

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

**ANÁLISE GEO-ESPACIAL DE SOBRENOMES EM GENÉTICA MÉDICA
POPULACIONAL NO BRASIL**

Marcelo Zagonel de Oliveira

Tese de doutorado submetida ao Programa de
Pós-Graduação em Genética e Biologia
Molecular da UFRGS.

Orientadora: Profa. Dra. Lavinia Schüler-Faccini

Porto Alegre, outubro de 2013.

Sumário

| | |
|--|----|
| LISTA DE ABREVIATURAS..... | 5 |
| RESUMO | 6 |
| ABSTRACT | 8 |
| CAPITULO 1 - INTRODUÇÃO | 10 |
| 1.1. GENÉTICA MÉDICA POPULACIONAL..... | 10 |
| 1.2 MIGRAÇÕES E DERIVA GENÉTICA..... | 11 |
| 1.2.1 Migrações..... | 11 |
| 1.2.2 Deriva genética..... | 12 |
| 1.3. SOBRENOMES NO ESTUDO DE POPULAÇÕES | 13 |
| 1.3.1. A utilização de sobrenomes na genética de populações | 13 |
| 1.3.2 Sobrenomes em estudos de estrutura populacional | 14 |
| 1.4 IMIGRAÇÃO ALEMÃ NO RIO GRANDE DO SUL | 17 |
| 1.4.1 Iniciativa privada na colonização do Rio Grande do Sul..... | 18 |
| 1.4.2 Colonização da região noroeste..... | 19 |
| 1.5 O MUNICÍPIO DE CANDIDO GODOI E A QUESTÃO DA GEMELARIDADE | 20 |
| CAPITULO 2 – JUSTIFICATIVA E HIPOTESE DE TRABALHO..... | 22 |
| CAPITULO 3 - OBJETIVOS..... | 23 |
| 3.1 Objetivo Geral | 23 |
| 3.2 Objetivos específicos..... | 23 |
| CAPITULO 4 – Artigo 1..... | 24 |
| So Close, So Far Away: Analysis of Surnames in a Town of Twins (Cândido Godói, Brazil). Annals of Human Genetics (2013) | 24 |
| CAPITULO 5 – Artigo 2..... | 36 |
| Histórias Cruzadas: Análise Histórica, Imigração e Distribuição Espacial de Sobrenomes no Rio Grande do Sul. (manuscrito em preparação)..... | 36 |
| CAPITULO 6 – DISCUSSÃO | 63 |
| A utilização de sobrenomes e geoprocessamento como indicadores de estrutura populacional | 63 |
| CAPITULO 7 - REFERÊNCIAS BIBLIOGRÁFICAS | 73 |
| CAPITULO 8 - ANEXOS | 79 |
| Anexo 1 - HISTORIA DA COLONIZAÇÃO ALEMÃ NO RIO GRANDE DO SUL..... | 79 |

| | |
|--|-----|
| Anexo 2 – Artigo produzido em colaboração - Genetic, Ancestry, Phenotypic diversity and Self-Perception of Ancestry in 6,938 Latin Americans (Aceito em Plos genetics)..... | 89 |
| Anexo 3 – Artigo produzido em colaboração - Twin Town in South Brazil: A Nazi’s Experiment or a Genetic Founder Effect? (Plos One, 2011) | 133 |
| Anexo 4 – Artigo produzido em colaboração - Li-Fraumeni and Li-Fraumeni-Like Syndrome among children diagnosed with pediatric cancer in Southern Brazil (Abstract, aceito em Cancer, 2013) | 141 |
| Anexo 5 – Artigo produzido em colaboração - Prevalence of the <i>TP53</i> p.R337H mutation in breast cancer patients in Brazil (abstract, submetido ao Journal of Clinical Oncology)..... | 142 |

Agradecimentos

Este trabalho não é resultado apenas de um esforço individual. Ele nasce de significativas contribuições que recolhi durante minha trajetória profissional, acadêmica e como cidadão, ao lidar com pessoas que foram fundamentais a essa construção.

Consciente de que é impossível listar todos que de uma forma ou de outra me acrescentaram conhecimentos e experiências essenciais à forma de ver o mundo e nele atuar, preciso expressar meu agradecimento por ter convivido e aprendido com pessoas como minha orientadora prof. Lavinia pela confiança depositada em mim, sempre acreditando e incentivando o nascimento de novas ideias. Aos meus pais Miriam e Leodato por sempre acreditarem que posso ser muito mais do que sou agora. A Debora Barbosa Bauermann pela oportunidade de poder compartilhar esta vida. A Virginia Ramallo pelas exaustivas revisões e orientações fundamentais para consolidação deste trabalho, sempre será um exemplo a ser seguido em minha vida profissional. Ao Nelson Jurandi da Rosa Fagundes pelo pensamento revolucionário, complexo e profundo da natureza.

De forma coletiva, preciso registrar a contribuição da equipe de trabalho ao longo deste tempo, com quem tive a oportunidade de unir uma convivência agradável, companheira e produtiva. Especialmente, agradeço ao Elmo por sua disponibilidade e competência, sempre disposto a ajudar.

LISTA DE ABREVIATURAS

GEMEPO - Genética Médica Populacional

IBGE – Instituto Brasileiro de Geografia e Estatística

IDH - Índice de Desenvolvimento Humano

INAGEMP - Instituto Nacional de Genética Médica Populacional

INCT – Instituto Nacional de Ciência e Tecnologia

LASERCA - Laboratório de Sensoriamento Remoto e Cartografia Digital

TRE - Tribunal Regional Eleitoral

PNUD - Programa das Nações Unidas para o Desenvolvimento

UFRGS – Universidade Federal do Rio Grande do Sul

UNISINOS - Universidade do Vale do Rio dos Sinos

Os sobrenomes são traços culturais transmitidos a partir de um ancestral a seus descendentes através de um mecanismo vertical, formando um sistema de herança único de nossa espécie. As bases de dados de sobrenomes conferem a vantagem de dispor de grandes amostras cobrindo grandes grupos populacionais, com rápida manipulação e o baixo custo nas análises.

A análise da distribuição espacial de sobrenomes nas mais diversas escalas geográficas (distritos, município, estado, país ou continente) pode servir como indicadores da estrutura de populações. Por outro lado, a análise geo-espacial ao longo do território ou em diferentes momentos históricos, contribui significativamente para as pesquisas em genética e saúde, pois permite ter uma dimensão espacial de variáveis em estudo.

A presente tese teve como objetivo utilizar a análise de sobrenomes e geo-espacial para fazer inferências sobre a estrutura populacional de comunidades no Brasil. Nossa proposta é de que estudos deste tipo podem ser importantes ferramentas auxiliares no campo de genética médica populacional.

Utilizamos como modelo municípios da região noroeste do estado do Rio Grande do Sul (Cerro Largo, Santo Cristo e Cândido Godói), especialmente o município de Candido Godoi, onde historicamente se conhece sua colonização por descendentes de imigrantes europeus, principalmente alemães. Nesta mesma região também foram desenvolvidos estudos em genética médica populacional fatores relacionados à ocorrência de altas taxas de gemelaridade ao longo do tempo.

Utilizamos a análise de isonimia e distribuição sobrenomes para identificar subgrupos populacionais dentro de 5.316 habitantes e 665 sobrenomes diferentes em Candido Godoi. Os índices de diversidade de sobrenomes de Fisher (α), e de consanguinidade apresentaram valores baixos e altos, respectivamente, o que está de acordo com o esperado para comunidades isoladas com altos níveis de deriva genética. De uma maneira geral, a análise de sobrenomes forneceu uma

estreita e adequada aproximação de fundo histórico e socioeconômico descrito para esta comunidade.

Grande parte das famílias com histórico de gêmeos morava inicialmente em Santo Cristo e Cerro Largo, antes de se fixar residência em Candido Godoi. Desta forma, o presente trabalho analisa a distribuição de sobrenomes também nesses dois municípios geograficamente vizinhos, antigas colônias que passaram pelo mesmo processo de criação. Baseado em análise de heredogramas, foram selecionados os dezoito nomes de família com maior frequência de nascimentos gemelares na cidade de Cândido Godói. Para cada sobrenome foi associada uma origem geográfica mais provável. Com o objetivo de verificar se as mesmas variáveis históricas e culturais que determinaram agrupamentos populacionais em Cândido Godói agem também no contexto geral dos três municípios e no estado do Rio Grande do Sul, foram empregadas duas ferramentas geoestatísticas: auto correlação espacial e a técnica de Krigagem.

Os resultados mostram uma clara diferença entre os três municípios da região, sendo Santo Cristo e Candido Godói formando um grupo com maior grau de isolamento, onde os índices de consanguinidade são maiores e Cerro Largo com menor grau de isolamento. Os sobrenomes das famílias que possuem registro de gêmeos apresentaram uma distribuição muito restrita ao nível do estado e também dentro dos municípios estudados. A explicação provavelmente está relacionada ao processo formador destas colônias.

A comparação do dados de nosso estudo com outros usando análise isonímica na América Latina, mostra diferenças entre os resultados obtidos que podem ser interpretados por diferentes processos históricos de colonização, com fortes implicações na estruturação genética das populações.

Surnames are cultural traits transmitted from ancestors to their descendants through a vertical mechanism, forming a heritage system unique to our species. An advantage of surname databases is that they possess large samples covering large population groups, which can be quickly manipulated, and the analyses can be performed at a low cost.

Spatial distribution analysis of surnames at various geographical levels (district, municipal, state, country, or continent) can serve as an indicator of population structure. Moreover, geospatial analysis throughout the territory or at different historical moments, significantly contributes to health and genetics research, since it enables a spatial dimension of the variables being studied.

This thesis aimed to use surname and geospatial analyses to make inferences about the population structure of communities in Brazil. Our proposition is that such studies can be important tools in the field of populational medical genetics.

Municipalities of the northeastern region of the state of Rio Grande do Sul (Cerro Largo, Santo Cristo, and Cândido Godói) were used as a model — in particular the municipality of Candido Godoi which is historically known for its colonization by descendants of European immigrants, especially Germans. In the same region, studies of populational medical genetics factors related to the occurrence of high rates of gemelarity over time were also developed.

We used isonymic analysis and distribution of surnames to identify population subgroups for the 5,316 inhabitants and 665 different surnames in Candido Godoi. The rate of surname diversity according to Fisher's (α), and the rate for consanguinity, had low and high values, respectively — this is consistent with what is expected for isolated communities with high levels of genetic drift. In general, the analysis of surnames provided a close approximation of the historical and socioeconomic background described for this community.

Most of the families with a history of twins originally lived in Santo Cristo and Cerro Largo, before taking up residence in Candido Godoi. Thus, this paper also

analyzes the distribution of surnames in these two geographically neighboring municipalities, both former colonies that went through the same creation process. Based on heredogram analysis, the eighteen surnames with the highest frequency of twin births in the city of Cândido Godói were selected and a most likely geographical origin was associated to each surname. In order to verify if the same historical and cultural variables that determined populational groupings in Cândido Godói also act in the general context of the three municipalities and in the state of Rio Grande do Sul, two geostatistical tools were employed: spatial self-correlation and the Kriging technique. The results of this analysis show a clear difference between the three municipalities of the region, with Santo Cristo and Cândido Godói forming a group with a higher degree of isolation, where the consanguinity rates are higher; and Cerro Largo, with a lower degree of isolation. The surnames of the families who have a record of twins showed a very limited distribution at the state level and also within the municipalities studied. The explanation is probably related to the formation process of these colonies.

The comparison, using isonymic analysis, of the data from our study with other data in Latin America, shows differences in the results obtained. This can be interpreted by different historical colonization processes, with large implications for the genetic structuring of the populations.

1.1. GENÉTICA MÉDICA POPULACIONAL

A Genética Médica Populacional (GEMEPO) é uma área da Genética Médica na qual o objeto de estudo e atenção médica é uma população (Castilla, 2005). A GEMEPO conjuga a Genética Médica com a Genética de Populações Humanas, fundamentando ações de saúde preventivas.

Historicamente, populações isoladas com alta frequência de doenças monogênicas, a maioria recessivas, vêm sendo estudadas em diversas partes do mundo. Um exemplo é o estudo da comunidade Amish nos Estados Unidos, iniciado por Victor McKusick (McKusick et al., 1964; McKusick, 1973) na década de 1960 e que até hoje vem sendo investigada (Payne et al., 2011). Os Amish, junto com os Menonitas e os Huteritas fazem parte do grupo religioso conhecido como anabatistas, que apareceram na Europa, na época da Reforma Protestante. Considerados hereges pelos católicos, eles fogem para a América do Norte, durante o século XVI onde se mantêm socialmente isolados (Francomano et al., 2003)

Fenômenos de deriva genética como o efeito fundador devido a sucessivas migrações aliados a um elevado endocruzamento favoreceram o aumento da prevalência de diversas doenças mendelianas e multifatoriais nessas comunidades. Por exemplo, a cistinose nefropática é um erro inato de metabolismo muito raro, com uma prevalência de 1/100.000 crianças no mundo e uma frequência de heterozigotos de 1/150. Comparativamente, entre os Amish em Ontario, Canadá, a prevalência de heterozigotos é de 1/6, ou seja, 25 vezes maior que a população em geral. (Payne et al., 2011)

Outro grupo muito bem estudado, que também apresenta uma história de migrações, com efeitos gargalo de garrafa e de fundador, e com certo isolamento cultural, são as comunidades judaicas derivadas das diversas diásporas. Diferentes mutações mendelianas com alta prevalência foram identificadas em grupos populacionais judeus, incluindo doenças lisossômicas de depósito (doenças de Tay-Sachs, Niemann-Pick, Gaucher e mucopolidose tipo IV),

glicogenoses, doenças de coagulação, defeitos de síntese de hormônios esteroides e genes relacionados a susceptibilidade a tumores (entre eles os BRCA1 e 2 relacionados a câncer de mama e ovário) (Ostrer, 2013).

Em Israel, Joel Zlotogora tem estudado há diversos anos doenças genéticas tanto em judeus como em não judeus (muçulmanos, árabes-cristãos e drusos) que vivem em Israel, sendo o responsável pelo “Banco de dados genéticos de Israel” - “Israeli National Genetic Database” (Zlotogora et al., 2009). No Brasil, os trabalhos de Salzano e Freire-Maia (1974) são pioneiros em investigar populações isoladas e índices de consanguinidade no Brasil.

Em 2009 foi criado o Instituto Nacional de Ciência e Tecnologia (INCT) em Genética Médica Populacional (INaGeMP – www.inagemp.bio.br). O programa do INaGeMP inclui ações coordenadas das equipes das instituições associadas e colaboradoras em pesquisa, prevenção, diagnóstico e tratamento de doenças com fortes componentes genéticos na sua etiologia, assim como formação de recursos humanos na área e ações de transferência de conhecimento para a sociedade através da divulgação científica e educação das comunidades afetadas.

1.2 MIGRAÇÕES E DERIVA GENÉTICA

1.2.1 Migrações

Migrar faz parte da história evolutiva humana. Trata-se de um fenômeno complexo, podendo consistir tanto um movimento de população dentro de uma comunidade e uma curta distância, quanto um movimento transoceânico.

A migração é um fenômeno variado e multiforme, tendo a mobilidade como ponto comum e a busca de melhores condições de sobrevivência por finalidade última. Um movimento migratório aparentemente homogêneo pode apresentar, motivações diversas, tais como econômicas, sociais, políticas, religiosas ou demográficas. “Migrar corresponde, portanto, a um processo de desajustamento-ajustamento que se traduz na recompensa subjetiva que proporciona. Objetivamente, migrar suprime a inadequação entre o sujeito ou grupo e a

sociedade de origem e restaura, nesta, o equilíbrio interno” (Martins, 1973). Assim, o imigrante vive numa constante dialética de expulsão e atração (Sayad, 1998).

As migrações têm um efeito evolutivo porque são capazes de promover alterações das frequências alélicas tanto nas populações da qual se originam os imigrantes, quanto naquelas que os recebem. De fato, se de uma dada população emigrar um grupo que não é uma amostra representativa dessa população em certas frequências alélicas, esse grupo emigrante poderá, dependendo do seu tamanho e da intensidade da corrente migratória, passar a sofrer as perdas gênicas provocadas pela emigração. Por outro lado, se o grupo que emigra desta população se estabelecer no território ocupado por outra população e diferir desta última quanto à frequência de certos alelos, tal situação também provocará uma alteração da frequência desses genes na população que recebe os imigrantes.

1.2.2 Deriva genética

A deriva genética é um processo natural, uma flutuação aleatória na frequência dos diferentes alelos presente numa população. Quando o tamanho efetivo da população é pequeno (como no caso das tribos indígenas e isolados genéticos) ou diminui muito por epidemias ou guerra (falamos, então, de um gargalo populacional), podem ser observadas variações importantes nas frequências alélicas de uma geração para outra por causa da deriva genética.

Populações isoladas ou parcialmente isoladas tendem a apresentar uma menor variabilidade genética do que outras não isoladas devido a ausência ou redução de fluxo gênico. Tais populações apresentam tamanho populacional reduzido e as frequências alélicas estão mais sujeitas à ação da deriva, podendo levar ao aumento ou a diminuição da frequência de determinados alelos (Futuyma, 1998). Nessas pequenas subpopulações, a variação aleatória pode provocar a eliminação ou, ao contrário, a fixação casual de um alelo, independentemente de seu coeficiente seletivo. Eventualmente, o resultado final da deriva genética é a mudança na frequência de um alelo desde 1 (fixado) até 0 (eliminado). Uma população separada a partir de outra maior também pode sofrer deriva genética aleatória, podendo apresentar frequências alélicas diferentes da população original (Freeman & Herron, 2009)

Um caso particular da deriva genética é o chamado efeito fundador, definido como o estabelecimento de uma nova população por meio de alguns poucos indivíduos que carregam consigo somente uma pequena fração da variação genética total da população parental (Ridley, 2003). Por exemplo, na maioria das populações no mundo é registrada alta variabilidade nos alelos dos grupos sanguíneos do sistema ABO, enquanto que na maioria das populações nativas americanas encontra-se quase exclusivamente o grupo O. A predominância deste grupo em nativos americanos é provavelmente consequência do efeito fundador, ocorrido durante o processo de diferenciação de linhagens asiáticas que deram origem à população fundadora do continente. (Estrada-Mena et al., 2010).

1.3. SOBRENOMES NO ESTUDO DE POPULAÇÕES

1.3.1. A utilização de sobrenomes na genética de populações

Os sobrenomes são traços culturais transmitidos a partir de um ancestral a seus descendentes através de um mecanismo vertical, formando um sistema de herança único de nossa espécie. Foi estabelecido desde a Idade Média na Europa, tornando-se ao longo do tempo um sistema mundial para relacionar pessoas aparentadas. Porém, as correlações entre famílias e sobrenomes podem não ser sempre diretas, já que o modo de escrita dos nomes pode sofrer mudanças e mutações, erros ao acaso que levam a emergência de novas variações. Dadas estas características, os sobrenomes podem ser comparados a um sistema genético com múltiplos alelos neutros (Guglielmino et al., 1991). Todos os indivíduos recebem um sobrenome e podem transmiti-lo aos seus descendentes, constituindo assim um recurso potencial na análise de diversidade humana. (Bedoya et al, 2006).

Pelo fato de não serem variáveis estritamente biológicas, os sobrenomes, podem sofrer mudanças com frequências desconhecidas ou estar fortemente influenciados por práticas culturais ou acontecimentos históricos. Representam uma medida relativa e não absoluta da afinidade biológica, tanto inter como intra-populacional (Dipierri, 2013). Com a metodologia estatística adequada, estas

restrições podem ser superadas, tornando a análise de sobrenomes um recurso metodológico robusto, podendo contar com grandes amostragens e permitindo uma visão global do comportamento demográfico ao longo de gerações. Esta abordagem global nem sempre é possível de realizar em estudos com marcadores genéticos, que são mais caros e que apresentam dificuldade de inclusão de um número significativo de indivíduos (Pinto & Guerra, 1988).

As bases de dados de sobrenomes conferem a vantagem de dispor de grandes amostras, cobrindo grandes grupos populacionais, com rápida manipulação e o baixo custo nas análises. Os sobrenomes podem ser obtidos nas mais variadas fontes, como os registros civis de casamentos, nascimentos e óbitos, documentos de propriedade, lista de contribuintes; dados paroquiais de casamentos, nascimentos e óbitos, e outras de procedência diversa com, por exemplo, listas telefônicas, registros eleitorais ou bancos de sangue (Guerra, 1987).

No Brasil há trabalhos pioneiros em pesquisa interdisciplinar envolvendo sobrenomes. Destacam-se os estudos de Freire-Maia (1952, 1958) com registros eclesiásticos dos estados de São Paulo e Minas Gerais para estimar a frequência de matrimônios consanguíneos, abrangendo várias gerações ao longo do tempo, entre 1838 até 1950. A informação proporcionada pelos arquivos históricos e demográficos tem se apresentado como fundamental no estudo da dinâmica populacional, permitindo conhecer o contexto histórico do seu desenvolvimento e sustentando estudos de estrutura genética, quer em grandes populações (Barbujani et al 1996; Barrai et al. 2002; Manni & Barrai 2001; Manni et al. 2004, 2005, 2008) e em pequenas ou isoladas (Blanco Villegas et al 2004; Boattini et al. 2006, 2007; Caravello & Tasso 1999, 2002; Esparza et al. 2006).

1.3.2 Sobrenomes em estudos de estrutura populacional

Todas as comunidades humanas possuem uma estrutura ou organização interna, aceitando ou rejeitando certos tipos de casamentos, seja por um número limitado de indivíduos ou por fatores culturais muitas vezes associados a um conjunto de valores particulares. A análise e distribuição de sobrenomes, o chamado método isonímico, pode ser uma fonte de informações quantitativas

sobre a estrutura genética da população. Este método, desenvolvido por Crow e Mange (1965), permite calcular o coeficiente de consanguinidade para grandes populações.

Desde então, diversos estudos utilizando distribuição de sobrenomes mostraram ser amplamente aplicáveis, constituindo-se em indicadores úteis na abordagem de estrutura populacional ou no estudo de fenômenos como migração, miscigenação ou isolamento (Jobling, 2001; King & Jobling, 2009; Graf et al, 2010) em vários níveis: comunal, regional, nacional ou continental (Dipierri et al, 2005, 2007;. Scapoli et al.2006; Tarskaia et al 2008; Bronberg et al, 2009, Chesire et al 2010). Os sobrenomes podem ser abordados até em grandes conjuntos populacionais, imersos em complexos contextos culturais. Especialmente interessante resulta a pesquisa isonímica na China, um caso paradigmático de país multiétnico (Liu et al., 2012). Sobre uma base de dados de 1,28 bilhões de pessoas, destaca-se a pouca variabilidade dos sobrenomes, os 100 mais frequentes cobrem 85% da população. Os autores acharam que regiões com altos valores de isonímia são coincidentes com a distribuição geográfica das minorias étnicas e que as distâncias calculadas inter-localidades em base aos sobrenomes estão fortemente correlacionadas a distância espacial entre elas.

Nas sociedades onde só o sobrenome paterno é herdado para todos os filhos, diversos estudos isonímicos pesquisaram a correspondência entre a linhagem dos sobrenomes e a linhagem do cromossomo Y (Qamar et al, 2002; McEvoy & Bradley, 2006; King et al. 2006; Mesa et al., 2010). Embora a associação entre sobrenomes e haplótipos possa ser confundida por eventos de filiação incorreta ou desconhecida, mutações tanto à nível molecular como na escrita, e adoção do mesmo sobrenome por vários fundadores (especialmente em populações com grande porcentagem de descendentes de escravos ou populações nativas, que receberam seus sobrenomes na hora do batismo), estas pesquisas fornecem informação muito útil quando grandes conjuntos populacionais são estudados. Seguindo esta linha de trabalho, King e Jobling (2009) acharam uma diferença significativa entre as linhagens masculinas na Inglaterra e na Irlanda, estando estas últimas mais afetadas pela deriva genética.

Os autores concluem que a abordagem combinada de sobrenomes e informação genética em grandes grupos permite atingir ancestralidades comuns nas últimas 20 gerações e teriam, portanto, grande relevância epidemiológica: Em termos de co-ancestralidade, a isonimia combinada com a pesquisa genética cobre a distancia entre o “*pedigree*” tradicional e a população mesma (McQuillan et al. 2008), tendo o potencial para atuar como um recurso auxiliar na identificação e monitoramento de doenças hereditárias. Recentemente, em forma muito interessante, surgiram estudos que fazem o caminho em sentido contrario, desde os haplótipos do cromossomo Y (conteúdos em grandes bases genéticas publicas) relacionando com outras informações como idade do voluntário e local de coleta, ate encontrar o sobrenome correspondente, permitindo a identificação pessoal (Gymrek et al., 2013).

Sobre as pesquisas no nosso país, a utilidade dos sobrenomes nos estudos de populações humanas não foi limitada só a analogia com os alelos de um sistema genético. Foram empregados para estimar miscigenação (Azevêdo 1980; Azevêdo et al., 1982), reconstruir rotas migratórias (Azevêdo et al., 1981), estudar associação com sistemas genéticos, tanto com grupos sanguíneos ABO (Tavares-Neto & Azêvedo, 1978), como com variabilidade ao nível do cromossomo Y (Barbosa et al., 1997), com marcadores autossômicos de tipo microssatélites (Barbosa et al, 2006) ou estudados como marcadores em si próprios, abordando a interessante relação entre os sobrenomes e diversos fenômenos sociais.

Dado que os nomes de família estão tão intimamente associados com crenças, religião e tradições são também uma ferramenta para reconstruir a historia cultural. Pesquisas com sobrenomes no estado de Bahia mostraram que representam um ponto de convergência biológico e cultural, tanto desde a gênese destas populações quanto ao longo das mudanças por elas experimentadas. No Noroeste do Brasil existem altas frequências de sobrenomes chamados devocionais, derivados da liturgia católica (nomes de santos, de festividades ou símbolos religiosos). As pessoas que os compartilham receberam estes sobrenomes no batismo. Elas não têm um ancestral biológico comum, mas sim uma historia comum, tratando-se de população escrava que foi logo liberada ou de

população nativa americana, que era registrada na sociedade colonial brasileira com um novo nome em português (Azevêdo et al., 1969).

Como antecedente no Rio Grande do Sul, podemos citar o trabalho de Dornelles et al. (1999), que combinaram dados de sobrenomes junto com pesquisa de variabilidade genética em 17 sistemas de proteínas e dados demográficos como local de nascimentos ou nacionalidade dos avós, analisando as distribuições nas sete mesorregiões do estado e parte de Santa Catarina. Estes pesquisadores encontraram que as populações que hoje habitam estes diferentes locais possuíam geralmente uma composição ancestral homogênea, sendo os sobrenomes portugueses os mais frequentes, seguidos de uma concentração de sobrenomes italianos na região noroeste e de sobrenomes alemães na região oeste, registrando miscigenação com população de origem africana (descendentes de escravos da região sudeste) e nativa americana em menor porcentagem.

Como visto, os sobrenomes compõem um excelente material para estudos biodemográficos e para analisar a estrutura genética de uma população. Estudos completos, incluindo a análise geo-espacial da distribuição ao longo do território ou em diferentes momentos histórico, contribuem significativamente para as pesquisas em Genética Médica Populacional. Permite ter uma dimensão espacial de todas as variáveis sob análise e fazer com os devidos embasamentos, planos de projetos futuros em articulação com a população local.

1.4 IMIGRAÇÃO ALEMÃ NO RIO GRANDE DO SUL

A primeira fase da imigração alemã no Rio Grande do Sul começa ano de 1824, período em que o governo imperial cria a colônia alemã de São Leopoldo, berço dessa imigração no Rio Grande do Sul.

De 1848 a 1874 o governo imperial institui diretrizes para o processo de colonização, e estabelece o primeiro estatuto que orientava a imigração, sendo esta constituída de indivíduos de origem germânica e visava à ocupação de territórios ainda não ocupados, balizando a encosta do planalto.

A partir de 1889, estendendo-se até 1914, o governo do estado, agora incumbido da colonização, dá-lhe novo estatuto e impulso e novas colônias são instaladas no planalto onde prosperaram rapidamente.

