ministry of Health, Vietnam

**Tran Duc Phan**  
Biomedical and Genetics Department, Hanoi Medical University, Vietnam

**Elizeu F. Carvalho**  
DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

**Carlos Vullo**  
DNA Forensic Laboratory, Argentinean Forensic Anthropology Team (EAAF), Córdoba, Argentina

**Leonor Gusmão<sup>a,b,c,\*</sup>**  
<sup>a</sup>DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

<sup>b</sup>IPATIMUP—Institute of Molecular Pathology and Immunology of the University of Porto, Portugal

<sup>c</sup>Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

\* Corresponding author at: Laboratório de Diagnóstico por DNA, Pavilhão Haroldo Lisboa da Cunha, UERJ—Universidade do Estado do Rio de Janeiro, São Francisco Xavier, 524. Maracanã, CEP: 20550-900 Rio de Janeiro, RJ, Brazil. Fax: +55 21 23340594. E-mail address: [lgusmao@ipatimup.pt](mailto:lgusmao@ipatimup.pt) (L. Gusmão).

Received 28 July 2015  
Received in revised form 29 January 2016  
Accepted 7 February 2016  
Available online 15 February 2016



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: [www.elsevier.com/locate/FSIGSS](http://www.elsevier.com/locate/FSIGSS)

## Y-STR haplotype background of Philippines: Comparison with other Southeast Asian populations



Fernanda Miranda-Barros<sup>a</sup>, Amaya Gorostiza<sup>b,c</sup>, Elizeu F. Carvalho<sup>a</sup>,  
Antonio González-Martín<sup>b</sup>, Leonor Gusmão<sup>a,d,e,\*</sup>

<sup>a</sup> DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

<sup>b</sup> Departamento de Zoología y Antropología Física, Facultad de Ciencias Biológicas, Universidad Complutense, Madrid, Spain

<sup>c</sup> Laboratorio Identificación Genética, GENOMICA S.A.U., Grupo Zeltia, Madrid, Spain

<sup>d</sup> IPATIMUP (Institute of Pathology and Molecular Immunology from the University of Porto), Porto, Portugal

<sup>e</sup> Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal

### ARTICLE INFO

#### Article history:

Received 29 August 2015

Accepted 20 September 2015

Available online 25 September 2015

#### Keywords:

Y chromosome

Male lineages

Forensic databasing

Population genetics

### ABSTRACT

Philippines is an archipelago in the Western Pacific Ocean. It is one of the most populated country in Southeast Asia, harboring multiple ethnic groups and cultures. Most of the non-indigenous population is composed by Chinese and Spanish immigrants. In order to characterize the Y-STR composition of Philippines, in the present study, we have collected 74 samples from unrelated males from the three main geographical divisions: Luzon, Visayas and Mindanao. These samples were typed for the 27 Y-STR loci included in the YFiler Plus kit. A high Y-STR haplotype diversity was found with all haplotypes being unique in our sample. Low diversity values were found for DYS391 and DYS438 (below 34%) and the multi-loci markers DYS385 and DYS387S1 were among the most diverse ones (above 95%). Since no data is available for the full set of YFiler Plus markers, a smaller set was used to calculate genetic distances between Philippines and other Southeast Asian populations. The results showed no significant differences between our samples and two other samples from Philippines, one from the general population and the other restricted to Luzon region. Significant differences were observed when comparing any of the three Philippines samples with Vietnam and Singapore populations.

© 2015 Elsevier Ireland Ltd. All rights reserved.

### 1. Introduction

The Philippines, officially known as the Republic of the Philippines, is the seventh-most populated country in Asia with approximately 93 million inhabitants [1]. It is an archipelago in the Western Pacific Ocean distributed in 7107 islands that are categorized broadly under three main geographical divisions: Luzon, Visayas and Mindanao [2]. Luzon and Mindanao are the largest island groups representing 95% of the square miles that make up Philippines of today [3].

Despite the high geographic and cultural heterogeneity of the country, few studies exist that characterize the genetic variability of Y-chromosomal haplogroups in Philippines. Data have been recently published for haplotypes comprising the Powerplex Y23 markers in two samples, one from Luzon and another from

general population [4]. Nevertheless, until now, no data were available for the full set of Y-STRs included in the YFiler Plus kit.

The aim of this study was to characterize the genetic background of Philippines for haplotypes defined by the markers included in the YFiler Plus in order to determine the usefulness of this kit to solve forensic cases concerning this population. Population comparison analysis was also performed to investigate the Y-haplotype diversity of Philippines in the context of Southeast Asian.

### 2. Materials and methods

A total of 74 samples were selected from unrelated males living in Philippines, from the three main geographical divisions: Luzon, Visayas and Mindanao. Samples were collected with written informed consent of the donors. This study, together with the informed consent, was approved by the Ethics Committee of the University Complutense of Madrid.

The Yfiler™ Plus PCR Amplification kit (Applied Biosystems™) was used to analyze 27 Y-STR loci: DYS19, DYS385 a/b, DYS387S1

\* Corresponding author at: DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil. Fax: +55 21 23340594.

E-mail address: [lgusmao@ipatimup.pt](mailto:lgusmao@ipatimup.pt) (L. Gusmão).

<http://dx.doi.org/10.1016/j.fsigs.2015.09.170>

1875-1768/© 2015 Elsevier Ireland Ltd. All rights reserved.