Os processos migratórios no país ocorreram de modo heterogêneo, cabem aqui os exemplos dos casos paulista e do sul do Brasil. No primeiro, os imigrantes substituíram gradativamente o braço escravo nos cafezais, inicialmente pelo sistema de parceria, e posteriormente como trabalhadores assalariados. Já no sul do país, tornaram-se proprietários de terras logo após a chegada.

No Rio Grande do Sul, o governo passa a apoiar à imigração espontânea, permitindo à atuação da iniciativa privada no processo. A partir deste período, a configuração dos novos núcleos coloniais passa a ser, em grande parte, da mesma origem geográfica e homogênea no aspecto religioso. Estas características são referidas na literatura como “imigração étnico-confessional” (Ver Capítulo 8 – Anexo 1).

A imigração espontânea, embora numericamente reduzida, voltada para a agricultura, transformava o imigrante de imediato em pequeno proprietário, evitando assim uma massa de população flutuante, migrando de um lugar para o outro, como ocorria, por exemplo, em São Paulo.

Por essa razão, no Rio Grande do Sul, a colonização constitui em uma exceção se comparada ao restante do país, pois não se caracterizou apenas pela importação de simples trabalhadores ou assalariados. Em outros estados, esse regime levou a formação de uma população flutuante, com número elevado de migrações e mobilidade.

1.4.1 Iniciativa privada na colonização do Rio Grande do Sul

As colônias particulares de povoamento essencialmente germânico adentram no planalto central subindo o Alto Jacuí ao longo da via férrea. De acordo com Roche (1969) são assim fundados os núcleos de Santa Clara (1896), Não-me-toque (1897), General Osório (1898), Dona Ernestina (1900), Selbach (1906), no município de Cruz Alta, Barra do Colorado (1897), Boi Prêto (1897), Neu Wurttemberg (1899), o de Santo Ângelo, Ijuí Grande (1892), Vitória (1900), Buriti (1908), Timbaúva e Boa Vista (1912), Steiglich (1914), o de São Luiz

Gonzaga, Cêro Azul (1902), Bela Vista (1903) e Dona Julia (1912); no de Erechim, Rio do Peixe (1911).

Essas colônias criadas por sociedades ou empresas particulares apresentavam características diferenciadas das fundadas pelo governo do estado. Como já foi descrito, reuniam indivíduos da mesma origem geográfica e muitas vezes da mesma confissão religiosa. O governo estadual tinha um maior cuidado em desenvolver os meios de comunicação entre as colônias por ele fundada, tornando assim mais isolados os núcleos particulares. Os alemães católicos agruparam-se em colônias como Boa Vista (hoje Santo Cristo) e Cerro Azul (hoje Cerro Largo), criadas por organismos profissionais e confessionais (Deutsche Bauer Verein), conservando o zelo tanto pela religião como pelas tradições germânicas, enquanto que os protestantes se dirigiram em grande maioria para Neu-Wurttemberger e colônias vizinhas, todas colônias-modelos fundada por H. Meyer. A força que esses blocos homogêneos adquiriram foi motivo de preocupações para o estado.

A partir de 1914, Brasil e Alemanha encontram-se em lados opostos nas guerras mundiais, surgindo desta forma novos problemas culturais e cívicos aos colonos, onde tinham uma dependência ao estado e uma fidelidade espiritual a uma nação.

1.4.2 Colonização da região noroeste

A região noroeste do estado, conhecida atualmente por região missioneira ou das missões, destaca-se pela presença de inúmeros descendentes de diferentes nacionalidades, porém entre essas a presença de descendentes de alemães é maior devido a migração mais recente. Na década de 1820 acontece a primeira tentativa de instalação de colonos alemães na região, intensificando-se a partir do início do século XX.

Devido principalmente ao superpovoamento e esgotamento das terras nas primeiras colônias (colônias velhas) do estado, houve a migração para outras áreas e o fluxo migratório direciona-se a colonização de regiões distantes e em terras com povoamento reduzido, a noroeste do estado. (Amstad, 1999).

Em 1900, a recém-fundada Associação de Colonos intitulada *Bauerverein*, com sede no município de São José do Hortêncio, interessou-se em colonizar uma área de terras na região noroeste, entre os rios Ijuí e Comandaí.

Segundo Lando e Barros (1980), em 1902 o “*Bauerverein*” adquiriu duas áreas de terra da *Nordwestbahn* e comprometeu-se a colonizar a colônia Serro Azul (hoje Cerro Largo) e a colônia Boa Vista (hoje Santo Cristo). Foi assim que inúmeras famílias de colonos descendentes de imigrantes alemães foram assentadas na região. A localidade originalmente chamou-se Serro Azul e abrangia os atuais municípios de Cerro Largo, Roque Gonzales, São Paulo das Missões, Salvador das Missões, Sete de Setembro, Porto Xavier e São Pedro do Butiá (Ramos, 2006).

1.5 O MUNICÍPIO DE CÂNDIDO GODOI E A QUESTÃO DA GEMELARIDADE

A cidade de Cândido Godói (27° 57' 9" Sul, 54° 45' 13" Oeste), com aproximadamente 6.000 habitantes, está localizada no Noroeste do Rio Grande do Sul e é conhecida como a “Cidade dos Gêmeos”, devido às suas altas taxas de nascimentos duplos e a alta recorrência dentro das famílias. Em um levantamento a partir do Datasus do Ministério da Saúde do Brasil (www.datasus.gov.br) entre os anos de 1994 a 2006, foi observado que Cândido Godói apresenta taxa média de nascimentos gemelares maior que a média nacional, ficando acima de 2%. Particularmente um distrito de Cândido Godói conhecido como Linha São Pedro, povoada basicamente por descendentes de alemães que ali se estabeleceram no início do século XX, apresenta prevalências de até 10% de gemelaridade (Matte e cols., 1996)

No início do ano de 2009 a imprensa leiga deu ampla divulgação à publicação de um livro no qual um historiador argentino implicava a alta natalidade de gêmeos a experiências do médico nazista Joseph Mengele que supostamente teria morado neste município no início da década de 1960. Este fato, apesar de desprovido de qualquer fundamentação científica, trouxe notoriedade ao município

(<http://www.telegraph.co.uk/news/worldnews/southamerica/brazil/4307262/Nazi-angel-of-death-Josef-Mengele-created-twin-town-in-Brazil.html>).

Interessado em desfazer esta imagem atribuída à população local de “cobaias de experimentos nazistas” o INAGEMP, em contato com membros da comunidade de Candido Godói, decide retomar um trabalho iniciado há mais de uma década por Matte e cols. (1996), procurando evidenciar fatores genéticos e ambientais envolvidos na alta taxa de gemelaridade neste município. Dois trabalhos, um de natureza epidemiológica e outro de investigação de genes candidatos, mostraram que provavelmente esta alta prevalência de gêmeos é decorrente de um efeito fundador (Tagliani-Ribeiro, 2011 e 2012)

CAPITULO 2 – JUSTIFICATIVA E HIPOTESE DE TRABALHO

A análise da distribuição espacial de sobrenomes nas mais diversas escalas geográficas (distritos, município, estado, país ou continente) pode servir como indicador do nível de endogamia de populações.

Por um lado, os sobrenomes compõem um excelente material para estudos biodemográficos e para analisar a estrutura genética de uma população. Por outro lado, a análise geo-espacial ao longo do território ou em diferentes momentos históricos, contribui significativamente para as pesquisas em genética e saúde, pois permite ter uma dimensão espacial de diversas variáveis em análise e projetar planos em articulação com a população local.

Nossa hipótese de trabalho se apoia na ideia que a análise de sobrenomes e geo-espacial pode ser uma ferramenta útil, ampla e relativamente barata para aplicação em genética médica populacional. Utilizamos como modelo municípios da região noroeste do estado do Rio Grande do Sul, especialmente o município de Candido Godoi, onde historicamente se conhece sua colonização por descendentes de imigrantes europeus, principalmente alemães. Nesta mesma região também foram desenvolvidos estudos em genética médica populacional pelo INaGeMP; como os já citados sobre fatores relacionados à ocorrência de gemelaridade .

3.1 Objetivo Geral

- Avaliar a análise de sobrenome associada a técnicas de geoprocessamento e análise espacial como ferramenta para a compreensão dos processos de migração e suas consequências na estruturação genética de populações humanas
- Mapear as diferentes variáveis sob análise em uma escala microrregional, tanto do ponto de vista geográfico como histórico.

3.2 Objetivos específicos

- Analisar a estrutura isonímica da cidade de Cândido Godói e a distribuição sobrenome dentro da cidade.
- Utilizar sobrenomes como elemento de análise da estrutura da população de Cândido Godoi, Santo Cristo, Cerro Largo.
- Identificar subgrupos populacionais e eventuais casos de isolamento que possam levar a um possível cenário no qual a deriva genética e o efeito fundador poderia ter ocorrido.
- Construção de uma base cartografia georeferenciada com as variáveis físicas e antropológicas da área de estudo.
- Localizar geograficamente as famílias com história de gêmeos dentro do município de Candido Godoi, para posteriormente relacionar com dados ambientais e históricos recolhidos.
- Investigar a estrutura genética da população quanto ao endocruzamento e distância genética entre os diferentes distritos do município de Candido Godói.
- Analisar conjuntamente os dados epidemiológicos, genéticos e ambientais, bem como os registros históricos da colonização, constituição demográfica daquela população, gerando uma explicação integrada para a alta ocorrência de gestações gemelares em Cândido Godói, utilizando para isso técnicas de geoprocessamento.

So Close, So Far Away: Analysis of Surnames in a Town of Twins (Cândido Godói, Brazil)

Marcelo Zagonel De Oliveira^{1,2}, Lavínia Schüler-Faccini^{1,2*}, Dario A. Demarchi³, Emma L. Alfaro⁴, José E. Dipierri⁴, Mauricio R. Veronez⁵, Marlise Colling Cassel^{1,2}, Alice Tagliani-Ribeiro^{1,2}, Ursula Silveira Matte^{1,6} and Virginia Ramallo^{1,2}

¹INAGEMP–Instituto Nacional de Genética Médica Populacional, Porto Alegre, Brazil

²Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

³Museo de Antropología, FFyH, Universidad Nacional de Córdoba, Córdoba, Argentina

⁴Instituto de Biología de la Altura, Universidad Nacional de Jujuy, San Salvador de Jujuy, Argentina

⁵LASERCA–Laboratório de Sensoriamento Remoto e Cartografia Digital – Programa de Pós-Graduação em Geologia, Universidade do Vale do Rio dos Sinos (UNISINOS), Brazil

⁶Unidade de Análise Molecular e de Proteínas, Centro de Pesquisas, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

Summary

Cândido Godói is a small Brazilian town known for high rates of twin birth. In 2011, a genetic study showed that this localized high rate of twin births could be explained by a genetic founder effect. Here we used isonymic analysis and surname distribution to identify population subgroups within 5316 inhabitants and 665 different surnames. Four clusters were constructed based on different twin rates ($P < 0.001$; MRPP test). Fisher's α and consanguinity index showed low and high values, respectively, corresponding with observed values in isolated communities with high levels of genetic drift. Values of A and B estimators confirmed population isolation. Three boundaries were identified with Monmonier's maximum difference algorithm ($P = 0.007$). Inside the isolated sections, surnames of different geographic origins, language, and religion were represented. With an adequate statistical methodology, surname analyses provided a close approximation of historic and socioeconomic background at the moment of colony settlement. In this context, the maintenance of social and cultural practices had strong implications for the population's structure leading to drift processes in this small town, supporting the previous genetic study.

Keywords: Cândido Godói, twins, population structure, isonymy, founder effect

Introduction

Surnames represent cultural features that are transmitted from ancestors to their descendants through a vertical mechanism similar to that of genetic inheritance (Guglielmino et al., 1991). They are historic features of identity (Manrubia & Zanette, 2002) characteristic of a family, a population, or a

group of related populations, and have been recognized as an inheritance system for our species

The analysis of surname distribution can substitute quantitative information on the genetic structure of a given population. All human communities are grouped into a certain structure, either because of their limited number of ancestors, or their acceptance or refusal of a specific kind of union. The selection of potential partners among a given group follows a pattern of preference based on a particular set of values that may vary according to opportunities, will, and the extension of the circle from which the selection is made.

Crow and Mange (1965) developed the isonymic method to calculate the consanguinity coefficient, including random and nonrandom components for large populations (for a review see Lasker, 1985). So far, studies on surname distribution

*Corresponding author: LAVÍNIA SCHÜLER-FACCINI, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500 – Prédio 43323M, Caixa Postal 15053, CEP 91501-970, Porto Alegre, RS, Brasil. Tel: 55 51 3308-6722; Fax: 55 51 3308-7311; E-mail: lavinia.faccini@ufrgs.br

have become widely and successfully applied to estimate the structure and level of migration, admixture, or isolation (Jobling, 2001; King & Jobling, 2009; Graf et al., 2010) in a given community, region, country, or continent (Dipierri et al., 2005, 2007, 2011; Scapoli et al., 2007; Tarskaia et al., 2008; Bronberg et al., 2009; Cheshire et al., 2010).

This work shows results obtained for the population of Cândido Godói, Brazil (27°57' 07"S; 54°45' 07"W) from a Population Medical Genetics approach. The main objective of this multidisciplinary approach is the study and potential medical care of a whole population. This discipline is globally underdeveloped and quite recent in Latin America. The present findings provide the basis for future education projects on health prevention to inform the population about its historical and current reproductive structure.

Cândido Godói is a small town (246.28 km²) in the north-west Rio Grande do Sul state, almost on the border between Argentina and Brazil, with a rural-based economy. This is a small municipality, divided into 24 sections (called "linhas"), with 6535 inhabitants in total and a demographic density of 26.54 persons/km² (Census 2010 IBGE- Brazilian Institute of Geography and Statistics). Almost 39% of the population is established in Linha Centro, the institutional seat and commercial center of Cândido Godói, 11.37 km² in size, and the rest of the population live in the peripheral areas. Most internal roads are in poor condition, making communication between people from different rural "linhas" difficult. The present population is largely of Polish or German ancestry and this characteristic is proudly emphasized in the oral memory of each family. Their ancestors were part of the great migration to the colonies of Rio Grande do Sul that took place during the 19th and beginning of the 20th century, and who then settled at Cândido Godói. At the time, the Brazilian government promoted immigration by allowing private undertakings. As a legacy of the recent past, many people still speak German, Polish, or dialects of both.

In this town, the most notable characteristic is the high rate of twin births (Tagliani-Ribeiro et al., 2011), close to 2%, considerably higher than that observed for the whole of Brazil [1% mean—Data obtained from DATASUS, national health records of the "Ministério da Saúde" for 1994–2006, see reference (Datusus 2010)]. It is known as the "TwinsTown" and this characteristic is a cause of celebration for the whole community. Every two years there is a Twins' Festival, a tourist event of economic relevance. This demographic phenomenon of a high twinning rate is not homogeneously distributed in Cândido Godói because, for instance, the prevalence rate in Linha São Pedro is between 7% and 10% (Matte et al., 1996; Tagliani-Ribeiro et al., 2011). Though most etiological aspects of these twin gestations have not yet been elucidated, several studies have been developed (Hall, 1996; Lummaa et al., 1998; Beemsterboer et al., 2006; Steinman, 2006), report-

ing geographic and time variation in twin birth rates (Hoekstra et al., 2008) involving both genetic and environmental factors (Al-Hendy et al., 2000; Hasbargen et al., 2000; Montgomery et al., 2004; Steinman 2006; Hoekstra et al., 2008, among many others).

Following the initiative of the community at Cândido Godói to ask for help from the University to understand the origins of the high incidence of twins, a research project involving baptism books and family stories on twin births was developed (Tagliani-Ribeiro et al., 2011). Authors supported the hypothesis of a genetic founder effect to explain the high prevalence rate of twin births.

Following this approach, this work analyzes the isonymic structure of the whole community at Cândido Godói and the surname distribution within the town. The purpose of this is to identify population subgroups and eventual cases of isolation that will in the end enable a scenario in which a founder effect could be said to have occurred.

Materials and Methods

Ethical Aspects

This work is part of the research project "Estudos genéticos e ambientais acerca da etiologia da gemelaridade em Cândido Godói, RS," already revised and approved by the Ethics Committee at Porto Alegre Clinical Hospital ("Hospital de Clínicas de Porto Alegre") under Protocol No.09–359. All data were analyzed preserving the identities of volunteers.

Data Sources Used

In Brazil, voting is mandatory for all citizens from 18 years of age. Each municipality in the country is divided into sectors and establishes a general voters' registry (Registro de Eleitores), comprising all residents within the sector over 18 years old. These registries constitute an excellent source of information in population studies because a large number of surnames are included at different levels of aggregation (electoral sector, municipality, state, and finally, an entire nation). In this work, the 2010 electoral registry for Cândido Godói and 23 sectors was analyzed.

Surnames can be indicative of migration, because they can be indicative of geographic origin. For this purpose, databases of the "Instituto de Imigração e Colonização da Polícia Marítima," "Arquivo Histórico Nacional," and "Instituto Histórico e Geográfico Brasileiro," were consulted. These files are records of migration since the beginning of the 19th century, when the immigration process in the northern region of Rio Grande do Sul began. At the time of entering the country, the immigration office recorded the name of

each family group, its members, and their ages, the ship on which they had traveled and which country they came from. These databases were the source for pinpointing a geographical origin for each surname analyzed in this study.

In addition, the National Census 2010 from IBGE (Brazilian Institute of Geography and Statistics) and health register databases from DATASUS (National Health Agency, Brazil) were also consulted. All geographic and demographic information was provided by municipal authorities of Cândido Godói and IBGE.

General Characteristics of the Sample

The total sample comprised 5316 adult inhabitants of Cândido Godói whose data were obtained from the 2010 Electoral Register, representing 81% of the total population.

Every section or *linha* represents an electoral district, except for *Linha Cascata* which was included in the neighboring district by the Regional Electoral Court because of its small number of residents (<30). Thus, data in this study were calculated over the total 23 *linhas* in the Register (“Registro de Eleitores”).

Similar to many other countries in Latin America, every person in Brazil has two surnames: maternal surname first, followed by the paternal surname. Though Cândido Godói shows a significant number of double surnames, most of its inhabitants only use their paternal surname. This practice is frequently found in colonies of European descendants. To make our sample homogeneous for global calculations, we considered only the paternal surname for each individual. Double surnames were separately analyzed, as detailed in the following section.

Statistics and Isonymy

To analyze the variability of surnames in the community, their possible distributional patterns in space and a potential population structure, we calculated a set of statistical indices based on data distribution for each *linha*. Unbiased isonymy was calculated according to the following formula (Rodríguez-Larralde et al., 1993):

$$I_{NS} = \sum_k (N_{ki}/N_i)^2 - (1/N_i)m,$$

where N_{ki} represents the absolute frequency of surname k for *linha* i , and N_i is the size of the population in the *linha*. The sum includes all surnames.

Fisher’s α index was calculated according to Barraí et al. (1996):

$$\alpha = 1/I_{NS}.$$

This parameter reflects the effective number of surnames and was applied to evaluate diversity. High values were observed for communities with high immigration rates, whereas isolated communities with high genetic drift showed low values.

Consanguinity was assessed for the overall population and for each *linha* in particular. This inbreeding coefficient was computed as

$$F = F_n + Fr(1 - F_n).$$

In this formula, Fr is the random component, a measure of isonymy that occurs from random unions within the population and is therefore a function of surname frequency only:

$$Fr = \sum (p_k q_k) / 4,$$

where p_k is the frequency of the k th surname in males, and q_k is the frequency of k th surname in females; summation is over all surnames.

F_n , the nonrandom component, resulting from the preference or intentional rejection of certain marriages.

$$F_n = (I - \sum p_k q_k) / 4(1 - \sum p_k q_k)$$

F_n can be either positive or negative, and immediately becomes zero with random mating (Crow & Mange, 1965)

Isonymy was also calculated among *linhas* to identify the eventual common origin of individuals (Scapoli et al., 2007):

$$I_{ij} = \sum_k P_{ki} P_{kj},$$

where P_{ki} and P_{kj} are the relative frequencies of surname k for *linhas* i and j , respectively. The sum includes all surnames. When two groups show no surnames in common, isonymy is equal to 0 (zero).

A and B estimators of population sedentarism and isolation were calculated (Rodríguez-Larralde, 1990). The first represents the percentage of the population with only one representative per surname and the second is the percentage of the population with the seven most frequent surnames. High values of B estimator were observed for isolated or small communities with high emigration rates, in which only few surnames were found in the majority of the population.

Linhas were divided into four groups according to the different twin rates observed (Tagliani-Ribeiro et al., 2011) and surname distances were then calculated (distance formulas are presented in the following paragraph). The multiresponse permutation procedure (MRPP) was used to test significance of *Linha* grouping according to twin birth rates, using PC-ORD 4.0 software. The MRPP is a nonparametric procedure for testing the hypothesis of no differences between two or

more groups of entities (in this case, populations); it is equivalent to discriminant analysis or one-way analysis of variance (McCune, 1991). As the probability value of an MRPP statistic is derived through a permutation argument, there are no distributional requirements on the data, such as multivariate normality and homogeneity of variances. A permutation is a specific arrangement or assignment of all N objects (in this case population samples) to the specified groups. The null hypothesis for the MRPP states that all the possible permutations are equally likely. The test statistic indicates the extent of differentiation between groups. The observed δ (the average of the within-group distances) is compared to an expected δ , which is calculated to represent the mean δ for all possible partitions of the data. Small values of δ indicate a tendency for clustering, whereas larger values of δ indicate a lack of clustering. The variance and skewness of δ are descriptors of the distribution of all possible values of δ corresponding to the possible partitions of the items. The probability value expresses the likelihood of obtaining a δ as extreme or more extreme than the observed, given the distribution of all possible deltas. For details, see Zimmerman et al. (1985).

Euclidean distance formula (Cavalli-Sforza & Edwards, 1967) between groups I and J is defined as:

$$E = \sqrt{(1 - \sum_k \sqrt{P_{ki} P_{kj}})},$$

and Sørensen's distance formula (Sørensen, 1948):

$$QS = 2C / A + B$$

where A and B are the total number of surnames for A and B linhas respectively, and C is the number of surnames shared by both linhas. The Sørensen index used as a distance measure was calculated as $1 - QS$. Though other distance formulas are available (Nei's, Laker's), those used in this work obtain results between 0 (absolute similarity) and 1 (no similarity). Matrixes of paired isonymic distances were calculated for all linhas, following both distance formulas.

The relationship of surname distribution among linhas was graphically expressed by means of a dendrogram based on the UPGMA (unweighted pair group method with arithmetic mean) algorithm using NTSYSpc 2.11S® software (Rohlf, 2000).

A matrix of geographic distances was constructed from relative central points inside each linha. Significance of the correlation between this matrix and that of surname distances was evaluated by the Mantel test (Mantel, 1967) and PASSaGE 2® software (Rosenberg & Anderson, 2011).

After describing the distribution of population variability in space, discontinuous zones indicate eventual potentially geographic, biological, or cultural barriers. A geometric method based on Monmonier's maximum differentiation al-

gorithm (Monmonier, 1973) was applied to a net connecting all linhas—relative central spots previously defined—by means of Delaunay's triangulation (Brassel & Reif, 1979) to identify potential barriers. According to the matrix of surname distances, barriers were calculated using BARRIERS 2.2® software (Manni et al., 2004), and MRPP was also used to evaluate the statistical significance of these clusters of linhas within boundaries.

According to the historical databases mentioned before, each surname from electoral lists was assigned to a certain geographic origin and the frequencies obtained were compared between linhas. These results were also compared with the distribution of barriers obtained by the Monmonier algorithm.

The Electoral Register of Cândido Godói includes 546 individuals with double surnames. As mentioned before, the first surname is inherited from the mother and the second one from the father. After marriage, women can take the family names of their husbands, adding it as a second surname, discarding her mother's surname and retaining the paternal surname. This inheritance system is useful for identifying marital preferences or residence patterns of the population under study. Every double surname refers to a particular type of union, either contemporary to the individual (her own marriage, for the case of women) or to the previous generation (parents' marriage). This subset of the sample (546 individuals with double surnames, 491 different family names in total) was analyzed separately, establishing the geographical origin of each surname, to identify possible trends toward a preferential kind of union.

The pattern of change of residence, even within the same community, may be different for men and women. If the migration rate varies according to sex, it is possible that the continuity of a certain pattern changes random isonymic values in the following generations, making maternal or paternal surnames more diverse (Pinto-Cisternas et al., 1990; Herrera Paz et al., 2010a; Herrera Paz et al., 2010b). The percentage of two variability models for the lists of doubles surnames was separately determined to estimate the trend in residence patterns.

Maternal percentage was calculated as follows:

$$(Ia1SN / (Ia1SN + Ia2SN))100.$$

Paternal percentage was calculated as follows:

$$(Ia2SN / (Ia1SN + Ia2SN))100.$$

where $Ia1SN$ and $Ia2SN$ represent random isonymy for the first and second surname, respectively.

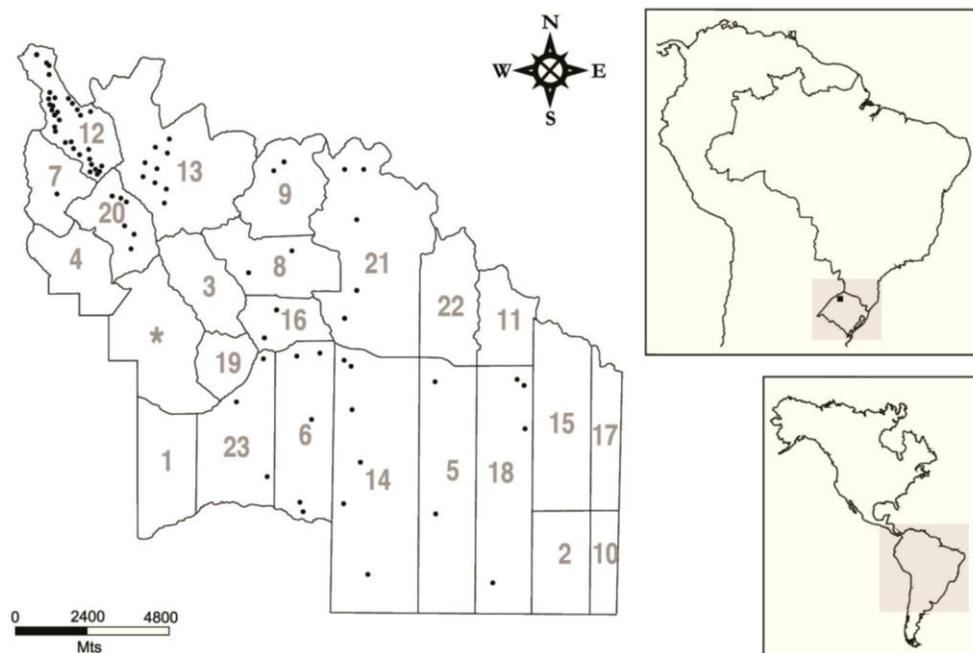


Figure 1 Districts of Cândido Godói and their respective twin birth registration. The distribution is based on twin's baptism records from 1959 to 2006 (Tagliani-Ribeiro et al., 2011). Each twin record is marked with a dot. * Symbol denotes Linha Cascata, not included in the general analysis. See details in "General characteristics of the sample."

Maps and Graphics

Geographic maps and data were obtained with the ArcGis 9.3® software and the cartographic base was geo-referenced to the SIRGAS geodesic system (Geocentric Reference System for the Americas) in UTM projection (Transverse Universal Mercator). Corel-DRAW X3® software (Corel Corporation, Ottawa, Canada) was used to edit all images obtained.

Results and Discussion

Possibly because of erroneous transcription, the electoral register shows alternative spellings for many surnames, with a mean of five variants for some of them. There were a total of 246 surnames with orthographic variants, organized into 49 groups according to graphic or phonetic similarity. This procedure requires careful and homogeneous criteria because the change in the final number of surnames to analyze may lead diversity to be under- or overestimated. Further interviews with local inhabitants enabled confirmation of group assignment. The interest shown by the Cândido Godói community was essential for this research study. Thus, differences were determined as historic mistakes at the time of register-

ing migrant families into Brazil. From a initial total of 862 surnames, an effective final number of 665 different surnames were recorded among 5316 individuals. Figure 1 shows a map of the town and its location in Rio Grande do Sul, Brazil, the different linhas in it and their respective twin birth registration (Tagliani-Ribeiro et al., 2011). Each twin record is marked with a dot.

The distribution by linha of the surname used in the analysis, with the isonymic parameters calculated, is given in Table 1. Data by linha are shown according to increasing Fisher's α values. In general, alpha shows low values, which suggest high consanguinity within sections, except for Centro and Timbauva, the two sections densely populated. Linha Centro, institutional seat and commercial center of Cândido Godói is the one which shows the highest variety of different surnames and number of inhabitants. This linha also shows the lowest isonymy values and therefore the lowest level of consanguinity. Linha Doze Norte is on the opposite side, with 159 inhabitants. Though population concentration in this linha is not the lowest, it shows the lowest α value. This parameter is calculated in direct relationship with unbiased isonymy rate and is a useful tool for the assessment of surname variety. Barrai et al. (2000) have reported the α value to be an effective indicator of the population structure and

Table 1 ID for linha, names, number of residents N, number of different surnames S, A estimator, and B estimator in Cândido Godói.

| ID | Name | N | S | I | α | F | A | B |
|----|----------------|------|-----|---------|----------|--------|------|------|
| 1 | Doze Norte | 159 | 28 | 0.0729 | 13 | 0.0198 | 2.6 | 62.3 |
| 2 | Abrantes | 128 | 24 | 0.0685 | 14 | 0.0191 | 5.5 | 63.3 |
| 3 | Castelo Branco | 61 | 19 | 0.0618 | 16 | 0.0196 | 9.8 | 67.2 |
| 4 | Secção C | 138 | 30 | 0.0541 | 18 | 0.0153 | 7.2 | 56.5 |
| 5 | Paranagua | 179 | 36 | 0.0516 | 19 | 0.0143 | 5.6 | 52.5 |
| 6 | Natal | 147 | 36 | 0.0485 | 20 | 0.0138 | 7.5 | 49.7 |
| 7 | São Bonifácio | 113 | 38 | 0.0428 | 23 | 0.0129 | 17.7 | 52.2 |
| 8 | São João | 107 | 33 | 0.0416 | 24 | 0.0127 | 10.3 | 52.3 |
| 9 | La Salle | 101 | 31 | 0.0394 | 25 | 0.0123 | 10.9 | 54.5 |
| 10 | Dr. Pederneira | 143 | 35 | 0.0380 | 26 | 0.0113 | 4.9 | 44.1 |
| 11 | Dos Louros | 156 | 43 | 0.0357 | 27 | 0.0105 | 9.6 | 51.3 |
| 12 | São Pedro | 214 | 45 | 0.0338 | 29 | 0.0084 | 3.7 | 41.6 |
| 13 | Boa Vista | 155 | 49 | 0.0332 | 30 | 0.0099 | 9 | 43.9 |
| 14 | Acre | 213 | 50 | 0.0307 | 32 | 0.0088 | 4.2 | 40.4 |
| 15 | Dr.P. Toledo | 159 | 41 | 0.0272 | 36 | 0.0084 | 5.7 | 39 |
| 16 | Esquina União | 167 | 46 | 0.0267 | 37 | 0.0082 | 9 | 41.3 |
| 17 | Treze de Maio | 114 | 33 | 0.02647 | 37 | 0.0088 | 9.6 | 48.2 |
| 18 | Silva Jardim | 274 | 56 | 0.02627 | 38 | 0.0075 | 2.9 | 35 |
| 19 | Secção B | 134 | 41 | 0.02562 | 39 | 0.0083 | 7.5 | 37.3 |
| 20 | São Miguel | 223 | 49 | 0.02477 | 40 | 0.0073 | 4 | 37.7 |
| 21 | Secção A | 217 | 57 | 0.02323 | 43 | 0.0070 | 8.8 | 38.7 |
| 22 | Timbauva | 342 | 88 | 0.01395 | 71 | 0.0042 | 6.4 | 26.3 |
| 23 | Centro | 1672 | 368 | 0.00545 | 183 | 0.0015 | 7.3 | 12.7 |
| | Total | 5316 | 665 | | | | | |

Comparison of isonymy parameters (Data by linha according to increasing Fisher's α values)

predictor of inheritance dynamics. In the same way, Linha Castelo Branco has only 61 inhabitants, a low α value and the higher value of B estimator. As detailed in the "Materials and Methods" section, this estimator is the percentage of the population with the seven most frequent surnames. In small communities, this is directly and strongly influenced by emigration rates. Low average values of both indices were found in seven linhas with high records of twin births (Acre, Boa Vista, La Salle, Natal, São Joao, São Pedro, São Miguel; ID numbers 14, 13, 9, 6, 8, 12, 20). Inside these linhas, between 38% and 55% of the inhabitants are related as the seven most frequent surnames are found here, indicating an important degree of isolation. A low α value in the present findings would be the consequence of a significant genetic drift process. In Linha São Bonifacio, the A estimator was 17.7%, i.e., people included in this percentage are the only representatives of their surname. In situations of expansion and rapid population growth (e.g., by factors of population attraction, such as an economic or industry improvement), a high frequency of unique representatives of surnames can be registered, because the new immigrants would introduce variation into the popu-

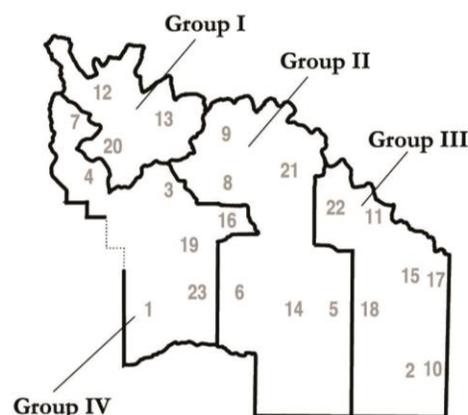


Figure 2 Groups of linhas classified according to different historical rates of twin births. ID numbers for linhas are given in Table 1.

lation. These are people who do not yet have descendants and are as yet the only representatives of their surnames. However, in this rural community, with a significant trend toward migration to urban areas (Censo 2010, IBGE), the value of the A estimator denotes the percentage of last representatives of family names still living in the linha.

Linhas were classified according to different historical rates of twin births, information on which was obtained from baptism records available in Cândido Godói churches (Tagliani-Ribeiro et al, 2011). In Figure 2 the four groups (I, II, III, and IV) obtained are shown. To test whether this grouping, based on a differential characteristic of the population (twinning), was accompanied by a random or non-random distribution of surnames, the isonymic distances within each group were measured, using the Euclidean and Sørensen formula.

Table 2 summarized the MRPP analysis for these four groups. The results reveal that the average distances observed between populations, within groups, are significantly lower than the distances expected for groups of populations randomly generated, both for Euclidean and Sørensen distances. A more detailed analysis shows a lack of clustering for group IV, because within-group distances were higher than those expected from a random distribution. This group is the most heterogeneous because it includes Linha Centro.

Sørensen's coefficient was observed to sensibly correspond with these groups of heterogeneous data and was slightly affected by extreme values. Group I showed the lowest isonymic average distances and included two adjacent linhas with high twin birth rates (São Miguel and São Pedro). Inhabitants in these sections were mostly interrelated; possibly the same lineage groups related by a common history. In fact, historical

Table 2 Results of MRPP test for linha grouping according to twin birth rates, based on Euclidean and Sørensen distance surnames.

| Euclidean Distance | | | |
|--------------------|-------------------------------|-------------|----------|
| Group | Average distance within group | Delta value | p |
| I | 0.0568 | | |
| II | 0.0651 | | |
| III | 0.0627 | | |
| IV | 0.0674 | 0.0630 | 0.000014 |

| Sørensen Distance | | | |
|-------------------|-------------------------------|-------------|----------|
| Group | Average distance within group | Delta value | p |
| I | 0.8625 | | |
| II | 0.9263 | | |
| III | 0.8815 | 0.9065 | 0.000009 |
| IV | 0.9334 | | |

Small values of delta indicate a tendency for clustering, and small *P* values suggest that the null hypothesis (all the possible permutations are equally probable) is unlikely to be true. As compared to Euclidean distance, Sørensen distance measures retain sensitivity in more heterogeneous data sets.

Group I: Boa Vista, São Miguel, São Pedro.

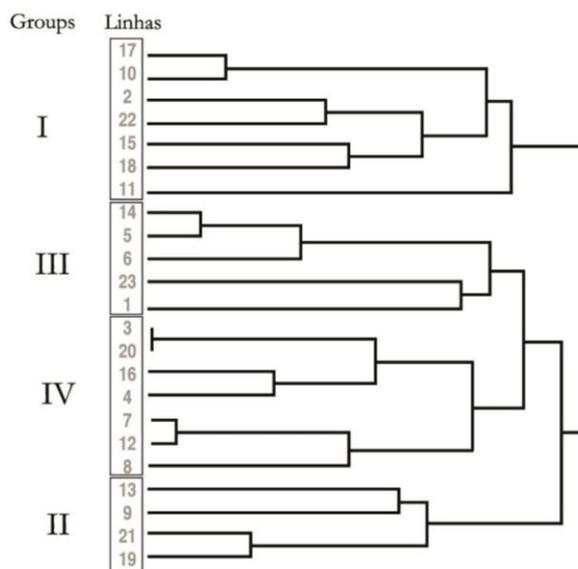
Group II: Acre, La Salle, Natal, Paranaguá, São João, Secção A.

Group III: Treze de Maio, Abrantes, Dos Louros, Pederneiras, Dr. Pedro Toledo, Silva Jardim, Timbauva.

Group IV: Castelo Branco, Centro, Doze Norte, Esquina União, São Bonifacio, Secção B, Secção C.

registers confirm that at the beginning of the 20th century Linha São Pedro was occupied by eight families of German descent who had first settled at Vale dos Sinos, located in the south of the state of Rio Grande do Sul and then moved to Cândido Godói. So far, this population has kept a stable size, a rural-based economy, and only a few other families have joined after its initial colonization. As mentioned before, the tendency to a low population density may be related to the lack of economic incentive in these peripheral regions. Over the years, this situation could be altered by differential migration rates, with individuals in the population moving outside the boundaries of the linha. This phenomenon is not restricted just to Linha São Pedro. Demographic data from the national census show a general trend toward population decrease and genetic drift caused by emigration from rural to urban areas during the last 30 years. In the 1980s, the population of Cândido Godói reached a peak of 8008 inhabitants and then fell steadily to the current population of 6535 (source IBGE).

Sørensen's distance formula was applied to compare pairwise linhas, only based on the surname distribution. From the distance matrix obtained, a dendrogram was constructed using the unweighted pair group method with arithmetic mean (UPGMA) method (Fig. 3), in which the relation-

**Figure 3** Dendrogram obtained from matrix of Sørensen's pairwise surname distances for 23 linhas. Calculated by UPGMA algorithm (matrix based on Sørensen's distance formula). ID numbers for linhas are given in Table 1.

ship between surnames from different sections was graphically represented. The same Sørensen's distance matrix was then compared with a geographic distance matrix (measured in kilometers between linhas) by the Mantel test. The correlation coefficient obtained was 0.421 ($P = 0.001$; after 1000 permutations), showing that both distances have a significant degree of relationship. However, geographical distance would not be the most important factor to explain the variability of surnames found from one linha to another. Some linhas belonging to the same cluster are not close in distance, e.g., Secção B (ID 19), which belongs to Group II, is mainly concentrated to the north of Cândido Godói. The information is more accessible if the typology of the dendrogram is transferred to the map of town, as is shown in Figure 4. Four linha groups can be observed again, though their components (linhas) are somewhat different from those found when they are grouped by twin birth rates (Fig. 2). Groups were numbered according to a decreasing order of inclusiveness. Significantly, São Miguel and São Pedro (ID numbers 20 and 12), two adjacent linhas with high records of twin birth, remained here in the same cluster, now named Group IV, and Group I contains the same linhas as Group III in Figure 2. This cluster remains unchanged in both analyses.

The discontinuities in the isonymic structure of Cândido Godói might be the consequence of directional migration, relocation of a major group (Sokal, 1992), or diverse barriers (linguistic, economic, religious, etc) interrupting genetic

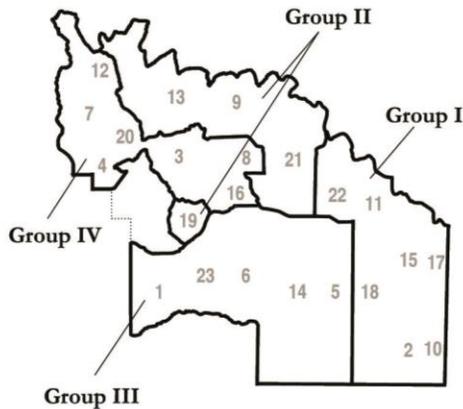


Figure 4 Groups of linhas classified according to dendrogram UPGMA. ID numbers for linhas are given in Table 1. Group numbers correspond to Figure 3. Note cluster II does not contain conterminous linhas.

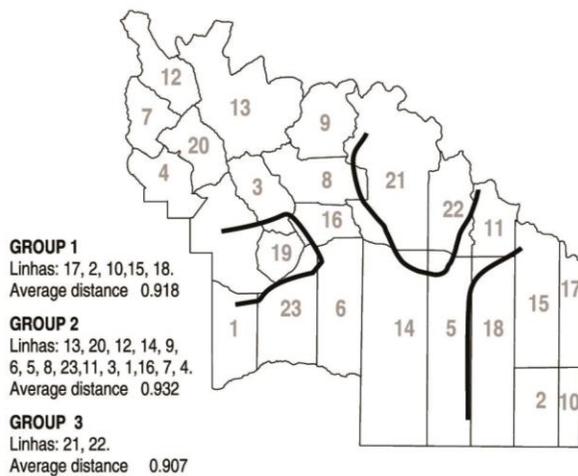


Figure 5 Barriers computed by Monmonier’s maximum difference algorithm, based on matrix of Sorensen’s pairwise surname distances for 23 linhas. Delta = 0.927, $P = 0.007$. Values of delta indicate a tendency for clustering, and small P values suggest that the null hypothesis (all the possible permutations are equally probable) is unlikely to be true. ID numbers for linhas are given in Table 1.

flow. Based on the same Sorensen’s pairwise distance matrix, Monmonier’s maximum difference algorithm enabled the identification of three barriers, defining three different groups of linhas (Fig. 5). The statistical accuracy of this differentiation was also checked by the MRPP test. The null hypothesis states that all the possible permutations (specific arrangement of linhas) are equally likely. The delta value obtained (0.927) was highly significant ($P = 0.007$) and the null hypothesis

rejected. Variability between inhabitants from different linhas was not only associated with inherited surnames but also with cultural characteristics related to those markers. Though quite difficult to adjust, these qualitative data dramatically affect any demographic phenomena involving population displacement or admixture (Barbujani & Sokal, 1990). One of these qualitative data types is religion. Based on the governmental records of Cândido Godói, eight linhas: Secção B, Secção A, Timbauva, Silva Jardim, Pedro Toledo, Abrantes, Treze de Maio, and Pederneiras (ID numbers 19, 21, 22, 18, 15, 2, 17, and 10, respectively, see Fig. 5), contain the highest percentage of Protestants. The three identified barriers separate this fraction from the rest of town, which is principally Catholic. Remarkably, five of these linhas clustered in the dendrogram using the UPGMA algorithm, belonging to Group I, and the same result was found in both analyses (see Figs 4 and 2).

Among the other qualitative data is the geographic origin of the families. German and Polish descendents comprise the two most numerous groups, unevenly distributed across the town. To elucidate this aspect, all the surnames in the sample were classified according to their geographic origin into six major groups: German, Polish, Portuguese, Italian, Minority (United Kingdom, France, Spain, Belgium, Croatia, Ukraine, Hungary), or Undetermined. The source for the designation of a geographical origin is described in “Materials and Methods.” The results of frequencies by linhas are shown in Table 3. German surnames registered high values (between 98.2% and 74.2%) in 21 linhas. This pattern changes only in Linhas Secção A and Secção B, where the majority of surnames are of Polish origin (63.6 and 47.8%, respectively). The remaining four categories (Portuguese, Italian, Minority, and Undetermined) presented considerably smaller frequencies, all less than 12%. Interestingly, barriers calculated before also isolate those linhas showing the highest percentage of Polish descendants, particularly Secção A, Secção B and Timbauva (ID numbers 21, 19, and 22, respectively). The results of these analyses are summarized in Figure 6, in which the frequency of surnames by geographical origin and the most common religious affiliation for each linha are shown. These differences at the level of religion and/or geographical origin could be potentially related to patterns of marriage.

On this topic, 546 individuals with double surnames were analyzed and a total of 491 different surnames were detected among the subset. Table 4 shows all maternal and paternal surnames, classified by geographical origin. The total includes everyone with two surnames. Each combination of surnames marks a particular type of marital union. The values presented in Table 4 indicate a tendency to establish marriage between individuals with surnames of the same geographical origin. It is remarkable to note that people with both German and Polish surnames are less likely to be related to people with Portuguese surnames, especially in unions involving women

Table 3 Frequencies of surnames by geographic origin in 23 linhas of Cândido Godói

| ID | Name | Germany | Poland | Portugal | Italy | Minority | Indeterminate |
|----|----------------|---------|--------|----------|-------|----------|---------------|
| 1 | Doze Norte | 96.2 | 2.5 | 1.3 | – | – | – |
| 2 | Abrantes | 95.3 | 4.7 | – | – | – | – |
| 3 | Castelo Branco | 95.1 | 3.3 | – | 1.6 | – | – |
| 4 | Secção C | 81.9 | 2.9 | 3.6 | 2.9 | 8.7 | – |
| 5 | Paranagua | 73.7 | 21.2 | 1.1 | 1.1 | 2.8 | – |
| 6 | Natal | 87.1 | 10.2 | 1.4 | 0.7 | – | 0.7 |
| 7 | São Bonifácio | 98.2 | 0.9 | 0.9 | – | – | – |
| 8 | São João | 86 | 12.1 | 0.9 | – | 0.9 | – |
| 9 | La Salle | 75.2 | 5.9 | 8.9 | – | 9.9 | – |
| 10 | Dr. Pederneira | 91.6 | 5.6 | – | – | – | 2.8 |
| 11 | Dos Louros | 74.4 | 9 | 3.2 | – | 9.6 | 3.8 |
| 12 | São Pedro | 93 | 2.3 | 2.8 | – | 1.9 | – |
| 13 | Boa Vista | 80.6 | 13.5 | 1.3 | – | 4.5 | – |
| 14 | Acre | 75.1 | 18.8 | 0.5 | 4.7 | 0.9 | – |
| 15 | Dr.P. Toledo | 74.2 | 22 | 3.8 | – | – | – |
| 16 | Esquina União | 91.6 | 3.6 | 2.4 | 1.2 | – | 1.2 |
| 17 | Treze de Maio | 79.8 | 7.9 | 5.3 | 7 | – | – |
| 18 | Silva Jardim | 74.8 | 21.2 | – | – | 2.2 | 1.8 |
| 19 | Secção B | 32.1 | 47.8 | 11.9 | 0.7 | 7.5 | – |
| 20 | São Miguel | 92.8 | – | 1.8 | – | 5.4 | – |
| 21 | Secção A | 24 | 63.6 | 4.1 | 0.9 | 7.4 | – |
| 22 | Timbauva | 77.5 | 16.4 | 4.1 | 1.5 | 0.6 | – |
| 23 | Centro | 77.6 | 6.2 | 6.6 | 2 | 4.8 | 2.8 |

Minority (United Kingdom, France, Spain, Belgium, Croatia, Ukraine, and Hungary).

with German/Polish surnames and men with a Portuguese surname. In this community, the surnames of Portuguese origin represent the fraction of the population that could be considered local, i.e., citizens of Brazilian descent of former migrants from Portugal, not related to migration events during the 19th and beginning of the 20th century, as opposed to the surnames of other origins that we might consider “foreign,” which have arrived more recently into the country. However, these data in Table 4 should be carefully evaluated. As presented in Table 3, most of the inhabitants of Cândido Godói have German names. Because this subset of double surnames is a fragment of the total sample, the frequencies may be biased.

The estimate of possible patterns of residence differently associated with each sex, show that the predominant residence pattern was masculine (56.9%). Maternal surnames are more diverse. This may be because of two causes: (i) women change their place of residence after marriage, which increases surname variability in the linha of her husband; or (ii) the general economy encourages male migration, which affects paternal surname variability. Both scenarios are compatible with a trend of marriage between individuals from the same group of origin.

This has probably been the tendency followed since the very foundation of Cândido Godói. During the 19th and 20th

centuries, the settlement pattern of colonies in Rio Grande do Sul was very particular because not only was the economic development of the area expected but also the tightening of Brazilian limits with Paraguay and Argentina was a factor. German immigration started in 1824. Foreigners were received as rural settlers: they were granted a piece of land and tools, and established colonial settlements quite different from Brazilian land holdings (Neumann, 2008).

As mentioned in the Introduction, the authorities promoted immigration by allowing private undertakings. This policy generated two types of communities: “Public Colonies” gathering populations of different origins and religions under government administration, and those in which private companies were mostly hired by religious institutions to organize “Homogeneous Communities” with faith and language as cohesive identity principles. Thus, after several religious cooperatives were organized by German immigrants (Catholic or Protestant-Lutheran) with the purpose of strengthening faith and providing social care and education (Rambo, 1988), a major education and religious structure was created. These associations made their own teaching material for rural schools and classes were given in German (Kreutz, 2008).

During the 20th century colonists moved to different areas in the state in search of better economic prospects, as

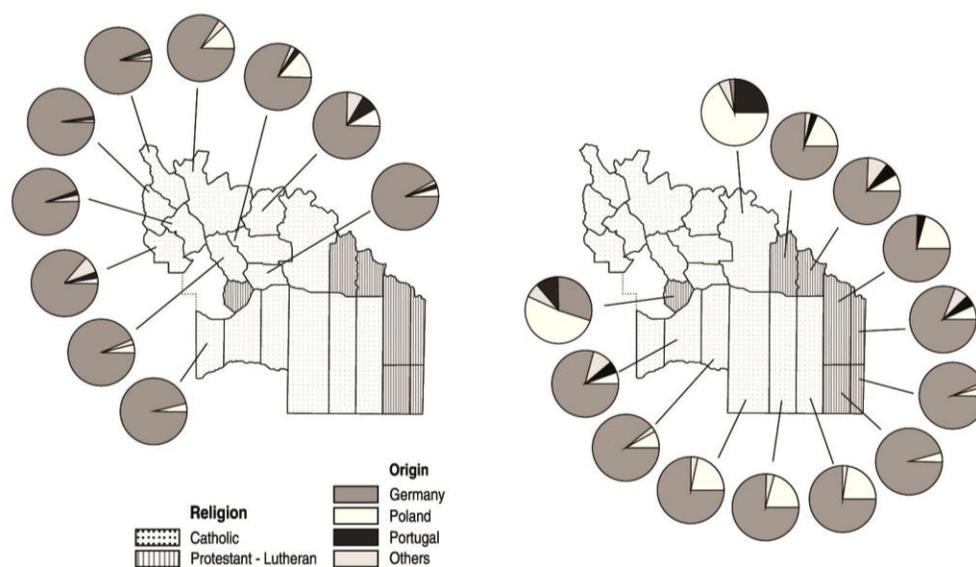


Figure 6 Surname origins and religion by linha. Most common religious affiliation for each linha is also shown.

Table 4 Marriage pattern by surname origin

| Maternal surname | Paternal surname | | | | | |
|------------------|------------------|----|----|----|----|----|
| | G | P | Po | It | M | Un |
| G | 282 | 26 | 23 | 4 | 14 | 2 |
| P | 19 | 19 | 4 | 1 | — | — |
| Po | 19 | 2 | 71 | 1 | 4 | — |
| It | 6 | 2 | 3 | 3 | 2 | — |
| M | 14 | 3 | 8 | 6 | 2 | — |
| Un | 3 | 2 | 1 | — | — | — |

Country codes: G, Germany; P, Poland; Po, Portugal; It, Italy; M, Minority (United Kingdom, France, Spain, Belgium, Croatia, Ukraine, and Hungary); Un, Undetermined-

mentioned for São Pedro, whose population derived from a displacement of families from Vale dos Sinos. This internal migration, mainly organized by religious associations, expanded colonial limits. Particularly since 1914, these new population clusters became more homogeneous and isolated to maintain their ancestral culture and religion. The period between both World Wars, in which Brazil and Germany were on opposing sides, seriously affected these colonies. School classes were taught in Portuguese, and education in the German language and using German texts was prohibited.

Effects of these global changes were also evident in the municipal town of Santa Rosa and consequently Cândido Godói, which was historically included in it. The new administrative status of Cândido Godói was defined only

in 1972. However, its current population has its roots far back in history and is not limited to only one territory or country.

Conclusions

Different research studies on migration, population genetics, and admixture dynamics have also used the isonymic method. There are examples of surname analyses in previously published research findings on the general population of Brazil (Azevêdo et al., 1969, 1980, 1982; Stueber-Odebrecht et al., 1985; Cabello & Krieger, 1991; Leal Barbosa et al., 2006) and on Rio Grande do Sul in particular (Dornelles et al., 1999), where the clustering of mesoregions and a similarity in the frequencies of German surnames was also observed. The low diversity of surnames indicated that the state's population was becoming more homogeneous and still reflected the effect of more recent European migrations.

The analysis of family names is also quite useful for Medical and Population Genetics studies, providing significant available information from state registers and documents, and allowing a large sample size that may even include the whole population for study. Supported by adequate statistical methodology, surname analyses are useful to identify the genetic-demographic structure of any population and enable accurate testing of different hypotheses on micro-evolutionary processes.

When applied to Cândido Godói these analyses provided a close approximation of the historic and socio-economic

background at the moment settlers established this rural colony and the rules they followed to keep their identity. Results obtained show that the maintenance of these rules over time has strong implications for the structure of the population, generating marked phenomena of isolation. In this scenario, processes of genetic drift can lead to significant changes in the variability of the population and the findings support the conclusions of previous investigations: the hypothesis of a genetic founder effect, which occurred during the colonization of Cândido Godói, is a valid alternative explanation of the high prevalence of twin births.

Acknowledgements

The authors wish to thank Dr. Eduardo Castilla for interesting discussions and ideas on surname analysis.

We also thank Mrs. Selia Heck from Cândido Godói municipality for her help with the voters' registry.

The authors acknowledge INAGEMP—National Institute of Population Medical Genetics (grant CNPq 573993/2008–4) for the support provided to this project

Conflict of Interest

The authors declare no conflict of interest.