**Table 1**

Matrix showing the pairwise  $F_{ST}$  and  $R_{ST}$  values (below diagonal) between the sample from Philippines included in this study and other Southeast Asian samples from Philippines, Singapore and Vietnam (data from Purps et al. [4]). The corresponding non-differentiation  $p$ -values are indicated in the above diagonal.

$F_{ST}$ results					
	Philippines (this study)	Vietnam	Philippines Luzon	Philippines total	Malay Singapore
Philippines_this study	*	<5E-06	0.1009	0.2796	<5E-06
Vietnam	0.03406	*	<5E-06	<5E-06	<5E-06
Philippines_Luzon	0.00237	0.04384	*	0.6903	<5E-06
Philippines_total	0.00117	0.04167	-0.00053	*	<5E-06
Malay_Singapore	0.01808	0.02777	0.02706	0.02582	*
$R_{ST}$ results					
	Philippines (this study)	Vietnam	Philippines Luzon	Philippines total	Malay Singapore
Philippines_this study	*	0.00050	0.10336	0.19176	0.00158
Vietnam	0.04040	*	<5E-6	<5E-6	<5E-6
Philippines_Luzon	0.00458	0.05092	*	0.22909	<5E-6
Philippines_total	0.00334	0.05085	0.00109	*	<5E-6
Malay_Singapore	0.01523	0.02952	0.02755	0.02099	*

Note:  $p$ -values were calculated after 10,100 permutations (s.e. < 0.0042).

a/b, DYS389 I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS449, DYS456, DYS458, DYS460, DYS481, DYS518, DYS533, DYS570, DYS576, DYS627, DYS635 (Y GATA C4) and YGATA H4. PCR was carried out following manufacturer's recommendations. ABI 3130 and ABI 3500 Genetic Analysers (Applied Biosystems) were used for genotyping and analysis was performed with the GeneMapper® ID Software v4.0 (Applied Biosystems). The Y-STR alleles were designated according to the ISFG recommendations [5].

Population comparisons, haplotype frequencies and genetic diversities according to Nei [6] were estimated with the software Arlequin 3.5.1.2 [7]. In population comparisons, the number of repetitions in DYS389I was subtracted from DYS389II and DYS385 was excluded. To match Arlequin STR data format, duplicated, null and non-consensus alleles were coded as missing data.

Population pairwise comparisons were conducted based on  $F_{ST}$  and  $R_{ST}$  genetic distances and the significance was tested with 10,000 permutations.

### 3. Results and discussion

In our sample of 74 unrelated individuals from Philippines, a high Y-STR haplotype diversity was found ( $1.0000 \pm 0.0022$ ) with all haplotypes being unique. Access to the full haplotype data can be obtained upon request to the first author, Fernanda Miranda (miranda.fernandam@gmail.com). Low diversity levels were found for DYS391 and DYS438 (below 34%) and the multi-loci markers DYS385 and DYS387 were among the most diverse ones (above 95%).

Since no data is available for the full set of Yfiler Plus markers, a smaller set of Y-STRs (those included in the Powerplex Y23) was used to calculate  $F_{ST}$  and  $R_{ST}$  genetic distances between Philippines and other populations from Southeast Asia [4]. The results showed no significant differences between our samples and two other samples from Philippines, one from the general population ( $n=161$ ) and the other restricted to Luzon region ( $n=629$ ) (Table 1). No population stratification was, therefore, observed, supporting the addition of the full Y-23Plex haplotype data available from Philippines to a single database.

In the comparison with other Southern Asian samples, the three samples from Philippines showed statistically significant differences when compared with both Vietnam ( $n=45$ ) and Singapore ( $n=303$ ) populations, the only Southeast populations with available data for the studied markers [4]. Although showing

statistically significant non-differentiation  $p$ -values, both  $F_{ST}$  and  $R_{ST}$  genetic distances between Philippines and Singapore were lower than between Philippines and Vietnam (Table 1).

### 4. Conclusions

In forensic genetics, it is important to accurately estimate the frequency of relevant profiles in reference populations, taking into account population boundaries and possible effects of sub-structure. In this study, we have reported the haplotype distribution for the recently released Yfiler Plus kit, in a population sample from Philippines. A Y-STR haplotype based analysis was performed showing significant differences between Philippines and two other populations from Southeast Asia, namely Vietnam and Singapore.

### Conflict of interest

None.

### Acknowledgements

We would like to acknowledge Life Technologies Corporation for providing the Yfiler Plus Kit and analysis support. Financial support was granted by Ministerio de Economía y Competitividad (España) through project: Atlas topogenético de las islas Filipinas (HAR2010-21063); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and DNA Program—State University and Justice Court of Rio de Janeiro, Brazil. IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT, the Portuguese Foundation for Science and Technology.

### References

- [1] The Philippines Population and Housing Census [Internet], 2010.
- [2] IBP Inc, Philippines Labor Laws and Regulations Handbook—Strategic Information and Basic Laws, vol. 1, International Business Publications, USA, 2015.
- [3] A. Giraldez, The Age of Trade: The Manila Galleons and the Dawn of the Global Economy, Rowman & Littlefield Publishers, 2015.
- [4] J. Purps, S. Siebert, S. Willuweit, et al., A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, Forensic Sci. Int. Genet. 12 (2014) 12–23.
- [5] L. Gusmão, J.M. Butler, A. Carracedo, et al., DNA commission of the International Society for Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, Int. J. Leg. Med. 120 (2006) 191–200.
- [6] M. Nei, Molecular Evolutionary Genetics, Columbia University Press, New York, 1987.
- [7] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.