References

- Al-Hendy, A., Moshynska, O., Saxena, A. & Feyles, V. (2000) Association between mutations of the follicle-stimulating-hormone receptor and repeated twinning. *Lancet* **356**, 914, doi:10.1016/S0140-6736(00)02687-8.
- Azevêdo, E. E., Morton, N. E., Miki, C. & Yee Am, S. (1969) Distance and kinship in northeastern Brazil. *Am J Hum Genet* **21**, 1–22.
- Azevêdo, E. S. (1980) The anthropological and cultural meaning of family names in Bahia, Brazil. *Curr Anthropol* **21**, 360–363.
- Azevêdo, E. S., Fortuna, C. M., Silva, K. M., Sousa, M. G., Machado, M. A., Lima, A. M., Aguiar, M. E., Abé, K., Eulálio, M. C., Conceição, M. M., Silva, M. C. & Santos, M. G. (1982) Spread and diversity of human populations in Bahia, Brazil. *Hum Biol* **54**, 329–341.
- Barbujani, G. & Sokal, R. R. (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci* **87**, 1816–1819.
- Barrai, I., Scapoli, C., Beretta, M., Nesti, C., Mamolini, E. & Rodríguez-Larralde, L. (1996) Isonymy and the genetic structure of Switzerland. The distribution of surnames. *Ann Hum Biol* **23**, 431–455.
- Barrai, I., Rodríguez-Larralde, A., Mamolini, E., Manni, F. & Scapoli, C. (2000) Elements of the surname structure of Austria. *Ann Hum Biol* **27**, 607–622.
- Beemsterboer, S. N., Homburg, R., Gorter, N. A., Schats, R., Hompes, P. G. & Lambalk, C. B. (2006) The paradox of declining fertility but increasing twinning rates with advancing maternal age. *Hum Reprod* **21**, 1531–1532.
- Brassel, K. E. & Reif, D. (1979) A procedure to generate Thiessen polygons. *Geogr Anal* **325**, 31–36.
- Bronberg, R., Dipierri, J. E., Alfaro, E. L., Barrai, I., Rodríguez Larralde, A., Castilla, E., Colonna, V., Rodríguez Arroyo, G. & Bailliet, G. (2009) Isonymy structure of Buenos Aires city. *Hum Biol* **8**, 447–461.
- Cheshire, J. A., Longley, P. A. & Singleton, A. D. (2010) The surname regions of Great Britain. *J Maps* **2010**, 401–409.
- Cabello, P. H. & Krieger, H. (1991) Note on estimates of the inbreeding coefficient through study of pedigrees and isonymous marriages. *Hum Biol* **63**, 719–723.
- Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967) Phylogenetic analysis models and estimation procedures. *Am J Hum Genet* **19**, 233–257.
- Crow, J. E. & Mange, A. P. (1965) Measurements of inbreeding from the frequency of marriages between persons of the same surnames. *Eugen Q* **12**, 190–203.
- Datasus. (2010) <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinasc/cnv/nvrs.def>. Accessed November 28, 2010.
- Dipierri, J. E., Alfaro, E. L., Scapoli, C., Mamolini, E., Rodríguez-Larralde, A. & Barrai, I. (2005) Surnames in Argentina: A population study through isonymy. *Am J Phys Anthropol* **28**, 199–299.
- Dipierri, J. E., Rodríguez-Larralde, A., Alfaro, E. L. & Barrai, I. (2007) Isonymy structure of the Argentine northwest. *Ann Hum Biol* **34**, 498–503.
- Dipierri, J. E., Rodríguez-Larralde, A., Alfaro, E. L., Scapoli, Ch., Mamolini, E., Salvatorelli, G., Caramori, G., De Lorenzi, S., Sandri, M., Carrieri, A. & Barrai, I. (2011) A study of the population of Paraguay through isonymy. *Ann Hum Genet* **75**, 678–687.
- Dornelles, C. L., Callegari-Jacques, S. M., Robinson, W. M., Weimer, T. A., Franco, M. H., Hickmann, A. C., Geiger, C. J. & Salzano, F. M. (1999) Genetics, surnames, grandparents' nationalities, and ethnic admixture in Southern Brazil—do the patterns of variation coincide? *Genet Mol Biol* **22**, 151–161.
- Graf, O. M., Zlojutro, M., Rubicz, R. & Crawford, M. H. (2010) Surname distributions and their association with Y-Chromosome markers in the Aleutian Islands. *Hum Biol* **82**, 745–757.
- Guglielmino, C. R., Zei, G. & Cavalli-Sforza, L. L. (1991) Genetic and cultural transmission in Sicily as revealed by names and surnames. *Hum Biol* **63**, 607–627.
- Hall, J. G. (1996) Twinning: Mechanisms and genetic implications. *Curr Opin Genet Dev* **6**, 343–347.
- Hasbargen, U., Lohse, P. & Thaler, C. J. (2000) The number of dichorionic twin pregnancies is reduced by the common MTHFR 677C→T mutation. *Hum Reprod* **15**, 2659–2662.
- Herrera Paz, E. F. & Mejía de Herrera, D. (2010a). Apellidos en Gracias a Dios: Estructura poblacional y patrones de residencia en la Moskitia Hondureña inferidos por el método de isonimia. <http://lahondurasvaliente.blogspot.com/2010/09/investigacion-apellidos-moskitia.html>. Accessed September 26, 2012.
- Herrera-Paz, E. F., Matamoros, M. & Carracedo, A. (2010b). The Garífuna (Black Carib) people of the Atlantic coasts of Honduras: Population dynamics, structure, and phylogenetic relations inferred from genetic data, migration matrices, and isonymy. *Am J Hum Biol* **22**, 36–44.
- Hoekstra, C., Zhao, Z. Z., Lambalk, C. B., Willemsen, G., Martin, N. G., Boomsma, D. I. & Montgomery, G. W. (2008) Dizygotic twinning. *Hum Reprod Update* **14**, 37–47.

- Jobling, M. A. (2001) In the name of the father: Surnames and genetics. *Trends Genet* **17**, 353–357.
- King, T. E. & Jobling, M. A. (2009) Founders, drift, and infidelity: The relationship between Y chromosome diversity and patrilineal surnames. *Mol Biol Evol* **26**, 1093–1102.
- Kreutz, L. (2008) Escolas de imigração alemã na Argentina, no Brasil e no Chile. In: *Campos múltiplos: Identidade, Cultura e História* (eds A. Sidekum, I. Grützmann & I. C. Arendt), pp. 153–168. São Leopoldo, Brazil: Oikos press.
- Lasker, G. W. (1985) *Surnames and Genetic Structure*. Cambridge University Press, Cambridge, United Kingdom.
- Leal Barbosa, A. A., Bispo Sousa, S. M., Abê-Sandes, K., Alonso, C. A., Schneider, V., Costa, D. C., Cavalli, I. J. & Azevêdo, E. E. (2006) Microsatellite studies on an isolated population of African descent in the Brazilian state of Bahia. *Genet Mol Biol* **29**, 23–30.
- Lummaa, V., Haukioja, E., Lemmetyinen, R. & Pikkola, M. (1998). Natural selection on human twinning. *Nature* **394**, 533–534
- McCune, B. (1991) *Multivariate Analysis on the PC-ORD System*. Oregon, United States of America: Corvallis press.
- Manni, F., Guérard, E. & Heyer, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by “Monmonier’s algorithm”. *Hum Biol* **76**, 173–190.
- Manrubia, S. C. & Zanette, D. H. (2002) At the boundary between biological and cultural evolution: The origin of surnames distributions. *J Theor Biol* **216**, 461–477.
- Mantel, N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**:209–220.
- Matte, U., Le Roux, M. G., Bénichou, B., Moisan, J. P. & Giugliani, R. (1996) Study on possible increase in twinning rate at a small village in south Brazil. *Acta Genet Med Gemellol* **45**, 431–437.
- Monmonier, M. (1973) Maximum-difference barriers: An alternative numerical regionalization method. *Geogr Anal* **3**, 245–261.
- Montgomery, G. W., Zhao, Z. Z., Marsh, A. J., Mayne, R., Treloar, S. A., James, M., Martin, N. G., Boomsma, D. I. & Duffy, D. L. (2004) A deletion mutation in GDF9 in sisters with spontaneous DZ twins. *Twin Res* **7**, 548–555.
- Neumann, R. M. (2008) A iniciativa privada na colonização do donoreste do Rio Grande do Sul: A colonizadora Meyer. In: *Campos múltiplos: Identidade, Cultura e História*. (eds A. Sidekum, I. Grützmann & I. C. Arendt), pp. 123–140. São Leopoldo, Brazil: Oikos press.
- Pinto-Cisternas, J., Rodriguez-Larralde, A. & Castro de Guerra, D. (1990) Comparison of two Venezuelan populations using the coefficient of relationship by isonymy. *Hum Biol* **62**, 413–419.
- Rambo, A. B. (1988) Associativismo teuto-brasileiro e os primórdios do cooperativismo no Brasil. *Persp Econ* **23**, 263–276.
- Rodriguez-Larralde, A. (1990) Distribución de los apellidos y su uso en la estimación de aislamiento y sedentarismo en los municipios del Estado Lara, Venezuela. *Acta Científica Venezolana* **41**, 163–170.
- Rodriguez-Larralde, A., Formica, G., Scapoli, C., Baretta, M., Mamolini, E. & Barrai, I. (1993) Microevolution in Perugia: Isonymy, 1890–1990. *Ann Hum Biol* **20**, 261–274.
- Rohlf, F. J. (2000) NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, version 2.11s. *Exeter Software, Setauket*, New York.
- Rosenberg, M. S. & Anderson, C. D. (2011) PASSaGE: Pattern analysis, spatial statistics and geographic exegesis. Version 2. *Meth Ecol Evol* **2**, 229–232.
- Scapoli, C., Mamolini, E., Carrieri, A., Rodriguez Larralde, A. & Barrai, I. (2007) Surnames in Western Europe: A comparison of the subcontinental populations through isonymy. *Theor Popul Biol* **71**, 37–48.
- Sokal, R. R., Harding, R. M., Lasker, G. W. & Mascie-Taylor, C. G. N. (1992) A spatial analysis of 100 surnames in England and Wales. *Ann Hum Biol* **19**, 445–476.
- Sørensen, T. (1948) A method for establishing groups of equal amplitude in plant sociology based on similarity of species content. *K Dan Vidensk Selsk Bio Skr* **5**, 1–34.
- Steinman, G. (2006) Can the chance of having twins be modified by diet? *Lancet* **367**, 1513–1519.
- Stueber-Odebrecht, N., Chautard-Freire-Maia, E. A., Primo-Parmo, S. L. & Carrenho, J. M. X. (1985) Studies on the CHE1 locus of serum cholinesterase and surnames in a sample from Santa Catarina (Southern Brazil). *Rev Brasil Genet* **8**, 535–543.
- Tagliani-Ribeiro, A., Oliveira, M., Sassi, A. K., Rodrigues, M. R., Zagonel-Oliveira, M., Steinman, G., Matte, U., Fagundes, N. J. R. & Schuler-Faccini, L. (2011) Twin town in south Brazil: A nazi’s experiment or a genetic founder effect? *PLoS One* **6**, e20328.
- Tarskaia, L., El’chinova, G. I., Scapoli, C., Mamolini, E., Carrieri, A., Rodriguez-Larralde, A. & Barrai, I. (2008) Surnames in Siberia: A study of the population of Yakutia through isonymy. *Am J Phys Anthropol* **138**, 190–198.
- Zimmerman, G., Goetz, H. & Mielke, P. (1985) Use of an improved statistical method for group comparisons to study effects of prairie fire. *Ecology* **66**, 606–611.

Received: 5 April 2012

Accepted: 16 September 2012

Histórias Cruzadas: Análise Histórica, Imigração e Distribuição Espacial de Sobrenomes no Rio Grande do Sul. (manuscrito em preparação)

Marcelo Zagonel de Oliveira¹, Lavínia Schüller-Faccini¹, Alice Tagliani-Ribeiro¹, Mauricio R. Veronez³, Virginia Ramallo¹.

¹ Departamento de Genética - Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.

² INAGEMP – Instituto Nacional de Genética Médica Populacional, CNPQ, Brasil.

³ LASERCA - Laboratório de Sensoriamento Remoto e Cartografia Digital – Programa de Pós-graduação em Geologia, Universidade do Vale do Rio dos Sinos (UNISINOS), Brasil.

INTRODUÇÃO

Os sobrenomes são traços culturais transmitidos a partir de um ancestral a seus descendentes através de um mecanismo vertical, constituindo um sistema de herança único de nossa espécie. O registro e a sucessão dos sobrenomes seguem regras particulares em cada estado ou país. Foi estabelecido desde a Idade Média na Europa, tornando-se ao longo do tempo um sistema mundial para relacionar pessoas aparentadas. Todos os indivíduos recebem um sobrenome e podem transmiti-lo aos seus descendentes. Porém, as correlações entre famílias e sobrenomes podem não ser sempre diretas, por exemplo, famílias diferentes podem receber o mesmo sobrenome por exercer a mesma atividade ou ofício ou por morar no mesmo local (topônimos). Até o modo de escrita dos nomes pode sofrer mudanças e mutações, erros ao acaso que levam a emergência de novas variações. Dadas estas características, os sobrenomes podem ser comparados a um sistema genético com múltiplos alelos neutros (Guglielmino et al., 1991).

Todas as comunidades humanas possuem uma estrutura de organização interna, aceitando ou rejeitando certos tipos de casamentos, seja por um numero limitado de indivíduos ou por fatores culturais associados a um conjunto particular de valores. A análise e distribuição de sobrenomes, o chamado método isonímico, pode ser uma fonte de informações quantitativas para a abordagem da estrutura genética da população. Este método, desenvolvido por Crow e Mange (1965), permite calcular o coeficiente de consanguinidade para grandes populações.

Desde então, diversos estudos utilizando distribuição de sobrenomes mostraram ser amplamente aplicáveis, constituindo-se em indicadores úteis na abordagem de estrutura populacional ou fenômenos como migração, miscigenação ou isolamento (Jobling, 2001; King & Jobling, 2009; Graf et al.,2010) em vários níveis: comunal, regional, nacional ou continental (Dipierri et al.,2005, 2007;. Scapoli et al.2006; Tarskaia et al 2008; Bronberg et al.,2009, Chesire et al 2010)

Este trabalho apresenta os resultados obtidos através da análise de sobrenomes na população de três municípios do estado do Rio Grande do Sul: Cândido Godói, Santo Cristo e Cerro Largo. Localizados no Noroeste do estado, quase no limite com a Argentina, os municípios contam com populações pequenas: Cândido Godói (6535 habitantes), Santo Cristo (14378 habitantes), e Cerro Largo (13289 habitantes) (Censo 2010, IBGE), sendo uma alta porcentagem (tendo variação entre 98,2 e 74,2 % nas 23 linhas) da população descendente de famílias provenientes da Alemanha, decorrentes dos movimentos migratórios ocorridos no final do século 19 e início do século 20.

O município de Cândido Godói foi alvo de estudos anteriores por uma característica muito peculiar: apresentar altas taxas de nascimentos de gêmeos, cerca de 2%, bastante superior à média de 1% para todo o Brasil (dados do Ministério da Saúde, Datasus, entre 1994 e 2006). A distribuição deste fenômeno demográfico não é homogênea no município: no distrito de Linha São Pedro, a prevalência é entre 7 e 10%, (Matte et al.,1996; Tagliani et al.,2011). Embora vários aspectos da etiologia das gravidezes duplas não foram ainda

elucidados, tem havido vários estudos no mundo (Hall, 1996; Lummaa et al.,1998; Beemsterboer et al.,2006; Steinman, 2006), considerando que as taxas de nascimento apresenta variação geográfica e temporal (Hoekstra et al.,2008) e envolvem diferentes graus de fatores genéticos e não-genéticos: idade materna avançada, aumento da paridade, a dieta, o aumento do índice de massa corporal, ascendência, etc. (Steinman, 2006; Al-Hendy et al.,2000; Hasbargen et al.,2000; Montgomery et al.,2004; Hoekstra et al.,2008).

No ano 2011 foi lançada uma investigação, solicitada pela própria comunidade de Cândido Godói, na tentativa de atingir um maior entendimento deste fenômeno ocorrente no município (Tagliani et al.,2011). Foram estudados os livros de batismo e as histórias familiares junto com variáveis ambientais (tipo de dieta, aceso a agua potável, exposição a diversos fatores durante a gravidez), comparando mulheres com registros de nascimentos únicos e duplos, concluindo que a hipótese de um efeito fundador explica adequadamente a alta prevalência de gemelaridade. Ao ano seguinte, os estudos genéticos com a mesma amostra de mães de gêmeos mostraram uma relação entre a presença do alelo P72 da proteína TP53 e a suscetibilidade à gravidez dupla. (Tagliani et. al.,2012)

Procurando abordar o contexto sócio histórico no qual ocorreu o efeito fundador, De Oliveira et al., (2013) analisaram a isonímia e a distribuição espacial dos sobrenomes em Cândido Godói, identificando sub-conjuntos populacionais em distritos isolados do município. Dentro destas seções, encontram-se estabelecidos moradores com sobrenomes da mesma origem geográfica, falando a mesma língua e com a mesma religião. Com uma metodologia estatística adequada, a análise de sobrenomes permitiu uma aproximação ao momento da criação deste núcleo populacional. Entre o fim do século XIX e início do século XX, estabeleceram-se colônias na região noroeste do estado que reuniam imigrantes europeus, especialmente alemães e praticantes da religião católica. Estas colônias, classificadas como de características étnico-confessionais, foram incentivadas pelas autoridades do governo brasileiro e organizadas pela iniciativa privada, onde cada empreendimento tinha um sistema de organização própria. A construção da identidade dentro das colônias ou dentro

de qualquer outro grupo humano não precisa de nenhuma particularidade biológica, é parte do contexto sociocultural de um conjunto de pessoas que são biologicamente heterogêneas. Esta construção envolve características discerníveis como língua, religião, padrões de interação social, etc (Crews & Bindon, 1991) e são estas mesmas características de identidade as que podem influenciar na estrutura genética da população através das práticas de casamento e reprodução, na estrutura biológica e ecológica através da nutrição, a transmissão de doenças, o padrão de mortalidade, etc. Neste contexto, a manutenção de certas práticas culturais teve fortes implicações, levando a um processo de deriva e favorecendo o efeito fundador proposto no estudo genético anterior.

Durante as entrevistas realizadas com mães de gêmeos surgiu um dado muito interessante: grande parte de suas famílias moravam inicialmente em Santo Cristo e Cerro Largo antes de fixar residência em Cândido Godói. Seguindo esta evidência, o presente trabalho analisa a distribuição de sobrenomes também nesses dois municípios geograficamente vizinhos, antigas colônias que passaram pelo mesmo processo de criação. Baseado nos heredogramas construídos por Tagliani et al, (2011), foram selecionados os dezoito nomes de família com maior frequência de nascimentos gemelares na cidade de Cândido Godói, sendo utilizados neste trabalho como indicadores populacionais. A cada sobrenome foi associada uma origem geográfica mais provável. Para verificar se as mesmas variáveis históricas e culturais que determinaram subconjuntos em Cândido Godói agem também no contexto geral dos três municípios e no estado como um todo foram empregadas duas ferramentas geoestatísticas: auto correlação espacial e a técnica de Krigagem, ambas apropriadas na identificação de possíveis agrupamentos.

MATERIAIS E MÉTODOS

1. Aspectos éticos

Este trabalho faz parte do projeto de pesquisa “Estudos genéticos e ambientais a cerca da etiologia da gemelaridade em Cândido Godói, RS”,

aprovado pelo comitê de Ética do Hospital de Clínicas de Porto Alegre, número do protocolo 09-359. Todos os dados analisados provem de bases públicas e foram tratados de forma anônima, global e não vinculante, respeitando assim o direito a privacidade. Para os dezoito sobrenomes foi empregada uma codificação sequencial, abrangendo as dezoito primeiras letras do alfabeto.

2. Características gerais da amostra

As coordenadas geográficas dos municípios são: Santo Cristo (27° 49' 32" Sul, 54° 39' 50" Oeste), Cândido Godói (27° 57' 9" Sul, 54° 45' 13" Oeste) e Cerro Largo (27° 57' 9" Sul, 54° 45' 13" Oeste). Santo Cristo e Cândido Godói são municípios vizinhos, compartilhando divisa administrativa (Figura 1).

Para a análise de sobrenomes foram considerados os cidadãos maiores de 18 anos incluídos no Registro de Eleitores do Rio Grande do Sul do ano 2010. Estas bases de dados são frequentemente empregadas nas análises isonímicas e frente a outras fontes de informação (como lista telefônica, por exemplo) apresentam a vantagem de ser mais representativas da população em estudo. O registro de eleitores da cidade de Candido Godoi inclui 5316 indivíduos (constituindo mais de 81% da população total), em Santo Cristo 12202 indivíduos (mais de 84 %) e 9638 indivíduos da cidade de Cerro Largo (mais de 72% da população).

Cada município contém um conjunto de diferentes distritos eleitorais. Segundo diretrizes do Tribunal Regional Eleitoral (TRE) os residentes em seções eleitorais onde o número de eleitores é menor que cinquenta votantes, são integrados ao distrito vizinho. Os cálculos para análise de distribuição de sobrenomes foram realizados em 23 distritos em Cândido Godói, 28 em Santo Cristo e 15 em Cerro Largo.

3. Estatística

No Brasil cada pessoa recebe um sobrenome de cada progenitor, escrevendo primeiro o nome materno seguido do paterno, porém ao casar as

mulheres podem optar por adotar o sobrenome de seu marido, retirando seu nome materno original.

Sendo assim, os sobrenomes duplos nos fornecem informações sobre diversas práticas de matrimônio. A informação pode ser da geração anterior, sobre o matrimônio dos pais de cada pessoa. No caso das mulheres casadas que trocaram seus sobrenomes, nos oferecem informação sobre um evento contemporâneo.

As amostras analisadas de Santo Cristo, Cândido Godói e Cerro Largo apresentaram um número elevado de sobrenomes duplos, mas a maior parte da população possui apenas um sobrenome, neste caso o paterno. Para homogeneização da amostra nos cálculos globais, decidimos trabalhar apenas com o sobrenome paterno, por este tratar-se o que será efetivamente transmitido para a próxima geração.

O cálculo de isonomia, definida como a probabilidade que duas pessoas compartilhem o mesmo sobrenome por ascendência comum, foi realizado segundo a fórmula (Rodríguez-Larralde et al., 1993)

$$I_{NS} = \sum_k (P_{ki} / N_i)^2 - (1 / N_i)$$

Onde P_{ki} é a frequência dos sobrenomes k na localidade e N_i é o tamanho populacional da localidade, sendo o somatório sobre todos os sobrenomes. Os valores finais podem variar entre 0 e 1, quanto mais próximo do zero maior variabilidade terá a população, já que menor será a probabilidade de duas pessoas ter um ancestral próximo comum.

O índice de α de Fisher calculou-se segundo Barrai et al., (1996).

$$\alpha = 1 / I_{NS}$$

Este índice avalia a diversidade nominal de cada localidade, pois é uma medida direta do número efetivo de sobrenomes. Valores baixos correspondem a localidades mais isoladas, sinalizando uma alta taxa de deriva

genética, enquanto valores altos são observados em localidades maiores, com taxas de imigração mais elevadas.

A consanguinidade da população por localidade foi calculada de acordo com a fórmula

$$F = F_n + F_r(1 - F_n)$$

onde F_n é o componente não aleatório, resultante da preferência ou rejeição intencional de determinados casamentos, podendo resultar em valores positivos ou negativos e tornando-se zero quando os matrimônios são ao acaso. F_r é o componente aleatório, uma medida de isonomia que ocorre a partir de uniões ao acaso dentro da população, estando portanto só em função da frequência dos sobrenomes:

$$F_r = \sum (p_k q_k) / 4$$

onde p_k é a frequência do sobrenome k em homens, e q_k é a frequência do sobrenome k em mulheres, sendo o somatório sobre todos sobrenomes.

Foram calculados os estimadores de sedentarismo (estimador A) e isolamento (estimador B) da população, conforme Rodriguez-Larralde (1990). O estimador A é o percentual da população coberta por sobrenomes únicos, com apenas um representante, o estimador B é o percentual da população coberta pelos sete sobrenomes mais frequentes. Valores altos de A são registrados quando o crescimento natural da população é negativo. Este índice traduz a diferença entre nascidos-vivos e óbitos: não havendo registro de novos nascimentos, não há possibilidade de herdar os sobrenomes. As pessoas fora da idade reprodutiva e que não tiveram filhos ficam como as únicas e, às vezes, últimas representantes de um sobrenome. Valores altos de B são observados em populações pequenas ou isoladas, onde encontram-se alguns sobrenomes muito comuns, compartilhados pela maior parte da população. Pode ser consequência do estabelecimento inicial de poucas famílias, cujos sobrenomes tornaram-se muito frequentes ao longo dos anos. Mais além desta descrição preliminar, os dois

estimadores fornecem dados muito informativos sobre a dinâmica demográfica e devem ser corretamente interpretados dentro do contexto particular de cada localidade, já que existe outra forma de crescimento associada à população migratória. Neste sentido foram especialmente interessantes os resultados obtidos para os três municípios, que são apresentados e discutidos ao longo deste trabalho.

A auto correlação espacial apresenta o nível de similaridade, proximidade, correlação ou aleatoriedade dos valores amostrais analisados. Foi medida através do índice I de Moran (Anselin, 1995), observando um nível de significância $p < 0,05$ e valores de dois desvios em relação à média (Z-score). Para identificar e representar padrões espaciais utilizamos os mapas de agrupamentos ou “*clusters*”, permitindo analisar também os “*outliers*”, aqueles casos atípicos que não são possíveis de serem agrupados.

A tipologia de padrões é dividida em agrupamentos com valores altos e próximos (High-High ou HH), agrupamentos com valores baixos e próximos (Low-Low ou LL), “*outliers*” com valores altos que não se agrupam, pois se encontram em meio de valores baixos (HL), “*outliers*” com valores baixos que não se agrupam, pois se encontram em meio de valores altos (LH) e não significativos quando não se enquadram em nenhum agrupamento.

Nos mapas de distribuição dos sobrenomes utilizamos a interpolação por krigagem, um processo de combinação dos pontos amostrados onde a cada ponto é atribuído um peso mediante uma matriz de covariância espacial (Camargo, 1997). A principal vantagem desta técnica é levar em conta a configuração geral da posição das amostras. Uma amostra mais próxima de um conjunto com distribuição heterogênea tem maior probabilidade de ter um valor parecido com o grupo, podendo também prever valores em locais não amostrados. Segundo Ribeiro Júnior (1995), o estimador de krigagem é o melhor estimador linear não enviesado de variância mínima, pois procura que a média dos erros (desvios entre o valor real e o estimado) seja a mínima possível.

MAPAS E GRAFICOS

Mapas e análises dos dados geográficas foram realizadas com o software ArcGis 10® e a base cartográfica foi georeferenciada ao sistema geodésico SIRGAS (Sistema de Referência Geocêntrico para as Américas) na projeção UTM (Universal Transversa de Mercator).

RESULTADOS

Os parâmetros de isonomia calculados por localidade, são resumidos nas tabelas 1 (Santo Cristo), 2 (Cândido Godói) e 3 (Cerro Largo). Observa-se em todas que o índice Alfa apresenta valores notavelmente mais altos nas três sedes municipais (distritos 28, 23 e 15 nas tabelas 1, 2 e 3, respectivamente), que constituem o centro comercial e industrial de cada cidade. Este índice está ligado a diversidade de sobrenomes, sendo maior seu valor quanto mais variável é a população. Como consequência desta diversidade, os cálculos de isonomia para estes distritos resultaram os mais próximos a zero.

O estimador de sedentarismo (índice A) apresentou valores maiores nos distritos de Santo Cristo, onde aproximadamente 964 vizinhos listados no registro de Eleitores são únicos representantes dos seus sobrenomes. Resultados similares foram obtidos para os distritos de Cândido Godói, com 364 pessoas como únicas portadoras dos seus sobrenomes morando no município. Estes valores representam o 7,9% e 6,9% do total de votantes em Santo Cristo e Cândido Godói, respectivamente. Em Cerro Largo, o índice A também resultou elevado, com 528 pessoas, um 5,5% do total do Registro de Eleitores. Por outro lado, os índices de consanguinidade (F) não resultaram valores homogêneos entre os municípios, encontrando-se os valores mais elevados nas localidades situadas em regiões predominantemente agrícolas, de baixa densidade demográfica e com baixa diversidade de sobrenomes (baixo índice Alfa). Consequentemente, nas mesmas localidades, o estimador de isolamento (B) revelou valores mais elevados: os moradores seriam descendentes do mesmo conjunto de famílias, com um acervo pouco variável de sobrenomes. A figura 2 apresenta mapas com os três índices interpolados pelo método de Krigagem.

As regiões limítrofes de Cândido Godói e Santo Cristo destacam-se pelos valores elevados para índices de consanguinidade (F) e isolamento (estimador B) e valores baixos do índice de diversidade (Alfa). Registra-se uma continuidade da população entre os dois municípios, não sendo limitados pela divisa administrativa. Foi realizada então uma análise dos dados em conjunto, desconsiderando os limites municipais. Os resultados obtidos indicam uma grande região de elevada consanguinidade, estendendo-se do Noroeste (NO) de Santo Cristo até Sudeste (SE) de Cândido Godói, (Figura 3).

Dos dezoito nomes de família com maior frequência de nascimentos gemelares, todos resultaram de origem alemã. A figura 4 apresenta as distribuições espaciais no território do estado com um gradiente de cores, sendo a cor azul equivalente à nenhuma ocorrência do sobrenome, mudando gradualmente até a cor vermelha, em aqueles locais onde tem frequências mais altas. Alguns sobrenomes tiveram ocorrência difusa (Sobrenomes C, F, H), entanto outros apresentaram maior concentração em determinadas regiões (A, B, E, G, I, Q, R), especialmente no Noroeste, Norte e Nordeste. Registraram-se menores ocorrências na região Leste e Sudeste e nenhuma nas regiões Sul, Sudoeste e Oeste.

Os sobrenomes A, B, E, G, I, Q registram nascimentos gemelares tanto na família materna quanto paterna, tendo estes matrimônios B-E, A-I e Q-G uma distribuição muito restrita no estado. Destaca-se o sobrenome E, que está presente em apenas dois locais no Rio Grande do Sul, apenas a leste (onde se inicia uma fase do processo de imigração alemã) e no Noroeste, onde estão localizados os municípios em estudo (Cândido Godói, Santo Cristo, Cerro Largo).

A figura 5 apresenta os resultados da análise de “*cluster*” destes 18 sobrenomes. Os sobrenomes A, B, E, G sempre ocorrem agrupados na região composta pelos três municípios e, interessantemente, registram-se “*clusters*” dos sobrenomes Q e I em outras regiões também ligadas ao mesmo processo de imigração alemã, no Noroeste do estado.

DISCUSSÃO

O uso combinado de análise de sobrenomes e geoprocessamento aplicadas a uma pequena região permitiu-nos realizar análises pormenorizadas com relativa precisão. Contar com múltiplas bases de dados faz com que seja possível interpretar os resultados dentro do contexto global atual, produto da história dos três municípios. Vale destacar que, embora a porcentagem da população coberta por sobrenomes únicos, com apenas um representante, tem sido elevada nas três localidades, o estimador A não seria índice de iguais situações acontecendo nestas populações. Como foi apresentado no início, as listas de eleitores de Cândido Godói e Santo Cristo abrangem entre 81-84 % do total dos moradores. Para o município de Cerro Largo, cobre só 72% da população, ficando 3651 pessoas fora do registro. Segundo o último censo (IBGE, 2010), Santo Cristo e Cândido Godói tiveram uma taxa estimada de crescimento para 2013 de 3% e 2%, respectivamente, sendo maior para Cerro Largo, de 4%. A diferença torna-se mais chamativa se consideramos que Cerro Largo é o município com menor superfície: somente 177,675 km², frente a Santo Cristo (366,886 km²) ou Cândido Godói (246,276 km²), com a consequência natural de ter a maior densidade demográfica: 74,79 hab./km². Neste contexto, os valores altos do estimador A não estão assinalando aos últimos representantes de um sobrenome. Bem pelo contrário, aponta ao crescimento demográfico tanto na forma natural como aquela associado a chegada de população migratória. Os 5,5% de representantes únicos dos seus sobrenomes no registro de eleitores de Cerro Largo, bem podem-se tratar dos pais de pessoas que ainda não atingiram a idade mínima para serem registrados como votantes, ou dos migrantes internos recém-chegados, que ainda não tiveram filhos e não contribuíram com seus sobrenomes para a variabilidade geral da população.

Os resultados revelam que há uma clara diferença entre os três municípios da região, sendo Santo Cristo e Cândido Godói formando um grupo com maior grau de isolamento, onde os índices de consanguinidade são maiores e Cerro Largo com menor grau de isolamento, sendo a evolução populacional mais rápida neste último.

Os sobrenomes das famílias que possuem registro de gêmeos apresentaram uma distribuição muito restrita ao nível do estado e também dentro dos municípios estudados. A explicação provavelmente está relacionada ao processo formador destas colônias, onde a imigração étnica confessional foi muito marcante (De Oliveira et al., 2013).

As sedes municipais das cidades em estudo e as regiões do entorno, por serem polos de atividade econômica, tornam-se atrativos laborais e financeiros, tendo como consequência maior densidade demográfica e maior diversidade de sobrenomes. Enquanto que nas regiões rurais, onde o processo de imigração ocorrido no início do século passado pode ter ficado cristalizado, o grau de isolamento é maior e menor a diversidade de sobrenomes. As práticas culturais diferenciadas criaram obstáculos ou barreiras que levaram à subdivisão da região. Este fenômeno não é raro e pode acontecer ainda em escala muito maior. Os resultados do estudo de isonimia de Scapoli et al. (2007) na Europa Ocidental, apontam para um forte relação entre os idiomas locais e a distribuição de sobrenomes, identificando “*clusters*” com divisas quase exatamente coincidentes com as nações subcontinentais.

No Rio Grande do Sul, os “*clusters*” achados para os 18 sobrenomes com registros gemelares, ficam localizados exclusivamente em zonas de colonização alemã. Vale lembrar que para todos eles também foi identificada uma origem alemã, contando-se o sobrenome H entre os 100 sobrenomes mais frequentes na Europa Ocidental, calculado sobre uma base de 26,2 milhões de pessoas (Scapoli et al., 2007). Até a segunda guerra mundial, nestas colônias do Rio Grande do Sul o idioma alemão era ensinado em muitas escolas (Rambo, 1988) e as famílias de origem alemã em sua grande maioria ainda o falam no âmbito doméstico, sendo o português adquirido como segunda língua, as festas tradicionais destas cidades ainda são muito ligadas a tradição germânica.

Barrai et al. (2003) também estudaram a variabilidade de sobrenomes num contexto multilíngue: Bélgica. Nesta nação, a diversidade cultural e ideológica é refletida na difícil história política e no complexo sistema de governo. A mesma variabilidade foi identificada na análise isonímica: a estrutura

dos sobrenomes está ligada em forma estrita as línguas locais, limitando fortemente processos populacionais como a migração ou a miscigenação. O poder de isolamento dos diferentes idiomas é pelo menos tão grande como o das barreiras físicas.

CONCLUSÕES

Demostramos que a estrutura genética e sociocultural pode ser abordada em uma escala microrregional, e que estas podem ser mapeadas tanto do ponto de vista geográfico como histórico. O conhecimento pormenorizado do passado das populações humanas é fundamental para o entendimento das distribuições da variabilidade. A análise de sobrenome associadas a técnicas de geoprocessamento e análise espacial é uma ferramenta econômica, rápida e eficiente para a compreensão dos processos de migração e suas consequências na estruturação genética. Este estudo apoia as conclusões anteriores de Tagliani-Ribeiro et al. (2011) e De Oliveira et al. (2013), reforçando a ideia de um processo de deriva genética seguida de um efeito fundador em Candido Godói, tendo como resultado a alta prevalência de gemelaridade nesta população.

Agradecimentos: Este trabalho foi financiado pelo projeto INAGEMP/CNPQ, proc. nr. 573993/2008-4. MZO recebeu bolsa de doutorado CAPES e VR, bolsa de Pós-doutorado do CNPQ.

REFERÊNCIAS

Al-Hendy, A., Moshynska, O., Saxena, A. & Feyles, V. (2000) Association between mutations of the follicle-stimulating-hormone receptor and repeated twinning. *Lancet* 356, 914.

Anselin L. (1995) Local indicators of spatial association – LISA. *Geographical Analysis* 27:93–115.

Barrai, I., Scapoli, C., Beretta, M., Nesti, C., Mamolini, E. & Rodriguez-Larralde, L. (1996) Isonymy and the genetic structure of Switzerland. I. The distribution of surnames. *Ann Hum Biol* 23, 431–455.

Barrai, I., Rodríguez-Larralde, A., Mamolini, E., Manni, F. & Scapoli, C. (2000) Elements of the surname structure of Austria. *Ann Hum Biol* 27, 607-622.

Barrai, I., Rodríguez-Larralde, A., Manni, F., Ruggiero, V., Tratari, D., Scapoli, C. (2003) Isolation by Language and Distance in Belgium. *Ann Hum Biol* 68, 1-16.

Beemsterboer, S.N., Homburg, R., Gorter, N.A., Schats, R., Hompes, P.G. & Lambalk, C.B. (2006) The paradox of declining fertility but increasing twinning rates with advancing maternal age. *Hum Reprod* 21, 1531-1532.

Bronberg, R., Dipierri, J.E., Alfaro, E.L., Barrai, I., Rodríguez Larralde, A., Castilla, E., Colonna, V., Rodríguez Arroyo, G. & Bailliet, G. (2009) Isonymy structure of Buenos Aires city. *Hum Biol* 8, 447-461.

Cheshire, J. A., Longley, P.A. & Singleton, A.D. (2010) The surname regions of Great Britain. *J MAPS*, 401-409.

Crews D. E. & Bindon J. R. (1991) Ethnicity as a taxonomic tool in biomedical research. *Ethnic. Dis.* 1, 42–49.

Crow, J.E. & Mange, A.P. (1965) Measurements of inbreeding from the frequency of marriages between persons of the same surnames. *Eugen Q* 12, 190-203.

Datasus. (2010) <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinasc/cnv/nvrs.def> [Acessado em Setembro de 2013]

De Oliveira, Marcelo Zagonel ; Schüler-Faccini, Lavínia ; Demarchi, Dario A. ; Alfaro, Emma L. ; Dipierri, José E. ; Veronez, Mauricio R. ; Colling Cassel, Marlise ; Tagliani-Ribeiro, Alice ; Silveira Matte, Ursula ; Ramallo, Virginia . So Close, So Far Away: Analysis of Surnames in a Town of Twins (Cândido Godói, Brazil). *Annals of Human Genetics*, v. 77, p. 125-136, 2013.

Dipierri, J. E., Alfaro, E. L., Scapoli, C, Mamolini, E., Rodriguez-Larralde, A. & Barraí, I. (2005) Surnames in Argentina: a population study through isonymy. *Am J Phys Anthropol* 28, 199–299.

Dipierri, J.E., Rodríguez-Larralde, A., Alfaro, E.L. & Barraí, I. (2007) Isonymy structure of the Argentine northwest. *Ann Hum Biol* 34, 498–503.

Graf, O.M., Zlojutro, M., Rubicz, R. & Crawford, M.H. (2010) Surname distributions and their association with Y-Chromosome markers in the Aleutian Islands. *Hum Biol* 82, 745-757.

Guglielmino, C.R., Zei, G. & Cavalli-Sforza, L.L. (1991) Genetic and cultural transmission in Sicily as revealed by names and surnames. *Hum Biol* 63, 607-627.

Hall, J.G. (1996) Twinning: mechanisms and genetic implications. *Curr Opin Genet Dev* 6, 343-347.

Hasbargen, U., Lohse, P. & Thaler, C.J. (2000) The number of dichorionic twin pregnancies is reduced by the common MTHFR 677C-->T mutation. *Hum Reprod* 15, 2659-2662.

Hoekstra, C., Zhao, Z.Z., Lambalk, C.B., Willemsen, G., Martin, N.G., Boomsma, D.I. & Montgomery, G.W. (2008) Dizygotic twinning. *Hum. Reprod. Update* 14, 37–47.

IBGE. Disponível em: <http://www.ibge.gov.br>. Acessado entre outubro e setembro/2013.

Jobling, M.A. (2001) In the name of the father: surnames and genetics. *Trends Genet* 17, 353–357.

Rodriguez-Larralde, A. (1990) Distribucion de los apellidos y su uso en la estimacion de aislamiento y sedentarismo en los municipios del Estado Lara, Venezuela. *Acta Cientifica Venezolana* 41: 163.170.

Lummaa, V., Haukioja, E., Lemmetyinen, R. & Pikkola, M. (1998). Natural selection on human twinning. *Nature* 394, 533-534.

Matte, U., Le Roux, M.G., Bénichou, B., Moisan, J.P. & Giugliani, R. (1996) Study on possible increase in twinning rate at a small village in south Brazil. *Acta Genet Med Gemellol* 45, 431-437.

Manni, F., and I. Barrai. 2001. Genetic structures and linguistic boundaries in Italy: A microregional approach. *Hum. Biol.* 73:335–347.

Montgomery, G.W., Zhao, Z.Z., Marsh, A.J., Mayne, R., Treloar, S.A., James, M., Martin, N.G., Boomsma, D.I. & Duffy, D.L. (2004) A deletion mutation in GDF9 in sisters with spontaneous DZ twins. *Twin Res* 7, 548-555.

Rambo, A.B. (1988) Associativismo teuto-brasileiro e os primórdios do cooperativismo no Brasil. *Persp Econ* 23, 263-276.

Rodriguez-Larralde, A. (1990) Distribucion de los apellidos y su uso en la estimacion de aislamiento y sedentarismo en los municipios del Estado Lara, Venezuela. *Acta Cientifica Venezolana* 41: 163.170.

Rodríguez-Larralde, A., Formica G., Scapoli C., Baretta, M., Mamolini, E. & Barraí, I. (1993) Microevolution in Perugia: isonymy, 1890–1990. *Ann Hum Biol* 20, 261–274.

Scapoli, C., Mamolini, E., Carrieri, A., Rodríguez Larralde, A. & Barraí, I. (2006) Surnames in Western Europe: a comparison of the subcontinental populations through isonymy. *Theor Popul Biol* 71, 37-48.

Stein, M.L. (1999) *Interpolation of Spatial Data*. Springer-Verlag, New York, Inc.

Steinman, G. (2006) Can the chance of having twins be modified by diet? *Lancet* 367, 1513-1519.

Tagliani-Ribeiro, A., Oliveira, M., Sassi, A.K., Rodrigues, M.R., Zagonel-Oliveira, M., Steinman, G., Matte, U., Fagundes, N.J.R. & Schuler-Faccini, L. (2011) Twin town in south Brazil: a nazi's experiment or a genetic founder effect? *PLoS ONE* 6, e20328.

Tagliani-Ribeiro, A ; Paskulin, D.D ; Oliveira, M. ; Zagonel-Oliveira, M. ; Longo, D. ; Ramallo, V. ; Ashton- Prolla, P. ; Saraiva-Pereira, M.L ; Fagundes, N. J. R. ; Schuler-Faccini, L. ; Matte, U. (2012) . High twinning rate in Candido Godoi: a new role for p53 in human fertility.. *Human Reproduction* (Oxford. Print), v. 27, p. 2866-2871, 2012.

Tarskaia, L., El'chinova G. I., Scapoli C., Mamolini E., Carrieri A., Rodríguez-Larralde A. & Barraí I. (2008) Surnames in Siberia: a study of the population of Yakutia through isonymy. *Am J Phys Anthropol* 138, 190–198.

Tabela 1 – Parâmetros de isonomia calculados em Santo Cristo

| ID | Nome | N | S | Alfa | I | F | A | B |
|----|---------------------------|------|-----|------|-------|-------|------|------|
| 1 | Linha Seca Alta | 61 | 23 | 12 | 0,083 | 0,021 | 19,7 | 63,9 |
| 2 | Linha Arnaldo | 263 | 49 | 18 | 0,056 | 0,014 | 5,7 | 50,6 |
| 3 | Linha Rolador Alto | 212 | 48 | 19 | 0,054 | 0,013 | 8 | 47,6 |
| 4 | Linha Torta | 121 | 32 | 19 | 0,052 | 0,013 | 8,3 | 54,5 |
| 5 | Linha Saturno | 146 | 46 | 21 | 0,047 | 0,012 | 15,1 | 51,4 |
| 6 | Linha Doze de Maio | 125 | 35 | 22 | 0,046 | 0,012 | 12,8 | 51,2 |
| 7 | Linha Castor | 210 | 51 | 26 | 0,039 | 0,01 | 8,6 | 43,8 |
| 8 | Linha Bom Fim | 221 | 56 | 27 | 0,037 | 0,009 | 8,2 | 40,7 |
| 9 | Linha Guaraipo | 248 | 49 | 27 | 0,038 | 0,009 | 3,6 | 43,6 |
| 10 | Linha Bernardo | 72 | 28 | 28 | 0,036 | 0,009 | 13,9 | 48,6 |
| 11 | Rolador Baixo | 242 | 69 | 29 | 0,035 | 0,009 | 14,5 | 44,2 |
| 12 | Linha Venus | 240 | 57 | 31 | 0,032 | 0,008 | 5,4 | 35,4 |
| 13 | Lajeado Vieira | 153 | 50 | 32 | 0,031 | 0,008 | 9,2 | 37,9 |
| 14 | Linha Larga | 229 | 47 | 33 | 0,03 | 0,008 | 3,5 | 37,1 |
| 15 | Linha Salto | 262 | 58 | 35 | 0,028 | 0,007 | 7,6 | 35,9 |
| 16 | Belinha Centro | 198 | 62 | 38 | 0,026 | 0,007 | 11,6 | 35,4 |
| 17 | Alma | 263 | 72 | 39 | 0,026 | 0,006 | 10,6 | 34,2 |
| 18 | Linha Bom Princípio Baixo | 388 | 94 | 39 | 0,026 | 0,006 | 7,7 | 32,5 |
| 19 | Herval Novo | 149 | 53 | 40 | 0,025 | 0,006 | 16,1 | 36,2 |
| 20 | Linha Orion | 139 | 49 | 43 | 0,024 | 0,006 | 13 | 32,4 |
| 21 | Linha Divisa | 279 | 83 | 47 | 0,021 | 0,005 | 13,6 | 29,7 |
| 22 | Linha Taquaraçú | 351 | 85 | 54 | 0,018 | 0,005 | 7,1 | 26,5 |
| 23 | Linha Dona Belinha | 433 | 93 | 57 | 0,018 | 0,004 | 6,2 | 25,2 |
| 24 | Vila Sirio | 680 | 130 | 63 | 0,016 | 0,004 | 5 | 26 |
| 25 | Linha Bom Principio Alto | 277 | 84 | 68 | 0,015 | 0,004 | 7,6 | 23,5 |
| 26 | Linha Laranjeira | 390 | 108 | 76 | 0,013 | 0,003 | 8,7 | 21,8 |
| 27 | Linha Revolta | 935 | 273 | 155 | 0,006 | 0,002 | 10,8 | 12,9 |
| 28 | Centro | 4915 | 901 | 309 | 0,003 | 0,001 | 6,6 | 7,9 |

ID por localidades, nomes, numero de eleitores N, numero de sobrenomes diferentes S, estimador A e estimador B.

Tabela 2 – Parâmetros de isonomia calculados em Cândido Godói.

| ID | Nome | N | S | Alfa | I | F | A | B |
|----|----------------|------|-----|------|-------|-------|------|------|
| 1 | Doze Norte | 159 | 28 | 13 | 0,073 | 0,02 | 2,6 | 62,3 |
| 2 | Abrantes | 128 | 24 | 14 | 0,069 | 0,019 | 5,5 | 63,3 |
| 3 | Castelo Branco | 61 | 19 | 16 | 0,062 | 0,02 | 9,8 | 67,2 |
| 4 | Secção C | 138 | 30 | 18 | 0,054 | 0,015 | 7,2 | 56,5 |
| 5 | Paranagua | 179 | 36 | 19 | 0,052 | 0,014 | 5,6 | 52,5 |
| 6 | Natal | 147 | 36 | 20 | 0,049 | 0,014 | 7,5 | 49,7 |
| 7 | São Bonifácio | 113 | 38 | 23 | 0,043 | 0,013 | 17,7 | 52,2 |
| 8 | São João | 107 | 33 | 24 | 0,042 | 0,013 | 10,3 | 52,3 |
| 9 | La Salle | 101 | 31 | 25 | 0,039 | 0,012 | 10,9 | 54,5 |
| 10 | Dr, Pederneira | 143 | 35 | 26 | 0,038 | 0,011 | 4,9 | 44,1 |
| 11 | Dos Louros | 156 | 43 | 27 | 0,036 | 0,011 | 9,6 | 51,3 |
| 12 | São Pedro | 214 | 45 | 29 | 0,034 | 0,008 | 3,7 | 41,6 |
| 13 | Boa Vista | 155 | 49 | 30 | 0,033 | 0,01 | 9 | 43,9 |
| 14 | Acre | 213 | 50 | 32 | 0,031 | 0,009 | 4,2 | 40,4 |
| 15 | Dr,P, Toledo | 159 | 41 | 36 | 0,027 | 0,008 | 5,7 | 39 |
| 16 | Esquina União | 167 | 46 | 37 | 0,027 | 0,008 | 9 | 41,3 |
| 17 | Treze de Maio | 114 | 33 | 37 | 0,026 | 0,009 | 9,6 | 48,2 |
| 18 | Silva Jardim | 274 | 56 | 38 | 0,026 | 0,008 | 2,9 | 35 |
| 19 | Secção B | 134 | 41 | 39 | 0,026 | 0,008 | 7,5 | 37,3 |
| 20 | São Miguel | 223 | 49 | 40 | 0,025 | 0,007 | 4 | 37,7 |
| 21 | Secção A | 217 | 57 | 43 | 0,023 | 0,007 | 8,8 | 38,7 |
| 22 | Timbauva | 342 | 88 | 71 | 0,014 | 0,004 | 6,4 | 26,3 |
| 23 | Centro | 1672 | 368 | 183 | 0,005 | 0,002 | 7,3 | 12,7 |

ID por localidades, nomes, numero de eleitores N, numero de sobrenomes diferentes S, estimador A e estimador B

Tabela 3 – Parâmetros de isonomia calculados em Cerro Largo

| ID | Nome | N | S | Alfa | I | F | A | B |
|----|-----------------------|------|------|------|-------|-------|-------|-------|
| 1 | Santa Barbara | 167 | 48 | 9 | 0,115 | 0,029 | 15,56 | 60,47 |
| 2 | São João Centro | 137 | 30 | 13 | 0,076 | 0,019 | 8,02 | 64,96 |
| 3 | Caçador | 150 | 35 | 13 | 0,079 | 0,02 | 6,6 | 60 |
| 4 | Santa Fé | 53 | 17 | 15 | 0,068 | 0,017 | 7,54 | 64,15 |
| 5 | Marreca | 121 | 29 | 18 | 0,056 | 0,014 | 2,48 | 56,2 |
| 6 | Poço Preto | 48 | 20 | 19 | 0,054 | 0,014 | 16,66 | 62,5 |
| 7 | Vila Tremonia | 285 | 56 | 21 | 0,047 | 0,012 | 8,07 | 52,28 |
| 8 | Santa Cruz | 175 | 50 | 26 | 0,038 | 0,009 | 12,57 | 43,43 |
| 9 | Esquina Sandri Moscon | 147 | 42 | 28 | 0,036 | 0,009 | 10,9 | 43,54 |
| 10 | Taquarucu | 173 | 52 | 28 | 0,036 | 0,009 | 13,3 | 44,5 |
| 11 | Atolosa | 220 | 53 | 33 | 0,031 | 0,008 | 7,72 | 37,28 |
| 12 | Vila Santo Antônio | 284 | 68 | 36 | 0,028 | 0,007 | 7,04 | 34,86 |
| 13 | Vila São Francisco | 323 | 70 | 37 | 0,027 | 0,007 | 5,57 | 34,05 |
| 14 | São Joao Norte | 212 | 52 | 38 | 0,026 | 0,007 | 4,71 | 35,84 |
| 15 | Centro | 7143 | 1042 | 218 | 0,005 | 0,001 | 4,43 | 11,85 |

ID por localidades, nomes, numero de eleitores N, numero de sobrenomes diferentes S, estimador A e estimador B

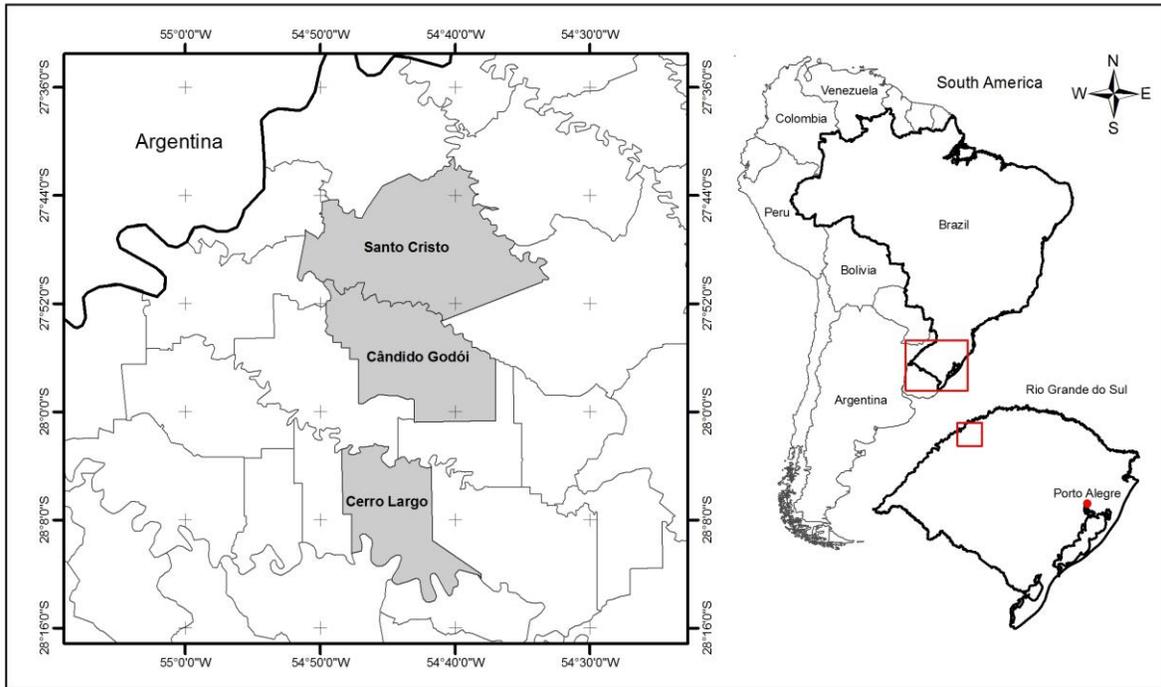


Figura 1 – Mapa de localização dos municípios de Santo Cristo, Cândido Godói e Cerro Largo.

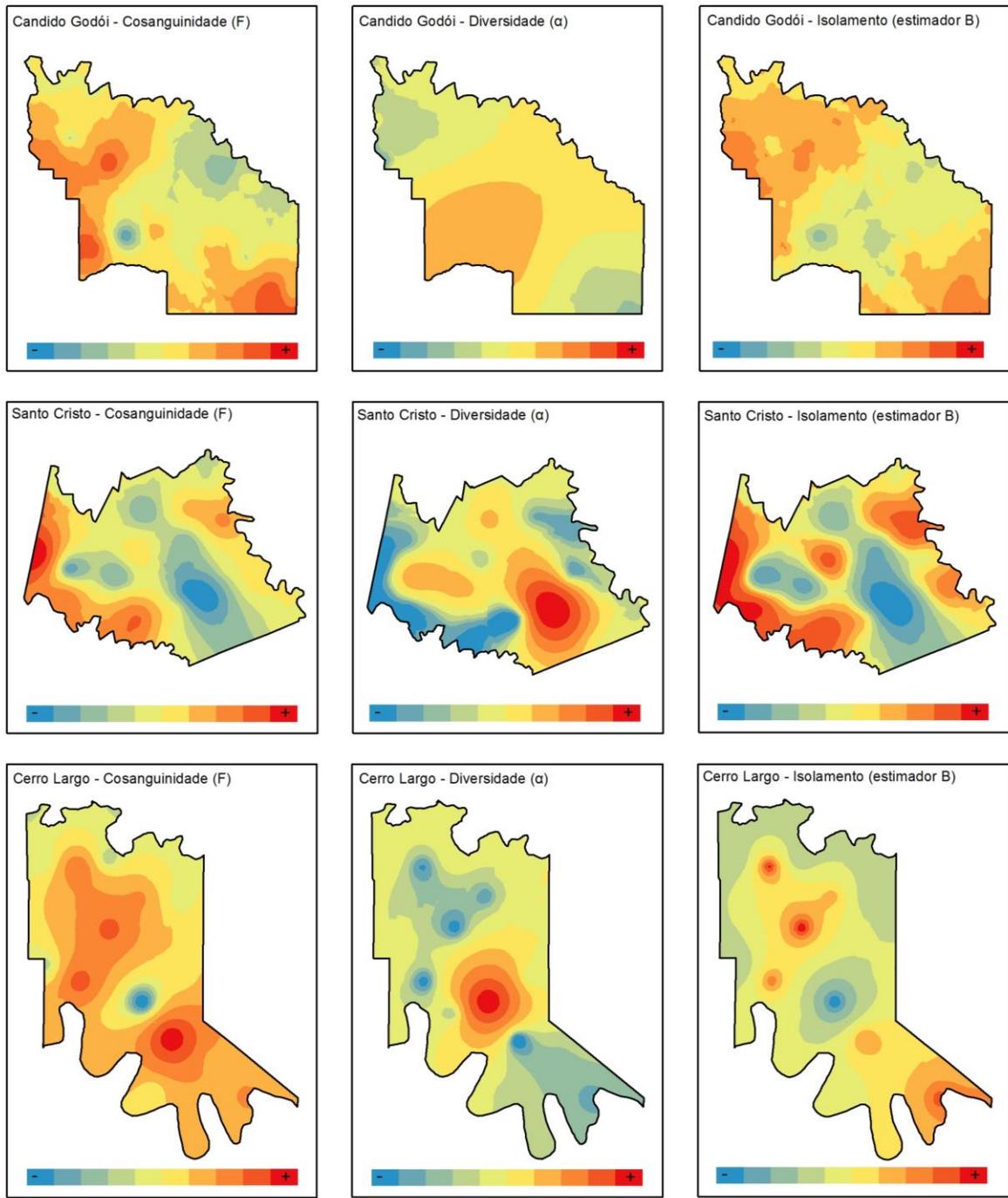


Figura 2: Índices de consanguinidade (f), diversidade (alfa) e estimador B, interpolados utilizando o método de Krigagem nos três municípios estudados.

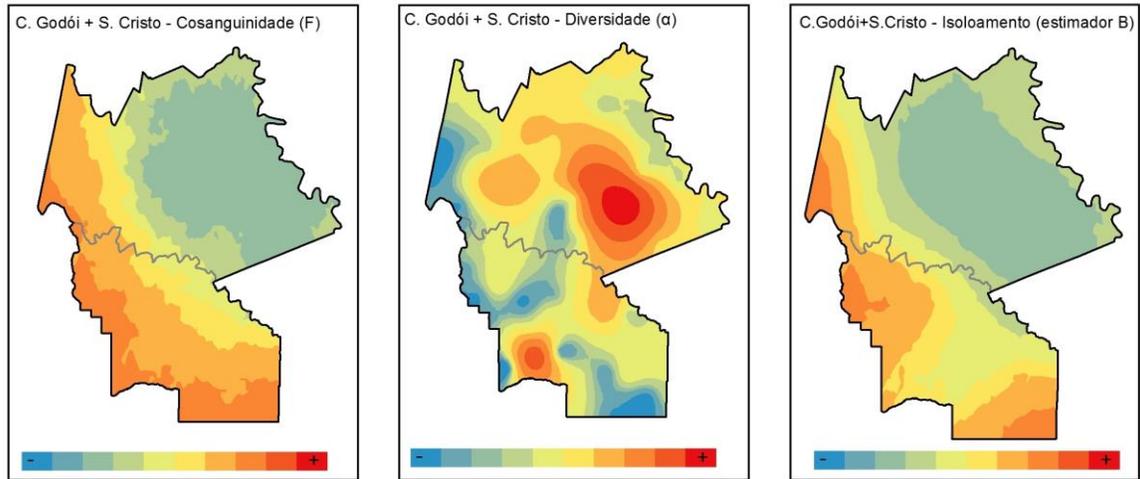


Figura 3 – Análise em conjunto dos índices de consanguinidade (F), diversidade (alfa) e isolamento (estimador B) para os municípios de Santo Cristo (S) e Cândido Godói (N), utilizando o método de krigagem.

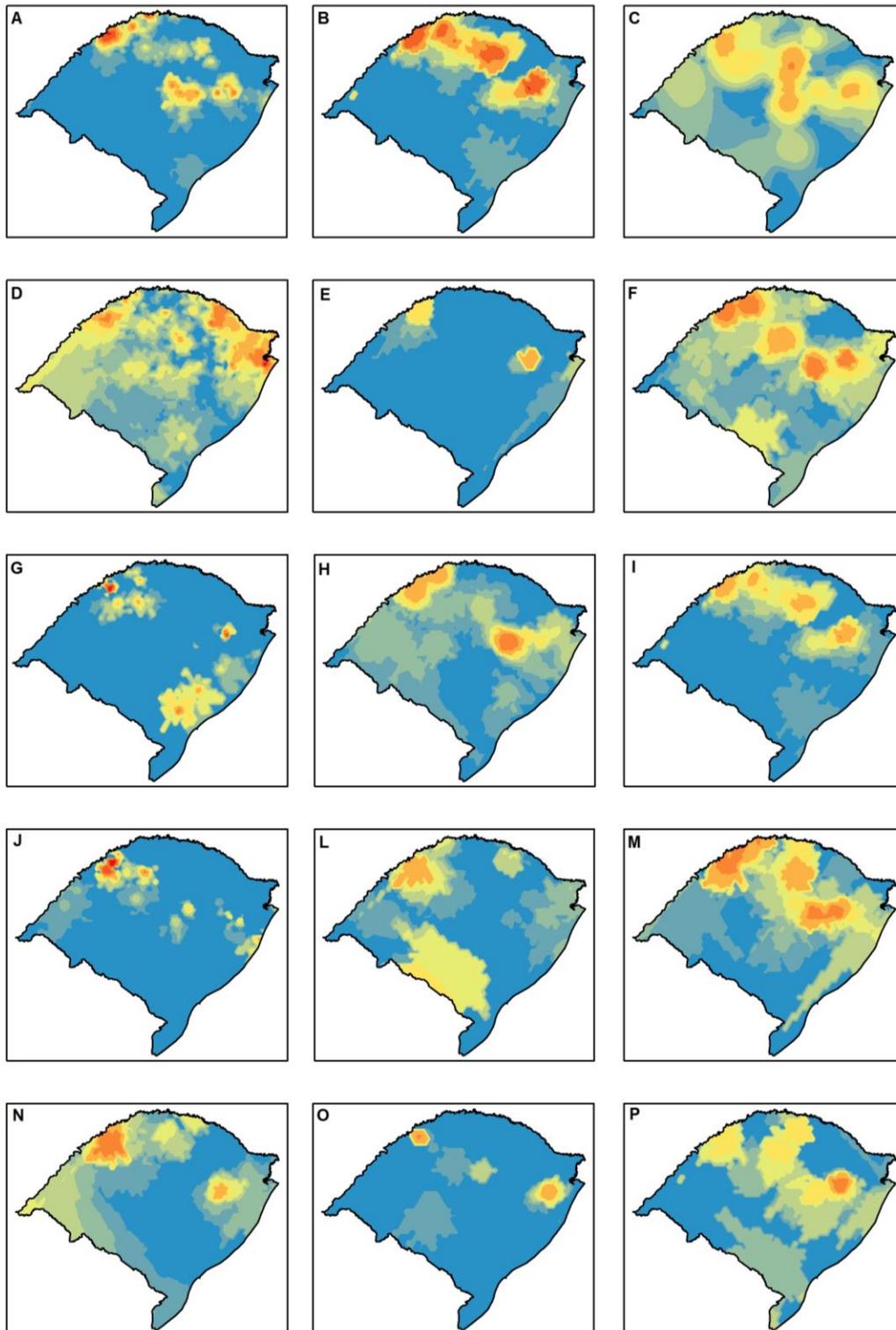


Figura 4 – Distribuição dos 18 sobrenomes com maior frequência de nascimentos gemelares, utilizando o método de krigagem. Azul representa nenhuma ocorrência e vermelho maior ocorrência.

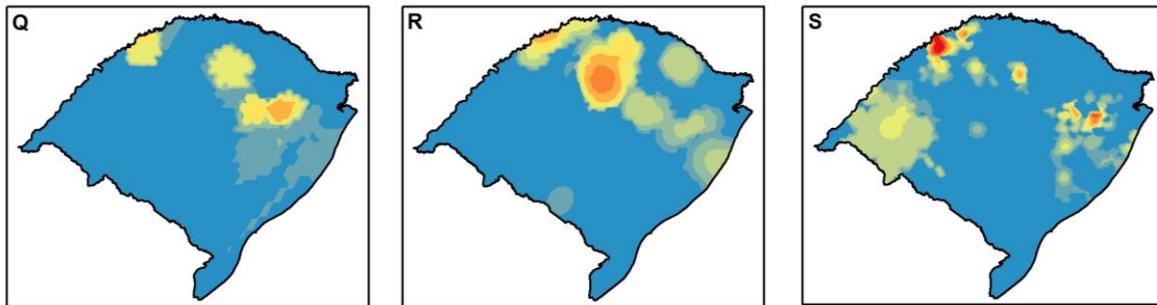


Figura 4 – Distribuição dos 18 sobrenomes com maior frequência de nascimentos gemelares, utilizando o método de krigagem. Azul representa nenhuma ocorrência e vermelho maior ocorrência. Continuação

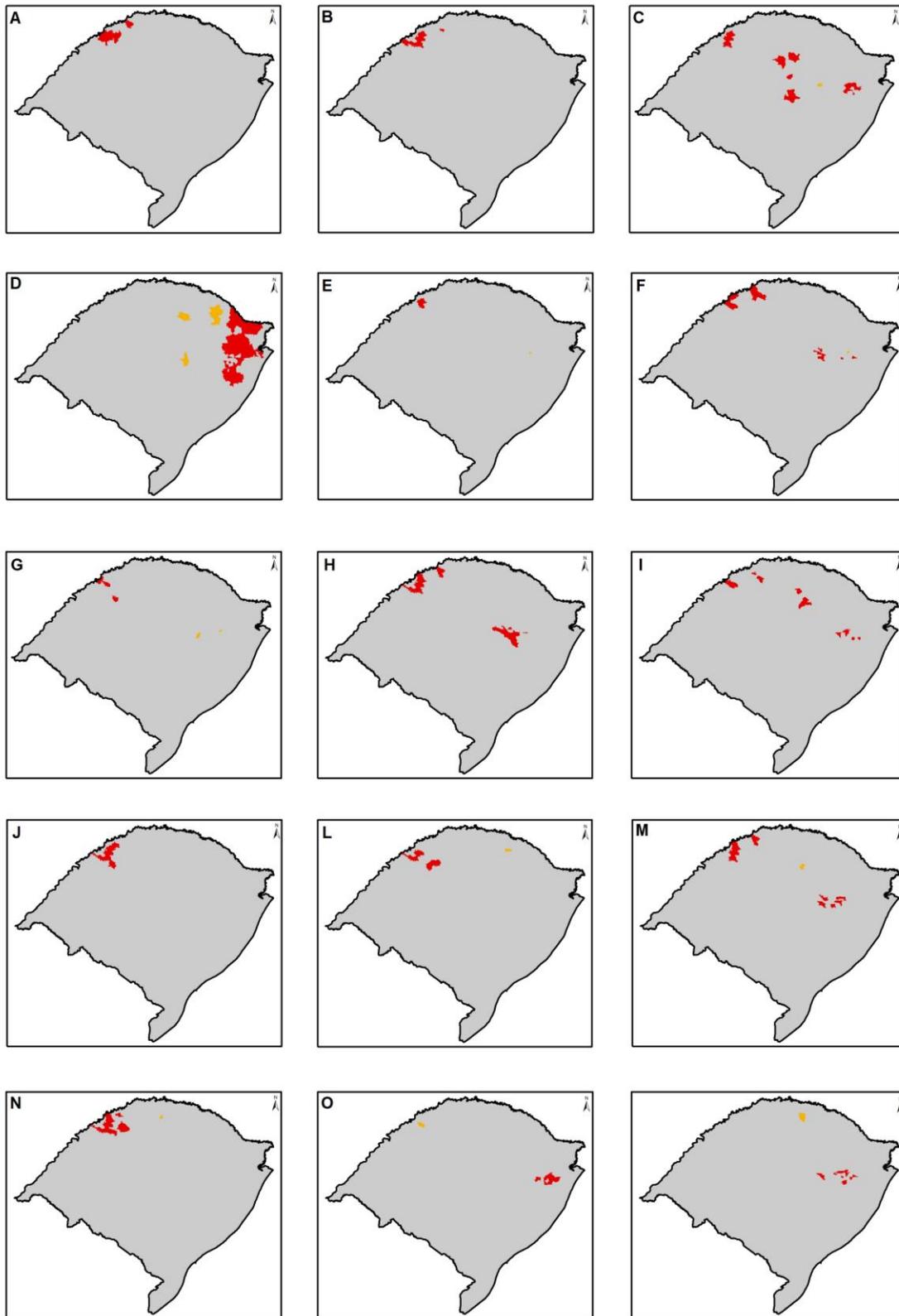


Figura 5 – Análise de Cluster dos 18 sobrenomes com maior frequência de nascimentos gemelares.

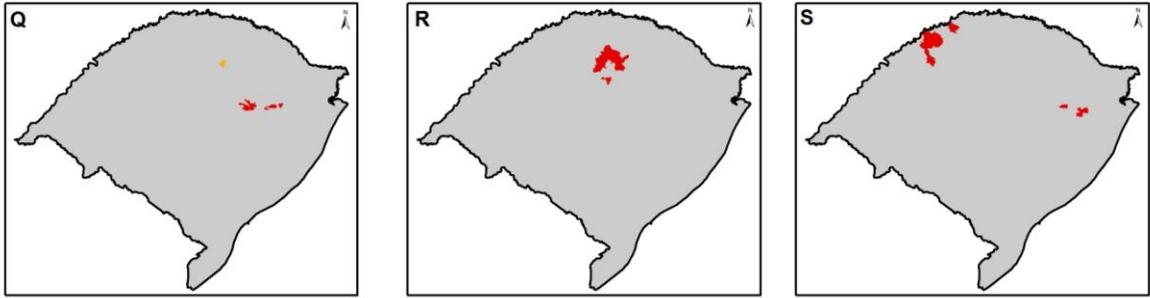


Figura 5 – Análise de Cluster dos 18 sobrenomes com maior frequência de nascimentos gemelares. Continuação.

A utilização de sobrenomes e geoprocessamento como indicadores de estrutura populacional

Os resultados da pesquisa combinando análise de sobrenomes e geoprocessamento nesta pequena região do noroeste do Rio Grande do Sul revelam uma clara divisão em duas partes: um conjunto populacional com maior grau de isolamento o noroeste e outro com menor grau de isolamento no sudeste. Esta distribuição não está vinculada ou afetada pelas divisas municipais atuais, sendo estes meros limites administrativos. Sendo assim, este fenômeno poderia ser melhor explicado pelo processo de imigração que deu origem a estas populações. No século passado, a iniciativa privada criou colônias homogêneas, com a mesma orientação étnico-confessional (De Oliveira et al. ,2013). Ainda hoje, o processo de imigração parece ter ficado cristalizado nas regiões rurais, criando barreiras que subdividem a região e diminuem o fluxo gênico.

Os sobrenomes das famílias que possuem ocorrência de gêmeos apresentam uma distribuição muito restrita no estado. Os *clusters* identificados coincidem com as zonas de colonização alemã. No Rio Grande do Sul, os sobrenomes ainda guardam uma relação próxima com a ancestralidade europeia, podendo ser usados como indicadores em estudos de genética de populações ou genética medica populacional.

Em uma análise micro histórica do processo de imigração, observou-se que o povoamento nestes municípios surgiu a partir do conflito de duas instituições religiosas. Enquanto a igreja protestante avançava em todo o mundo (inclusive no estado do Rio Grande do Sul), a igreja católica tentava barrar este avanço, principalmente na América Latina. Estas circunstâncias fizeram que o catolicismo alemão voltasse suas ações para os alemães católicos que viviam fora da Alemanha; surgindo assim, em 1911, em Dresden, a “União para os Alemães Católicos do Exterior” - *Katholisches Deustchtum im Auslande* (Gertz, 1992). Conforme Schallenberger (2001), esta organização se assentou nos objetivos de 1) dar assistência viva na educação entre a pátria (Alemanha) e os colonos católicos que vivem no exterior; 2) zelar pela manutenção e pelo fomento da

língua, dos costumes, da cultura e da religião; 3) Fornecer subsídios para o desenvolvimento espiritual e material dos alemães que estiverem distantes de sua pátria. Esses objetivos figuravam a busca de um elo de coesão social onde o elemento identificador da unidade esteve envolto em torno do conceito de cultura, neste caso consignado em base aos eixos étnico-linguístico e religioso (Nascimento, 1999).

Ao longo dos anos, estas características continuaram agindo como elementos de coesão e, simultaneamente, como obstáculos à miscigenação. No nosso estudo observou-se que a religião católica é predominante em Candido Godói, declarada por 71% da população (IBGE, 2010) estando as diferentes religiões claramente separadas dentro do município (ver figura 6 do Artigo 1, capítulo 4). O mesmo papel tiveram as diversas línguas ou dialetos falados desde a formação da colônia. Na figura 5 do Artigo 1 (Capítulo 4) estão apresentadas as barreiras que isolam os locais onde os sobrenomes de origem polonesa são mais frequentes, ficando separados do resto da população com maioria de sobrenomes de origem alemã. Este pequeno exemplo a nível muito regional correlaciona-se aos achados em outros estudos sobre isonimia na Europa.

Nos países onde a imigração tem sido relativamente baixa nos últimos séculos, os sobrenomes estão estritamente ligados à diferenças linguísticas (Barrai et al. 2000). O poder de coesão do idioma é marcante ainda em migrações não planejadas. Vienna et al.(2001) estudaram a isonimia dos descendentes de refugiados da Albânia, Croácia e Grécia, que hoje habitam várias aldeias em onze estados do sul da Itália e Sicília. Alguns destes movimentos migratórios ocorreram há muito tempo, nos séculos XIV e XV, outros durante o período pós-segunda guerra mundial. Todas as comunidades resultantes consideram-se minorias étnicas, caracterizadas pela sua própria língua, cultura, história e localização geográfica em relação à população italiana circundante. No entanto, é importante notar que a linguagem falada representa hoje o mais importante, e em muitos casos o único elemento culturalmente característico. Essa particularidade é achada também nos sobrenomes, tendo um sistema de escrita diferenciado dos sobrenomes italianos.

O processo de imigração alemã no Rio Grande do Sul não pode analisar-se como sendo uniforme ao longo do espaço e do tempo. As dinâmicas demográficas estão influenciadas pela economia do estado e da nação durante os diferentes períodos. O desenvolvimento das pequenas e grandes indústrias regionais podem abrir mercados de trabalho, atrativos ainda para a população mais afastada do foco econômico. Mas as possibilidades de migrar estarão limitadas pelos contextos sociais. A receptividade no local do destino pode ser uma variável que garanta o sucesso ou o fracasso da empresa. Os atuais municípios derivados das colônias privadas mantêm o perfil homogêneo desde sua origem, registrando-se nas populações de Candido Godoi e Santo Cristo pouco crescimento demográfico e, depois da Segunda Guerra Mundial, certa relutância as mobilizações e deslocamentos internos. Na figura 5 do Artigo 2 (Capítulo 5) vemos que a distribuição dos dezoito sobrenomes com maior frequência de gêmeos não é ao acaso, mas sim um reflexo das histórias familiares. Os clusters registrados nos falam da continuidade no tempo destes conjuntos populacionais, reunidos desde o século XIX dentro do contexto de coesão cultural dado pela mesma religião e a mesma língua. Cerro Largo, no entanto, passa por um período de expansão cujos resultados ainda não estão refletidos na variabilidade do registro de eleitores, com uma porcentagem importante dos sobrenomes pouco representados numericamente.

As mesmas ferramentas metodológicas empregadas até agora a nível microrregional para a análise isonímica, permitem fazer comparações a escala macrorregional. Os estados de Rio Grande do Sul, Santa Catarina e Paraná formam a denominada Região Sul do Brasil. É a menor em superfície, ocupando cerca de 6,8% do território nacional, mas é a segunda mais desenvolvida economicamente e sua população total é três vezes maior que o número de habitantes das regiões Norte e Centro-Oeste (IBGE, 2010). Forma um conjunto geográfico, climático e populacional diferenciado do resto do Brasil. É a única região localizada fora de zona tropical, registrando as menores temperaturas durante o inverno e a maior amplitude climática ao longo do ano. Sua história foi fortemente marcada pela imigração europeia e o desenvolvimento econômico

voltado à área agrícola e pecuária. As terras do norte e oeste do Paraná e do oeste de Santa Catarina foram as últimas regiões a serem povoadas, com a criação de colônias rurais. A Região Sul forma parte da fronteira exterior do Brasil, limitando ao sul com o Uruguai e a oeste com a Argentina e o Paraguai. Estudos de isonímia em estes dois últimos países (Dipierrri et al. 2005; 2011) permitiram fazer o quadro comparativo da tabela 1 e figura 1, onde estão apresentados o número de eleitores, a quantidade de sobrenomes diferentes e os índices alfa e índices F de consanguinidade, comparados entre os diferentes estados, “departamentos” ou “províncias”, todas unidades administrativas ao mesmo nível de hierarquia.

O índice alfa é uma medida direta do número efetivo de sobrenomes, encontrando valores baixos em localidades mais isoladas e apresentando uma correspondência inversa com o índice F, cujos valores geralmente são altos em populações pequenas, sem fluxo gênico nem migração. Este comportamento é muito claro nas “províncias” da Argentina, contrastando especialmente nos registros para Capital Federal, sede central do governo e principal cidade do país (ver Figura 2). A maior consanguinidade ocorre nas “províncias” de Santiago del Estero, Corrientes e La Rioja, fenômeno que poderia estar relacionado ao menor desenvolvimento econômico. Os índices para os “departamentos” do Paraguai também variam, mas sem apresentar contrastes tão marcantes como na Argentina. Interessantemente, a maior variabilidade de sobrenomes acontece no “departamento” de Boqueron (ver Figura 3), que também é o maior em superfície (91.669 km²), mas que possui baixa densidade demográfica (0,67 hab./km²). O clima da região é semiárido, o que dificulta o estabelecimento de economias produtivas e explicaria o baixo IDH (Índice de Desenvolvimento Humano) da região (0,487 PNUD 2011). O valor alto de Alfa pode ser um vestígio de um momento inicial de expansão populacional, que não foi logo seguido do crescimento demográfico. Os três estados do Brasil que formam a Região Sul apresentam, em comparação, os valores mais extremos do gráfico, com baixa variabilidade de sobrenomes e elevada consanguinidade, especialmente no estado do Paraná. No conjunto, os três registros reúnem 20.572.786 de eleitores,

uma amostra representativo do total de 28.795.762 habitantes da região (IBGE, censo 2010). Aqui, diferente das “províncias” da Argentina (Santiago del Estero, Corrientes e La Rioja), os altos valores de F estariam influenciados pelo fenômeno dos sobrenomes polifiléticos. Segundo Garza-Chapa et. al. (2010), os sobrenomes são designados como parafiléticos (que contém o ancestral comum e alguns, mas não todos, os descendentes), quando todos seus portadores podem ser considerados como aparentados ao mesmo grupo ou evento populacional, como os sobrenomes alemães ou italianos das migrações mais recentes. Os sobrenomes polifiléticos são aqueles de origem múltipla, tal como os sobrenomes portugueses, que durante o tempo da colônia imperial eram repetidamente introduzidos na região, tornando-se comuns ao longo dos séculos. O índice B serve para medir o impacto que tais sobrenomes têm na população geral, expressado como a porcentagem de eleitores que compartilham os sete sobrenomes mais frequentes. Para Santa Catarina, essa porcentagem foi de 17,49%, no Rio Grande do Sul de 19,8% e mais alto ainda para o Paraná, 24,7%. No total, 4.341.933 de pessoas, sendo Silva o sobrenome com maior número de portadores: 1.471.383.

Para conseguir compreender a dinâmica demográfica em nível de município, é necessário compreender o papel dos movimentos históricos e socioculturais dentro de um contexto microrregional. Os processos migratórios ocorridos em Cândido Godói, Cerro Largo e Santo Cristo tiveram características diferentes, especialmente marcantes e junto com o isolamento posterior, fortes implicações na estruturação genética das populações. Nossos achados apoiam os estudos anteriores (Tagliani et al., 2011; De Oliveira et al., 2013), estes subgrupos populacionais podem definir-se como cenários possíveis nos quais a deriva genética e o efeito fundador poderiam ter ocorrido.

Tabela 1 - Tabela de distribuição dos índices Alfa e F para Argentina (por estado), Paraguai (por estado) e Brasil (estados da região sul)

| PAIS | ESTADO | N | S | Alfa | (F x 10 ⁵) | REFERÊNCIA |
|---------------------|--------------------|-----------|---------|-------|------------------------|------------------------|
| ARGENTINA | Capital Federal | 2,552,359 | 175,39 | 907 | 110,2 | Dipierri et al. (2005) |
| | Santa Fé | 2,199,212 | 81,074 | 778 | 128,5 | Dipierri et al. (2005) |
| | Rio Negro | 283,874 | 29,644 | 641 | 155,9 | Dipierri et al. (2005) |
| | Chubut | 220,892 | 23,539 | 638 | 156,6 | Dipierri et al. (2005) |
| | Buenos Aires | 8,210,176 | 274,832 | 622 | 160,7 | Dipierri et al. (2005) |
| | La Pampa | 188,751 | 17,18 | 592 | 168,9 | Dipierri et al. (2005) |
| | Cordoba | 1,982,250 | 75,243 | 562 | 177,8 | Dipierri et al. (2005) |
| | Tierra de Fuego | 49,452 | 12,602 | 565 | 178,9 | Dipierri et al. (2005) |
| | Entre Rios | 721,791 | 32,138 | 548 | 182,2 | Dipierri et al. (2005) |
| | Neuquen | 237,979 | 24,188 | 492 | 203,2 | Dipierri et al. (2005) |
| | Santa Cruz | 120,835 | 14,305 | 438 | 228,3 | Dipierri et al. (2005) |
| | Mendoza | 927,366 | 37,274 | 436 | 229 | Dipierri et al. (2005) |
| | Misiones | 471,176 | 30,331 | 341 | 292,5 | Dipierri et al. (2005) |
| | Salta | 647,329 | 20,169 | 310 | 321,7 | Dipierri et al. (2005) |
| | Jujuy | 301,331 | 11,357 | 268 | 372,5 | Dipierri et al. (2005) |
| | San Juan | 400,949 | 11,325 | 258 | 386,8 | Dipierri et al. (2005) |
| | Chaco | 544,698 | 22,964 | 238 | 419 | Dipierri et al. (2005) |
| | Tucuman | 759,446 | 21,234 | 252 | 395,8 | Dipierri et al. (2005) |
| | Formosa | 240,643 | 13,615 | 238 | 419,4 | Dipierri et al. (2005) |
| | Catamarca | 174,852 | 6,394 | 235 | 424,8 | Dipierri et al. (2005) |
| San Luis | 240,545 | 16,669 | 223 | 447,7 | Dipierri et al. (2005) | |
| Santiago del Estero | 502,57 | 12,296 | 197 | 507,3 | Dipierri et al. (2005) | |
| Corrientes | 539,063 | 19,523 | 178 | 560,3 | Dipierri et al. (2005) | |
| La Rioja | 148,656 | 7,259 | 174 | 572,8 | Dipierri et al. (2005) | |
| BRASIL | Santa Catarina | 4681005 | 93691 | 138 | 181,84 | Presente estudo |
| | Rio Grande do Sul | 8259308 | 97079 | 101 | 250 | Presente estudo |
| | Paraná | 7632473 | 145595 | 71 | 351,757 | Presente estudo |
| PARAGUAI | Boqueron | 30,491 | 2777 | 214,8 | 117 | Dipierri et al. (2011) |
| | Asunción (Capital) | 600,599 | 12,521 | 188,4 | 132 | Dipierri et al. (2011) |
| | Central | 1,240,855 | 11,479 | 156,5 | 159 | Dipierri et al. (2011) |
| | Alto Paraguay | 13,502 | 938 | 149,4 | 168 | Dipierri et al. (2011) |
| | Pdte. Hayes | 74,325 | 2815 | 147,1 | 169 | Dipierri et al. (2011) |
| | Alto Parana | 472,428 | 10,96 | 144,9 | 171 | Dipierri et al. (2011) |
| | Neembucu | 84,647 | 1494 | 145,2 | 171 | Dipierri et al. (2011) |
| | Concepcion | 168,159 | 2425 | 143,3 | 173 | Dipierri et al. (2011) |
| | Paraguari | 230,595 | 2608 | 142,4 | 174 | Dipierri et al. (2011) |
| | Itapua | 392,877 | 9998 | 139,1 | 179 | Dipierri et al. (2011) |
| | Misiones | 114,117 | 2070 | 134,2 | 185 | Dipierri et al. (2011) |
| | Cordillera | 255,835 | 3094 | 130,4 | 190 | Dipierri et al. (2011) |
| | Amambay | 95,993 | 2344 | 129 | 193 | Dipierri et al. (2011) |
| | Canindeyu | 111,652 | 4462 | 128,9 | 193 | Dipierri et al. (2011) |
| | Caaguazu | 366,958 | 4327 | 120,1 | 206 | Dipierri et al. (2011) |
| | San Pedro | 278,217 | 3432 | 119,6 | 207 | Dipierri et al. (2011) |
| | Caazapa | 134,435 | 2039 | 101,1 | 245 | Dipierri et al. (2011) |
| Guaira | 178,183 | 2675 | 96,7 | 256 | Dipierri et al. (2011) | |

N = Numero de eleitores / S = Número de sobrenomes

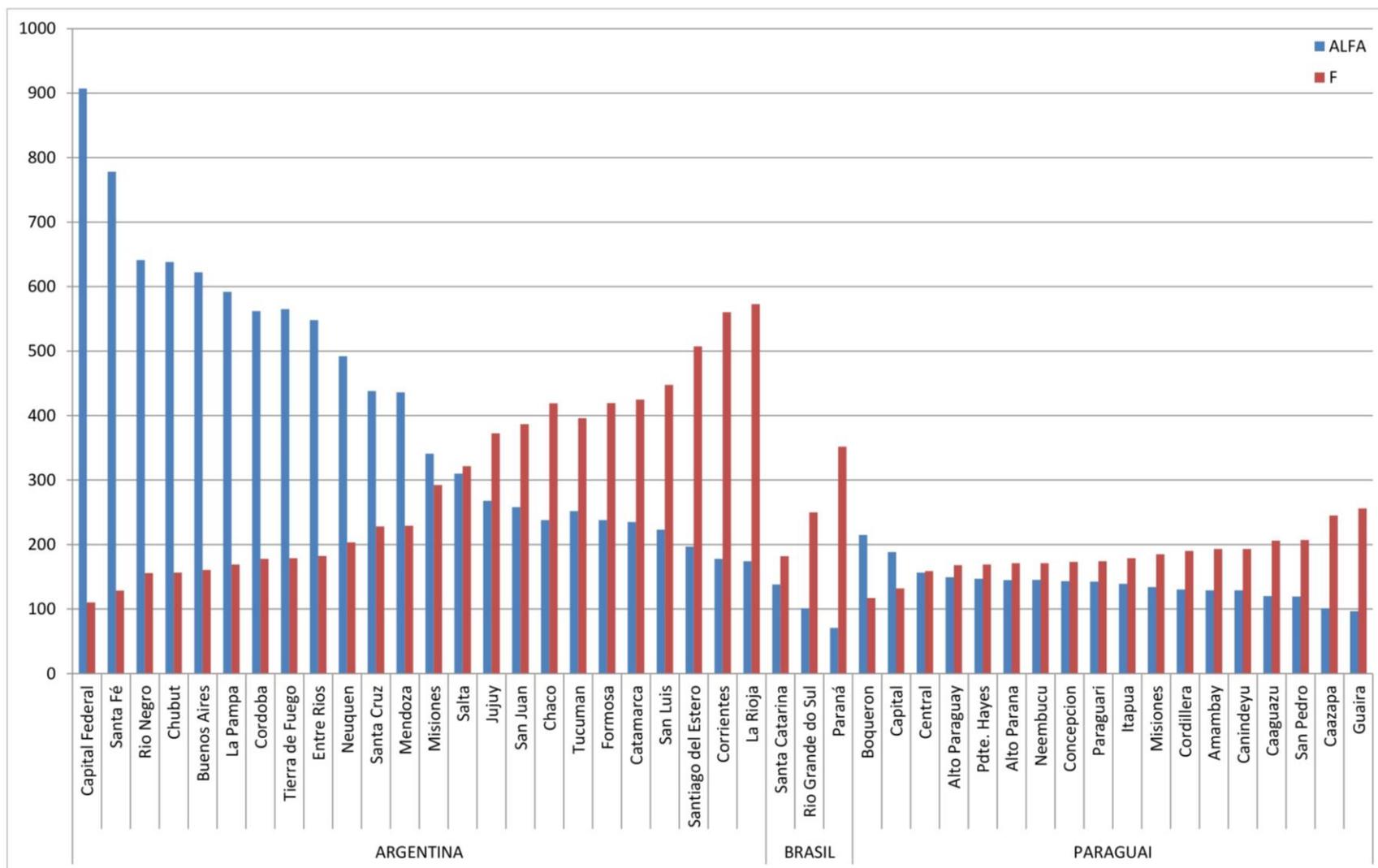


Figura 1 - Gráfico de distribuição dos índices Alfa e F para Argentina (por estado), Paraguai (por estado) e Brasil (estados da região Sul)

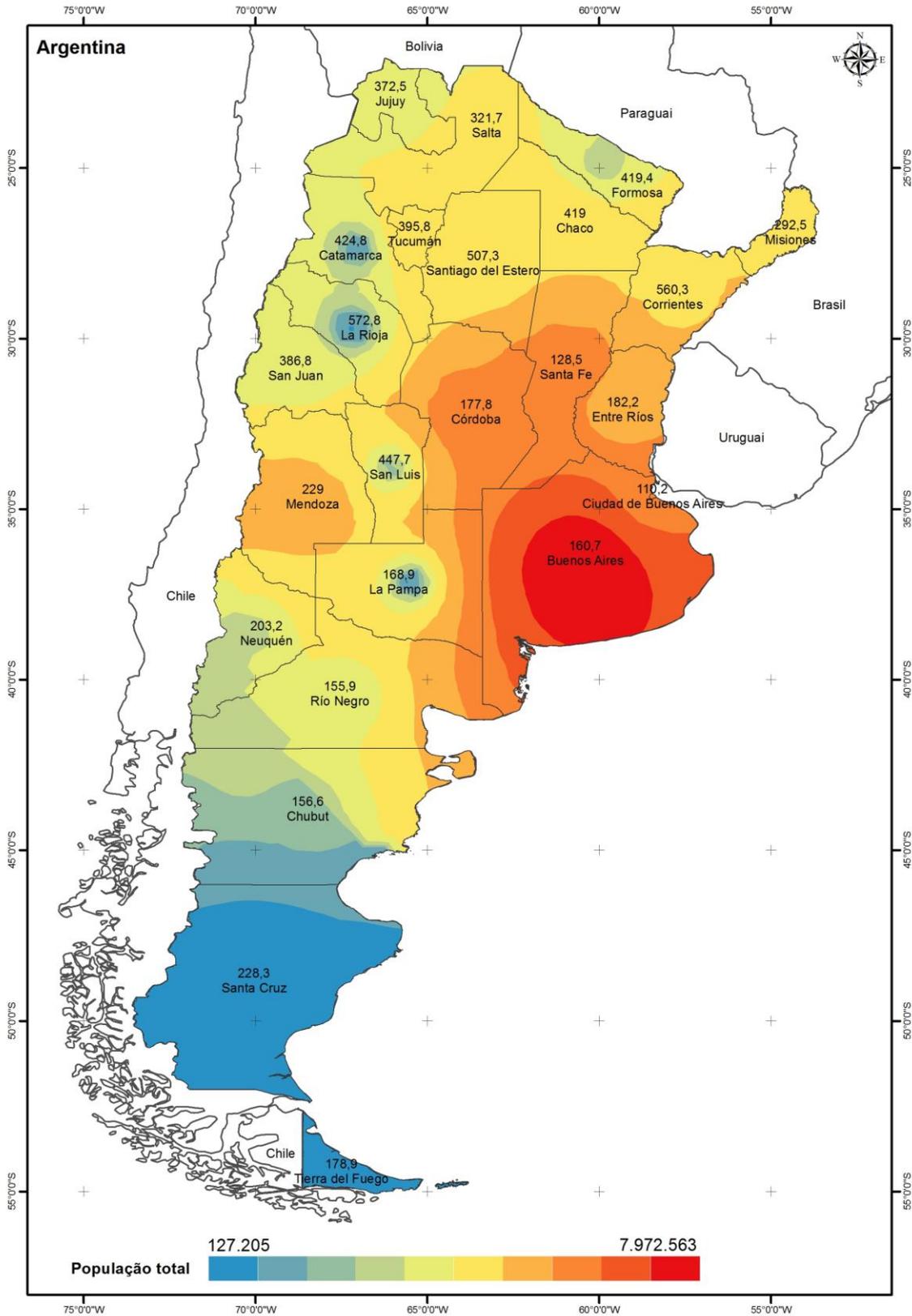


Figura 2 – Mapa de distribuição da população total e índice F, para as províncias da Argentina.

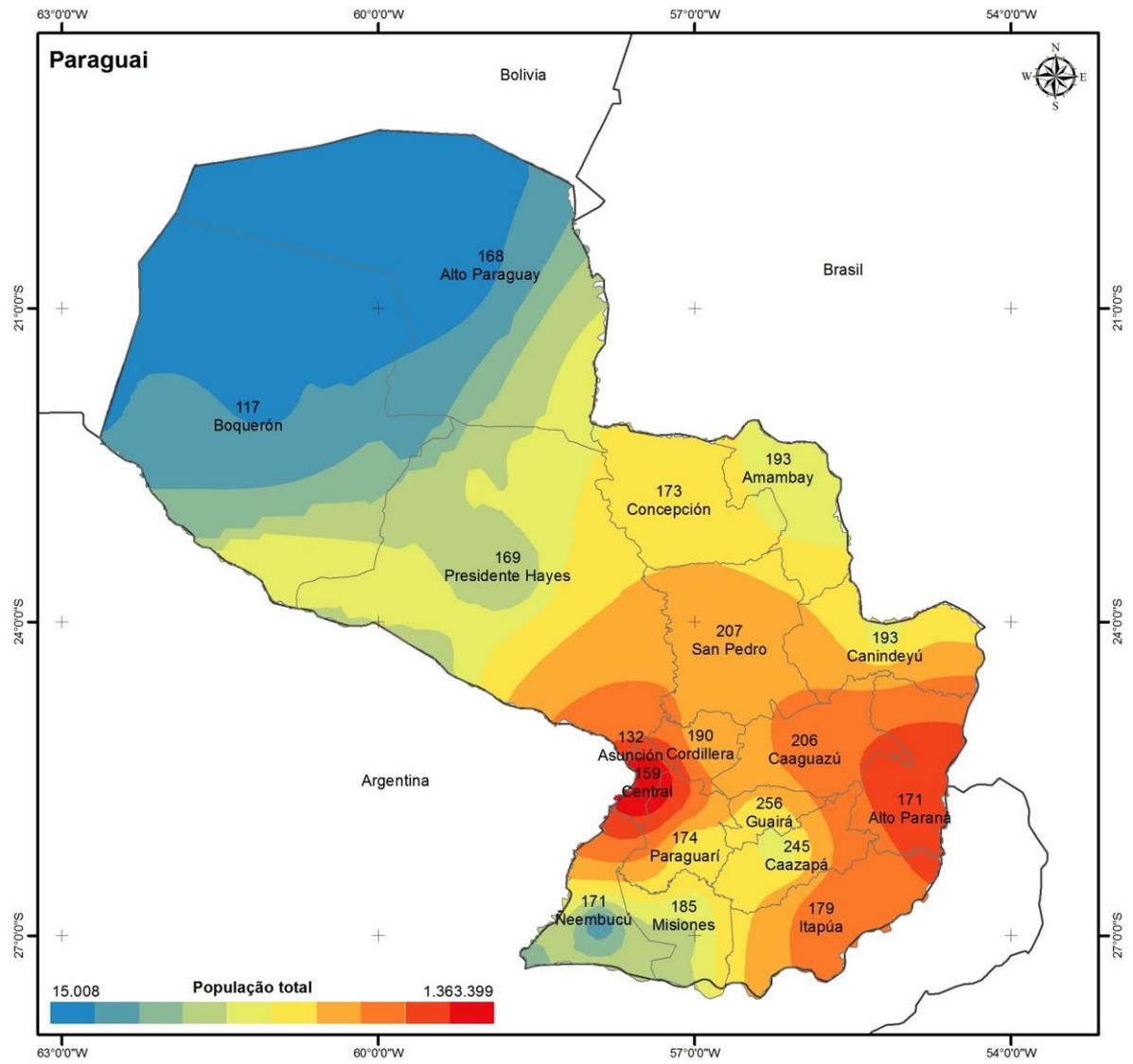


Figura 3 – Mapa de distribuição da população total e índice F, para os departamentos do Paraguai.

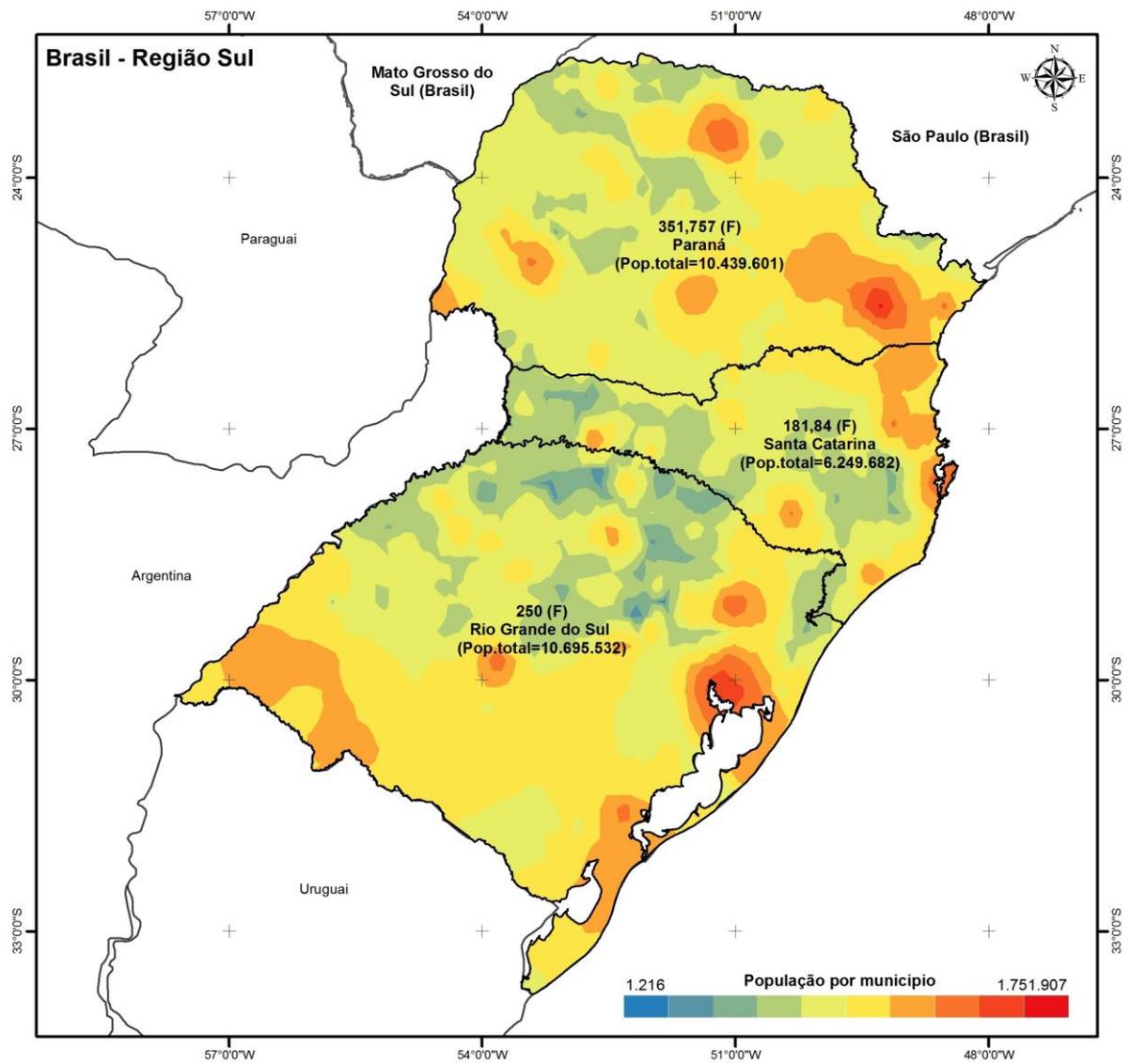


Figura 4 – Mapa de distribuição da população por município da Região Sul do Brasil (Paraná, Santa Catarina e Rio Grande do Sul) e índice F por estado.

CAPITULO 7 - REFERÊNCIAS BIBLIOGRÁFICAS

- Amstad, T. **Cem anos de germanidade no Rio Grande do Sul 1824-1924**. São Leopoldo: UNISINOS, 1999.
- Azevêdo, E.S.; Morton NE, Miki C, Yee S. (1969). **Distance and kinship in northeastern Brazil**. *American Journal of Human Genetics* 21:1-22
- Azevêdo, E.S. (1980). **The anthropological and cultural meaning of family names in Bahia**. *Curr. Anthropol.* 21: 360-363.
- Azevêdo, E.S., Silva, K.M.C., Silva, M.C.B.O., Lima, A.M.V.M.D., Fortuna, C.M.M. and Santos, M.G. (1981). **Genetic and anthropological studies in the Island of Itaparica, Bahia, Brazil**. *Hum. Hered.* 31: 353-357.
- Azevêdo, E.S., Fortuna, C.M.M., Silva, K.M.C., Souza, M.G.F., Machado, M.A., Lima, A.M.V.M.D., Aguiar, M.E., Abé, K., Eulálio, M.C.M.N., Conceição, M.M., Silva, M.C.B.O. and Santos, M.G. (1982). **Spread and diversity of human populations in Bahia, Brazil**. *Human Biol.* 54: 329-341.
- Barbosa, A.A.L., Iglénir J. Cavalli , Kiyoko Abé, Maria G. Santos and Eliane S. Azevêdo.(1997) **Family names and the length of the Y chromosome in Brazilian blacks**. *Braz. J. Genet.*, Ribeirão Preto , v. 20, n. 1, Mar.
- Barbosa, A.A.L.; Sandra Mara Bispo Sosal, Kiyoko Abé-Sandes; Carlos Alberto Alonso; Vicente Schneider; Denise C.C. Costa; Iglénir João Cavalli; Eliane Elisa Souza Azevêdo. (2006) **Microsatellite studies on an isolated population of African descent in the Brazilian state of Bahia**. *Genet. Mol. Biol.* [online]. vol.29, n.1, pp. 23-30. ISSN 1415-4757
- Barbujani, G. & Sokal, R.R. (1990) **Zones of sharp genetic change in Europe are also linguistic boundaries**. *Proc. Natl. Acad. Sci.* 87, 1816-1819.
- Barral, I., Rodríguez-Larralde, A., Mamolini, E., Manni, F. & Scapoli, C. (2000) **Elements of the surname structure of Austria**. *Ann Hum Biol* 27, 607-622.
- Barral I, Rodríguez-Larralde A, Manni F, and Scapoli C (2002) **Isonymy and isolation by distance in the Netherlands**. *Hum Biol* 74:263-281.
- Barral, I., Rodríguez-Larralde, A., Manni, F., Ruggiero, V., Tratari, D., Scapoli, C. (2003) **Isolation by Language and Distance in Belgium**. *Ann Hum Biol* 68, 1-16.
- Bedoya G, et al (2006). **Admixture dynamics in hispanics: a shift in the nuclear genetic ancestry of a south american population isolate**. *Proceedings of the national academy of sciences, estados unidos.* 103:7234–7239.

Bronberg, R., Dipierri, J.E., Alfaro, E.L., Barraí, I., Rodríguez Larralde, A., Castilla, E., Colonna, V., Rodríguez Arroyo, G. & Bailliet, G. (2009) **Isonymy structure of Buenos Aires city.** *Hum Biol* 8, 447-461.

Castilla EE. (2005) **On Being a Medical Geneticist.** *Am J Med Genet* 138A:197-198.

Caravello, G., and M. Tasso. (1999). **An analysis of the spatial distribution of surnames in the Lecco area (Lombardy, Italy).** *Am. J. Hum. Biol.* 11:305–315.

Caravello, G., and M. Tasso. (2002). **Use of surnames for a demo-ecological analysis: A study in southeast of Sardinia.** *Am. J. Hum. Biol.* 14:391–397.

Cheshire, J. A., Longley, P.A. & Singleton, A.D. (2010) **The surname regions of Great Britain.** *J MAPS*, 401-409.

Crow, J.F., and P. Mange. (1965). **Measurement of inbreeding from the frequency of marriages between persons of the same surname.** *Eugen. Quart.*12:199–203.

De Oliveira, Marcelo Zagonel ; Schüler-Faccini, Lavínia ; Demarchi, Dario A. ; Alfaro, Emma L. ; Dipierri, José E. ; Veronez, Mauricio R. ; Colling Cassel, Marlise ; Tagliani-Ribeiro, Alice ; Silveira Matte, Ursula ; Ramallo, Virginia. (2013) **So Close, So Far Away: Analysis of Surnames in a Town of Twins (Cândido Godói, Brazil).** *Annals of Human Genetics*, v. 77, p. 125-136.

Dipierri, J. E., Alfaro, E. L., Scapoli, C, Mamolini, E., Rodriguez-Larralde, A. & Barraí, I. (2005) **Surnames in Argentina: a population study through isonymy.** *Am J Phys Anthropol* 28, 199–299.

Dipierri,J.E., Rodríguez-Larralde, A., Alfaro, E.L. & Barraí, I. (2007) **Isonymy structure of the Argentine northwest.** *Ann Hum Biol* 34, 498–503.

Dipierri,J.E., (2013) **Consanguinidad por isonimia y factores socioeconómicos en Argentina: analisis poblacional y especial.** Tesis doctoral (Inédita). Facultad de Ciencias Médicas. Universidad Nacional de Cuyo. Argentina

Dornelles, C.L., Callegari, S.M., Robinson, W.M., Weiner, T. A., Franco, M.H.L., Geiger, C.J., e Salzano, F.M. (1999) **Genetics, surnames, grandparents nationalities and ethnic admixture in southern Brazil – Do the patterns of variation coincide?** *Genetics and Molecular Biology*, 22, 2, 151-161.

Estrada-Mena B, Estrada FJ, Ulloa-Arvizu R, Guido M, Méndez R, Coral R, Canto T, Granados J, Rubí-Castellanos R, Rangel-Villalobos H, García-Carrancá A. (2010) **Blood group O alleles in Native Americans: implications in the peopling of the Americas.** *Am J Phys Anthropol.* 142:85–94.

Francomano C, Mckusick Va, Biesecker L. (2003). **Medical Genetic Studies In The Amish: Historical Perspective.** Am J Med Genet C Semin Med Genet;121c:1-4

Freire-Maia, N. (1952) **Frequency of consanguineous marriages in Brazillian populations.** American Journal of Human Genetics. Vol 4, Nro 3.

Freire-Maia, N. (1958) **Consanguineous marriages in Brazil.** Eugenics Quaterly, Vol. 5, Nro 2.

Freeman, S; Herron, J. C. (2009). **Análise Evolutiva.** 4a. ed. Porto Alegre: ArtMed Editora. 848 p.

Futuyma, D.J. (1998). **Biologia evolutiva.** 3^a ed. Sociedade Brasileira de Genética, Ribeirão Preto.

Jobling, M.A. (2001) **In the name of the father: surnames and genetics.** *Trends Genet* **17**, 353–357.

Garza-Chapa, R., Rojas-Alvarado M. A. e Cerda-Flores R. M.,(2000) **Prevalence Of Niddm In Mexicans With Paraphyletic And Polyphyletic Surnames.** American Journal Of Human Biology 12:721–728.

GERTZ, René.(1992) **Catolicismo social no Rio Grande do Sul: a União Popular.** Veritas,Porto Alegre, v. 37, nº 148.

Griffiths Anthony J.F., Susan R. Wessler, Richard C. Lewontin, Sean B. Carroll. (2008) **Introduction to Genetic Analysis** ISBN 978-1429233231. Ninth Edition

Graf, O.M., Zlojutro, M., Rubicz, R. & Crawford, M.H. (2010) **Surname distributions and their association with Y-Chromosome markers in the Aleutian Islands.** *Hum Biol* **82**, 745-757.

Guglielmino, C.R., Zei, G. & Cavalli-Sforza, L.L. (1991) **Genetic and cultural transmission in Sicily as revealed by names and surnames.** *Hum Biol* **63**, 607-627.

Guerra D. C. (1987) **Uso de apellidos para determinar estructura de población en pueblos negros de la costa norcentral de Venezuela.** Tesis de Grado Magister Scientiarum en Biología. Centro de Estudios Avanzados. IVIC, Caracas.

Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., Erlich, Y. (2013) **Identifying Personal Genomes by Surname Inference.** Science 339, 321.

IBGE – Instituto Brasileiro de Geografia e Estatística (2010). **Disponível em:** <http://www.ibge.gov.br/cidadesat/xtras/perfil.php?codmun=430430>. Acesso em 10 de Junho de 2013

King TE, Ballereau SJ, Schürer K, Jobling MA. (2006). **Genetic signatures of coancestry within surnames.** *Curr Biol.* 16:384–388.

Lando, Aldair M; Barros, Eliane C.(1980) **Capitalismo e colonização: os alemães no Rio Grande do Sul.** In: DACANAL, José H. (Org.). RS: Imigração e Colonização. Porto Alegre: Mercado Aberto. p. 09-46.

Nascimento, J C. **A cultura ocultada.** Londrina: Ed.UEL.

Manni, F., and I. Barrai. (2001) **Genetic structures and linguistic boundaries in Italy: A microregional approach.** *Hum. Biol.* 73:335–347.

Martins, José de S. (1973). **A imigração e a crise do Brasil agrário.** São Paulo: Biblioteca Pioneira de Ciências Sociais

Mckusick Va, Hostetler Ja, Egeland Ja. (1964) **Genetic Studies Of The Amish, Background And Potentialities.** *Bull Johns Hopkins Hosp. Sep;*115:203-22.

McEvoy B, Bradley DG. (2006). **Y-chromosomes and the extent of patrilineal ancestry in Irish surnames.** *Hum Genet.* 119:212–219.

McQuillan R, Leutenegger AL, Abdel-Rahman R, et al. (21 coauthors). (2008). **Runs of homozygosity in European populations.** *Am J Hum Genet.* 83:359–372.

McKusick VA. (1973) **Genetic studies in American inbred populations with particular reference to the Old Order Amish.** *Isr J Med Sci. Sep-Oct;*9(9):1276-84.

Mesa, M. S., A. M. López-Parra, D. R., C. Baeza, E. L. Alfaro, F. Castilla, M. D. Marrodán, M. Tirado, P. Iturbe, C. Gamba, E. Fernández, J. E. Dipierrri y Arroyo-Pardo E. (2010) **La aplicación de los Y-STRs a la investigación de los linajes familiares. em: Diversidad Humana y Antropología Aplicada** (A. Gutiérrez-Redomero, A. Sánchez-Andrés y V. Galera Olmo, eds.), pp. 129-136. Universidad de Alcalá; Alcalá de Henares. ISBN 978-84-614-2257-9.

Michael Payne, C Anthony Rupar, Geoffrey M Siu, BAsC, and Victoria Mok Siu, MD (2011) **Amish, Mennonite, and Hutterite Genetic Disorder Database.** *Paediatr Child Health.* March; 16(3): e23–e24.

Nascimento, J C. (2007) **A cultura ocultada.** Londrina: Ed.UEL.

Neumann, R.M. (2003) **Quem nasce no Brasil, é brasileiro ou traidor”. A Campanha de Nacionalização nas Colônias Alemãs.** São Leopoldo: UNISINOS, 2003. Dissertação (Mestrado em História), PPGH, UNISINOS.

Ostler H, Skorecki K. (2013) **The population genetics of the Jewish people.** Hum Genet. Feb;132(2):119-27.

Pianovski, M. A., Maluf, E. M., de Carvalho, D. S., Ribeiro, R. C., Rodriguez-Galindo, C., Boffetta, P., Zancanella, P. e Figueiredo, **B. C.** (2005) **Mortality rate of adrenocortical tumors in children under 15 years of age in Curitiba, Brazil.** Pediatric Blood & Cancer 47, 56–60.

Pinto Cisternas J, Castro de Guerra D (1988) **Utilidad de los apellidos en estudios de biología humana.** Rev Med Chile 116:1191-1197.

PNUD - Programa das Nações Unidas para o Desenvolvimento 2010

Disponível em

http://www.undp.org/content/undp/es/home/librarypage/hdr/human_developmentreport2011.html. Acessado em Setembro de 2013.

Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ. (2002). **Ychromosomal DNA variation in Pakistan.** Am J Hum Genet. 70:1107–1124.

Rambo, A.B. (1988) **Associativismo teuto-brasileiro e os primórdios do cooperativismo no Brasil.** *Persp Econ* 23, 263-276.

Ramos, A. D. (2007) **A Formação Histórica Dos Municípios Da Região Das Missões Do Brasil.** Santo Ângelo: Publicação On-Line, V. 1. 100 p.

Roche, Jean. (1969) **A colonização alemã e o Rio Grande e o Sul.** Porto Alegre: Editora Globo.

Sayad, A. (1988) **A Migração ou os paradoxos da alteridade.** São Paulo: Editora da Universidade de São Paulo.

Scapoli, C., Mammolini, E., Carrieri, A., Rodriguez Larralde, A. & Barraí, I. (2006) **Surnames in Western Europe: a comparison of the subcontinental populations through isonymy.** *Theor Popul Biol* 71, 37-48.

Schallenberger, E. (2001). **O Associativismo cristão no Sul do Brasil: a contribuição da Sociedade União Popular e da Liga das Uniões Coloniais no desenvolvimento e na organização social sul-brasileira.** Porto Alegre: PUCRS, v. 15. 593p

Tarskaia, L., El'chinova G. I., Scapoli C., Mamolini E., Carrieri A., Rodriguez-Larralde A. & Barraí I. (2008) **Surnames in Siberia: a study of the population of Yakutia through isonymy.** *Am J Phys Anthropol* 138, 190–198.

Tagliani-Ribeiro, A., Oliveira, M., Sassi, A.K., Rodrigues, M.R., Zagonel-Oliveira, M., Steinman, G., Matte, U., Fagundes, N.J.R. & Schuler-Faccini, L.

(2011) **Twin town in south Brazil: a nazi's experiment or a genetic founder effect?** *PLoS ONE* 6, e20328.

Turi E. King and Mark A. Jobling. (2009) **Founders, Drift, and Infidelity: The Relationship between Y Chromosome Diversity and Patrilineal Surnames.** *Mol. Biol. Evol.* 26(5):1093–1102.

Vienna, A., Peña Garcia, J. A. , Mascie-Taylor, C. G. N. , Biondi, G. (2001) **The ethnic minorities of Southern Italy and Sicily: relationships through surnames** *J. biosoc. Sci.* (33): 25–31

Wright, S. (1930). **The genetical theory of natural selection. A review.** *J. Hered.* 21, 349-56.

Zlotogora J, Van Baal S, Patrinos Gp. (2009) **The Israeli National Genetic Database.** *Isr Med Assoc J.* Jun;11(6):373-5

Liu, Y. Chen I., Yan Y. and Chen J.(2012) **A study of surnames in china through isonymy.** *American Journal of Physical Anthropology* 148:341–350

Anexo 1 - HISTORIA DA COLONIZAÇÃO ALEMÃ NO RIO GRANDE DO SUL

1. IMIGRAÇÃO E COLONIZAÇÃO NO RIO GRANDE DO SUL

A primeira fase da imigração alemã no Rio Grande do Sul começa ano de 1824, período em que o governo imperial cria a colônia alemã de São Leopoldo, berço dessa imigração no Rio Grande do Sul. No entanto, em 1930 a oposição parlamentar, suspende essa colonização e a guerra civil paralisa a província até 1845.

De 1848 a 1874 o governo imperial institui diretrizes para o processo de colonização, e estabelece o primeiro estatuto que orientava a imigração, sendo esta constituída de indivíduos de origem germânica e visava à ocupação de territórios ainda não ocupados. Neste momento os imigrantes sendo que a mesma c puramente germânica, balizando a encosta do planalto. No período de 1874 a 1889 o governo provincial abandona a colonização porque a opinião publica lhe é desfavorável, mesmo assim o governo imperial funda apenas algumas colônias com imigrantes italianos.

A partir de 1889, estendendo-se até 1914, o governo do estado, agora incumbido da colonização, dá-lhe novo estatuto e impulso e novas colônias são instaladas no planalto onde prosperaram rapidamente. Após 1914 o estado limita a imigração e reorganiza os serviços de colonização, liquidando as questões pendentes e prepara a integração das colônias na comunidade nacional, visto que se acha quase concluída a distribuição das terras devolutas.

Com intensão de redirecionar o fluxo emigratório que acontecia na Europa no século XIX, a colonização passa a ser tratada como questão oficial do império brasileiro, com investimentos em propaganda do país no exterior, redirecionamento de imigrantes à medida que chegavam nos portos, administração, organização das colônias assim como subsídios nos primeiros meses, fazendo assim o povoamento de locais estratégicos do país. Os processos migratórios no país ocorreram de modo heterogêneo, cabem aqui os exemplos dos casos paulista e do sul do Brasil. No primeiro, os imigrantes substituíram gradativamente o braço escravo nos cafezais, inicialmente pelo

sistema de parceria, e posteriormente como trabalhadores assalariados. Já no sul do país, tornaram-se proprietários de terras logo após a chegada.

Outro fato histórico relevante para os processos migratórios ocorreu com o fim do Império, as terras devolutas passam para os estados assim como a política de imigração e colonização. No Rio Grande do Sul, o governo passa a apoiar a imigração espontânea, retirando gradualmente qualquer tipo de subsídio anteriormente fornecido pelo império, permitindo a atuação da iniciativa privada nos processos de imigração. Cabe aqui salientar a relevância deste fato, já que conformará as práticas administrativas do processo de imigração deste período em diante, no sentido da configuração dos novos núcleos coloniais que passam a ser, em grande parte, confeccionais e com tendências étnicas. Segundo Neumann (2003), essa política não afastava os imigrantes do estado, uma vez que eram contínuos os pedidos de colonos já estabelecidos para que fosse facilitada a vinda de seus parentes e patrícios.

Por essa razão, no Rio Grande do Sul, a colonização constitui em uma exceção se comparada ao restante do país, pois não se caracterizou apenas pela importação de simples trabalhadores ou assalariados. Em outros estados, esse regime levou a formação de uma população flutuante, com número elevado de migrações e mobilidade.

Conforme o posicionamento do governo, a imigração espontânea, embora numericamente reduzida, voltada para a agricultura, transformava o imigrante de imediato em pequeno proprietário, evitando assim uma massa de população flutuante, migrando de um lugar para o outro, como ocorria, por exemplo, em São Paulo.

O desenvolvimento da imigração obrigaria à fundação de novos núcleos coloniais, mais afastados dos centros consumidores e de exportação, onde a falta de meios de comunicação “obrigaria o colono a arrastar uma vida de quase miséria, por não achar comprador para os seus produtos e, quando achasse, seria por valor tal que talvez não desse para a parca manutenção da família” (Nascimento, 2007).

Em maio de 1914, o governo estadual rescindiu o acordo celebrado com a União em agosto de 1908, referente à instalação de imigrantes introduzidos por ela no país pelo Serviço de Povoamento do Solo, cessando, a partir de então, as entradas de novas levas. Explicando o fato, a Diretoria de Terras e

Colonização justificava que o governo tinha por objetivo dar maior atenção aos imigrantes já instalados, apontando que as medidas gerais a serem tomadas em relação ao serviço de colonização eram de duas naturezas: “medidas relativas à instalação dos agricultores e medidas relativas à nacionalização gradual dos de origem estrangeira”. Salientava que, dada a elevada população colonial no estado, a medida foi bem recebida por parte dos colonos até os próprios empresários particulares de colonização. (Nascimento, 2007).

Havia ainda as razões de ordem política e moral, igualmente contrárias à rápida elevação artificial da população do Estado. Logo, a realocação do elemento colonial não implicava em ônus para o estado, cujos recursos poderiam ser aplicados no melhoramento das colônias. No mesmo relatório de 1915, informava ainda que a população de estrangeiros e descendentes formava mais de 1/3 da população total do estado. Desses, 60% de origem germânica e eslava, e 40% latina. A assimilação desse contingente à população nacional era urgente e necessária. Para viabilizar esse processo, apontava como medidas: cessar a instalação de mais imigrantes; instalação e proteção do elemento genuinamente nacional, bem como a organização de colônias mistas; desenvolvimento da viação nas colônias para permitir maior contato entre a população; e difusão da instrução primária.

1.2. Avanço da imigração para a região norte do estado

Na prática, essas medidas políticas referentes à imigração e realocação do excedente populacional da antiga zona colonial, implicaram no avanço da fronteira agrícola para a região Norte do estado, onde se formou a nova zona colonial ou as chamadas colônias novas. Caracterizou-se pela atuação da iniciativa privada, seja individual ou de empresas; e iniciativa pública, no caso estadual ou municipal.

Atendendo aos propósitos oficiais, ou seja, à rápida nacionalização dos elementos estrangeiros, as colônias deveriam ser etnicamente mistas – o caso exemplar de Ijuí (1890). Contudo, no que se refere à iniciativa privada, esse quesito nem sempre foi regra, e o governo não dispunha de mecanismos eficientes para interferir, permitindo assim a formação de colônias étnicas, como foi o caso das colônias fundadas pela Empresa de Colonização Dr. Herrmann Meyer, com a alemães luteranos.

1.2.1 Ação da Colonização particular

As colônias particulares de povoamento essencialmente germânico adentram no planalto central subindo o Alto Jacuí ao longo da via férrea. De acordo com Roche, 1969 são assim fundados os núcleos de Santa Clara (1896), Alto Jacuí (1897), Não-me-toque (1897), General Osório (1898), Dona Ernestina (1900), Selbach (1906), no município de Cruz Alta, Barra do Colorado (1897), Boi Prêto (1897), Neu Wurttemberg (1899), o de Santo Ângelo, Ijuí Grande (1892), Vitória (1900), Buriti (1908), Timbaúva e Boa Vista (1912), Steiglich (1914), o de São Luiz Gonzaga, Cêrro Azul (1902), Bela Vista, Bela Vista (1903) e Dona Julia (1912); no de Erechim, Rio do Peixe (1911).

Essas colônias fundadas por empresas particulares ou por sociedades apresentavam características diferenciadas das fundadas pelo governo do estado que tinha um maior cuidado em desenvolver os meios de comunicação entre as colônias por ele fundada, tornando os núcleos particulares e de sociedades, mais isolados do que os fundados pelo estado, reunindo indivíduos da mesma origem étnica e muitas vezes da mesma confissão. Os alemães católicos agruparam-se em colônias como Boa Vista (hoje Santo Cristo) e Cerro Azul (hoje Cerro Largo), criadas por organismos profissionais e confessionais (Deutsche Bauer Verein), conservando o zelo tanto pela religião como pelas tradições germânicas, enquanto que os protestantes se dirigiram em grande maioria para Neu-Wurttemberger e colônias vizinhas, todas colônias-modelos fundada por H.Meyer. A força que esses blocos homogêneos adquiriram foi motivo de preocupações para o estado.

A partir de 1914, a imigração passa por um período que ocorre entre duas guerra mundiais, onde Brasil e Alemanha, encontram-se em lados opostos, surgindo desta forma novos problemas culturais e cívicos aos colonos, onde tinham uma dependência ao estado e uma fidelidade espiritual a uma nação.

1.3 Instituições relevantes no processo de colonização

1.2.1 A união popular para a Alemanha católica – *Volksverein für das katholische Deutschland*

Num contexto de efervescência da ação católica³, a sob a liderança de intelectuais, de políticos, de empresários e do alto clero, foi fundada a União Popular para a Alemanha Católica - *Volksverein* -, em agosto de 1890 que mais tarde viria a ser a mais representativa organização associativa católica da Alemanha. (Schallenberger, 2001). Esta instituição tinha como discurso a defesa do catolicismo, a sua difusão e promoção como objetivos básicos. A *Volksverein* deveria atingir a todos os católicos, portanto, na concepção dos líderes não poderia ser uma associação com altas contribuições, visando manter o caráter e a simpatia popular. Conforme Gertz (1992)., esta deveria ser, sobretudo, “a associação do povo católico para promover e divulgar as ideias do catolicismo social mediante uma obra gigantesca de educação popular para lutar contra o socialismo e pela liberdade religiosa”.

A grande mobilização e a vitalidade do social-catolicismo alemão fizeram com que a sua ação se voltasse para os alemães católicos que viviam no exterior. Surgiu assim, em 1911, em Dresden, a “União para os Alemães Católicos do Exterior” - *Katholisches Deustchtum im Auslande*. Esta organização se assentou nos objetivos de:

- Dar assistência viva na educação entre a pátria (Alemanha) e os colonos católicos que vivem no exterior;
- Zelar pela manutenção e pelo fomento da língua, dos costumes, da cultura e da religião;
- Fornecer subsídios para o desenvolvimento espiritual e material dos alemães que estiverem distantes de sua pátria.

Esses objetivos figuravam a busca de um elo de coesão social onde o elemento identificador da unidade nacional esteve envolto em torno do conceito de cultura, neste caso consignado em base aos elementos étnico-lingüísticos e religiosos. (Nascimento, 2007).

1.3.1 O *Bauernverein* (associação dos agricultores)

O *Bauernverein* – Associação dos Agricultores – foi fundado no II *Katholikentag* (2º Congresso Católico) realizado em Feliz (RS), em 1900, e seus objetivos direcionavam-se principalmente aos problemas econômicos e técnicos dos agricultores. Uma das principais ideias defendidas pela

associação era o cooperativismo, cujo idealizador foi o padre jesuíta Theodor Amstad. De acordo com seus estatutos, essa associação foi pensada como uma organização interétnica e interconfessional. As pretensões, portanto, não se limitavam à solução dos problemas dos teuto-brasileiros católicos apenas. Tratava-se de um projeto de promoção humana que não deveria excluir nenhuma das vertentes étnicas ou religiosas presentes no estado.

O *Bauernverein* fundou a colônia Serro Azul (Cerro Largo), em 1902. Já em 1912, na assembleia dos católicos em Venâncio Aires, houve uma divisão, e foi fundada a *Volkverein für die Deutschen Katholiken in Rio Grande do Sul* (Sociedade União Popular para os Católicos Alemães do Rio Grande do Sul), também idealizada pelo padre Theodor Amstad SJ., logo, de caráter confessional, enquanto que os de confissão protestante continuaram com o *Bauernverein*, sob a designação de Liga Colonial (Rambo, 1992; Schallenberger, 2001).

As Uniões Coloniais remanescentes do *Bauernverein* constituíram, em 1929, a Liga das Uniões Coloniais, que, sob marcante liderança evangélico-luterana, manteve-se próxima do Estado e exerceu papel vital na organização sindical e no desenvolvimento do cooperativismo de produção. Até a edição do Decreto Federal Nº 581/1938, que definiu nova estrutura para o cooperativismo, vinculando-o ao controle dos órgãos públicos federais, haviam sido constituídas em torno de 50 cooperativas de crédito, organizadas pelo *Volkverein*, e 289 de produção, fomentadas pela Liga. O associativismo promovido pelos teuto-brasileiros marcou o cenário sociocultural e econômico do sul do Brasil.

1.4. Colonização da região noroeste

A região noroeste do estado, conhecida atualmente por região missioneira ou das missões, destaca-se pela presença de inúmeros descendentes de diferentes etnias, porém entre essas a presença de descendentes de alemães é maior devido a migração mais recente. Após a primeira tentativa de instalação de colonos alemães na região na década de 1820, apenas no início do século XX é que os colonos germânicos retornam.

Devido principalmente ao superpovoamento e esgotamento das terras nas primeiras colônias (colônias velhas) do estado, houve a migração para

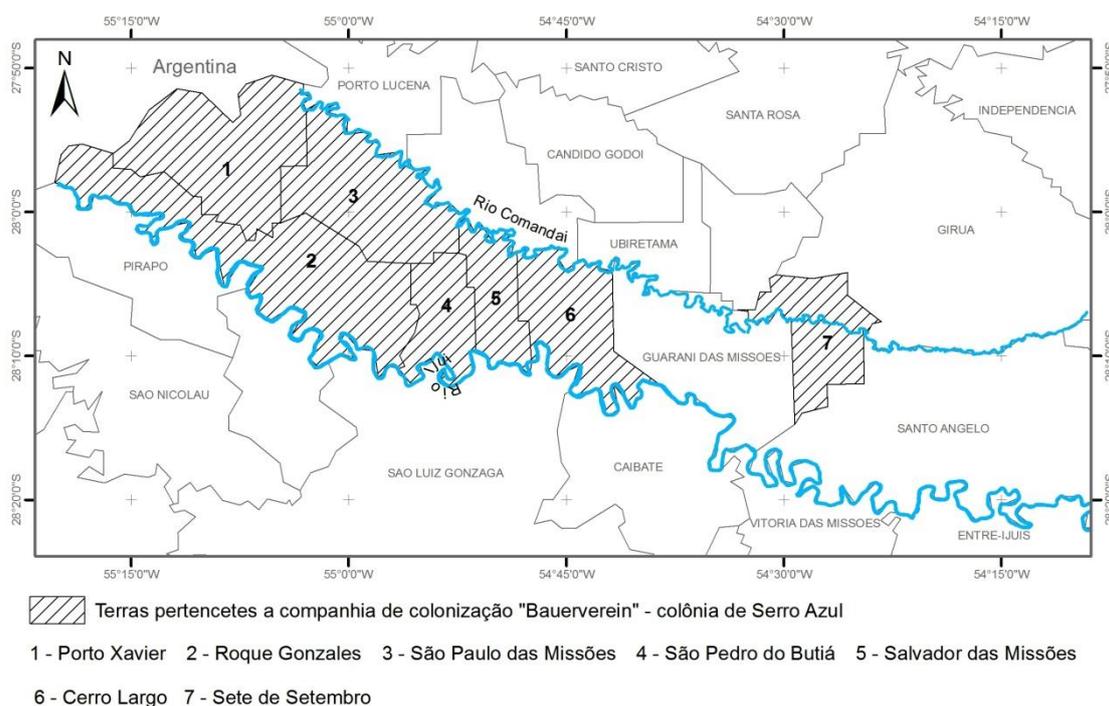
outras regiões e o fluxo migratório direciona-se a colonização de regiões distantes e em terras com povoamento reduzido, a noroeste do estado.

Dessa forma, em 1900, a recém-fundada Associação de Colonos intitulada *Bauerverein*, com sede no município de São José do Hortêncio, interessou-se em colonizar uma área de terras na região noroeste, entre os rios Ijuí e Comandáí, que abrangia terras de propriedade da *Nordwestbahn*, uma companhia férrea alemã que havia adquirido as terras do governo do Estado.

Segundo Nascimento (2007), em 1902 o "*Bauerverein*" adquiriu duas áreas de terra da *Nordwestbahn* e comprometeu-se a colonizar a colônia Serro Azul, atual município de Cerro Largo e a colônia Boa Vista, atual município de Santo Cristo.

Através da Companhia de Colonização "Bauerverein", que promoveu a venda de lotes de terras na região das missões, inúmeras famílias de colonos descendentes de imigrantes alemães foram assentadas na região. Esse assentamento que primeiro chamou-se Serro Azul abrangia os atuais municípios de Cerro Largo, Roque Gonzales, São Paulo das Missões, Salvador das Missões, Sete de Setembro, Porto Xavier e São Pedro do Butiá (Ramos, 2006).

...no Rio Grande do sul fora fundada a Companhia de Estrada de Ferro Alemã (Nortwestbahn) para construir uma ferrovia. (...) Em troca da construção, a Companhia recebia a concessão para colonizar uma faixa de terra de 5 Km ao longo da estrada podendo repassar ao preço de 2000 réis o hectare. Mas não obteve sucesso. Até 1900, essa não chegara a assentar uma dúzia de colonos e o projeto foi repassado à Bauernverein, que através de um acordo com a Nortwestbahn e o Governo Estadual, comprometia-se a colonizar a gleba Serro Azul, no prazo de 10 anos, para assentar famílias de origem alemã. (...) Esses pioneiros de serro Azul adquiriram lotes de terra na área compreendida entre os rios Ijuí e Comandáí, que foi repartida em lotes segundo critérios diversos. Dividia-se em faixas de terra de um a dois quilômetros de largura, chamados de "linha". (...) Cada localidade, desde o início, procurou obter autosuficiência. Por isso era necessário capela, escola e clube da comunidade, além de uma casa comercial. (2000: 14 – 17)



1.5 A colonização étnica e confessional

Projetos de colonização étnica foram bem recebidos pela ala germanista no estado, enquanto criticava a formação de colônias mistas, pela sua despreocupação com questões étnicas e confessionais, o que comprometia o seu desenvolvimento como um todo. Dentre os maiores críticos, estavam os jesuítas católicos, envolvidos diretamente na assistência religiosa da zona colonial alemã e italiana e ligados ao Bauernverein, defenderam abertamente a colonização confessional e étnica, tendo como seus expoentes os padres jesuítas Theodor Amstad, João Evangelista Rick, Max von Lassberg. Nessa perspectiva, foram fundadas as colônias Serro Azul (Cerro Largo-RS) e Porto Novo (Itapiranga-SC).

Para Amstad (1999), a vantagem da colonização particular estava no fato de “ser executada com relativa rapidez, com custos adicionais insignificantes e, na maioria dos casos, adotando a separação étnica e confessional”. Distanciava-se, nesse ponto, do modelo de colonização mista adotada pelo governo republicano, que tinha por propósito estabelecer diferentes grupos étnicos em uma mesma colônia. Na contramão, vários empreendimentos de colonização particular, de capital nacional ou estrangeiro,

fundaram colônias étnicas ou confessionais, e em raros casos, de caráter político.

Dentre os projetos de colonização confessionais, destaca-se a fundação da colônia de Serro Azul, hoje Cerro Largo, na região das Missões, em 4/10/1902, pela Associação de Agricultores, a *Bauernverein*, na pessoa de Pe. Max von Lassberg, SJ, juntamente com Karl Culmey. A partir dessa experiência, o mesmo religioso fundou ainda ao lado Santo Cristo, e na província de Misiones, Argentina, as colônias Puerto Rico e San Alberto. Na continuidade, e a partir da experiência de Serro Azul, a Sociedade União Popular Católica ou *Volksverein* fundou, na pessoa de Pe. João Evangelista Rick, SJ, a colônia Porto Novo, hoje Itapiranga, em Santa Catarina, na margem norte do rio Uruguai, em 31/7/1926, terras estas adquiridas da Empresa Chapecó-Pepery. Como já abordado anteriormente, a *Volksverein* tinha como finalidade promover os interesses materiais e ideais do povo católico de fala alemã. Seu programa, sobretudo na sua concepção ideal, avançou muito mais do que a antiga Associação de Agricultores.

A *Bauernverein* era interconfessional e tinha a melhoria material como objetivo principal. No breve tempo de sua existência realizou muito com a fundação de caixas de poupança, melhoria dos métodos de trabalho, introdução de novas raças de gado e de modo especial pela fundação da esplêndida colônia de Cerro Largo. A mistura confessional, porém, continha o gérmen de sua dissolução. (Rambo, 2004; Schallenberger, 2001).

Na avaliação do padre jesuíta Balduino Rambo, era preciso separar não só por confissões como também por nacionalidades. As colônias mistas de alemães, italianos e poloneses não logravam organizar uma vida comunitária como era de desejar. Em muitos casos o segmento mais fraco migrava para outro lugar, fato que criava também sérios obstáculos para o desenvolvimento econômico (Rambo, 2004).

Conforme Amstad 1999, Ijuí e São Leopoldo apresentaram traços e papéis similares. Como São Leopoldo fora o centro de irradiação das chamadas colônias antigas, ou colônias velhas. Assim Ijuí seria o centro de irradiação das colônias novas, no norte e noroeste do Estado.

A partir daí, a região se transformou num grande laboratório de experiências de colonização. Entre os anos de 1890 e 1930, foram

implantados, no mínimo, quatro modelos de colonização. Ijuí foi uma iniciativa do governo federal, Santa Rosa, do governo estadual, e Santo Ângelo, um projeto municipal. Todas as demais colonizações foram empreendimentos particulares, ou de empresas, ou de associações ou de pessoas físicas.

Ao lado das colônias oficiais, propagavam-se as particulares, com características próprias. Roche (1969) afirma que elas não desenvolveram meios de comunicação na mesma proporção que as do governo, resultando num isolamento mais acentuado, primando pela homogeneidade étnica e confessional.

Anexo 2 – Artigo produzido em colaboração - Genetic, Ancestry, Phenotypic diversity and Self-Perception of Ancestry in 6,938 Latin Americans (Aceito em Plos genetics)

GENETIC ANCESTRY, PHENOTYPIC DIVERSITY AND SELF-PERCEPTION OF ANCESTRY IN 6,938 LATIN AMERICANS

Andrés Ruiz-Linares¹, Kaustubh Adhikari¹, Victor Acuña-Alonzo^{1,2}, Mirsha Quinto³, Claudia Jaramillo⁴, William Arias⁴, Macarena Fuentes⁵, María Pizarro⁵, Paola Everardo^{2,6}, Francisco de Avila^{2,6}, Jorge Gómez-Valdés⁶, Paola León-Mimila⁶, Tábita Hunemeier⁷, Virginia Ramallo⁷, Caio C. Silva de Cerqueira⁷, Mari-Wyn Burley¹, Marcelo Zagonel de Oliveira⁷, Mauricio Roberto Veronez⁸, Marta Rubio-Codina⁹, Orazio Attanasio^{9,10}, Sahra Gibbon¹¹, Carla Gallo¹², Giovanni Poletti¹², Javier Rosique¹³, Lavinia Schuler-Faccini⁷, Francisco M. Salzano⁷, Maria-Cátira Bortolini⁷, Samuel Canizales-Quinteros^{6,14}, Francisco Rothhammer⁵, Gabriel Bedoya⁴, David Balding¹ and Rolando Gonzalez-José³

¹ Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London, UK

² National Institute of Anthropology and History, Mexico City, México

³ Centro Nacional Patagónico, CONICET, Puerto Madryn, Argentina

⁴ Universidad de Antioquia, Medellín, Colombia

⁵ Instituto de Alta Investigación Universidad de Tarapacá, Programa de Genética Humana ICBM Facultad de Medicina Universidad de Chile and Centro de Investigaciones del Hombre en el Desierto, Arica, Chile

⁶ Facultad de Medicina and Facultad de Química, UNAM, Mexico City, México

⁷ Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

⁸ Remote Sensing and Digital Imaging Laboratory, Graduate Program on Geology, Vale do Rio dos Sinos University, São Leopoldo, Brazil

⁹ The Institute for Fiscal Studies, London, UK.

¹⁰ Department of Economics, University College London, UK.

¹¹ Department of Anthropology, University College London, London, UK.

¹² Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Perú

¹³ Departamento de Antropología. Facultad de Ciencias Sociales y Humanas. Universidad de Antioquia, Medellín, Colombia

¹⁴ Instituto Nacional de Ciencias Médicas y Nutrición “Salvador Zubirán”, Mexico City, México

ABSTRACT

We report a characterization of the genetic ancestry (African, European and Native American), phenotypic diversity and self-perception of ancestry in 6,938 Latin American individuals recruited in Brazil, Chile, Colombia, México and Perú. Phenotypes examined include pigmentation (of hair, eyes and skin), hair type (curliness) and greyness, male pattern baldness, and a range of anthropometric and facial features. We observe extensive variation in ancestry between individuals and geographic regions. We find significant effects of ancestry ($P < 10^{-3}$) on most phenotypes examined, consistent with the geographic structure of these traits in Africa, America and Europe. Significant association was detected for two of the markers typed: rs260690 (in the Ectodysplasin A Receptor, *EDAR* gene) is associated with hair type (P-value 2×10^{-16}) and rs717225 (in the Rho guanine nucleotide exchange factor 1, *ARHGEF1* gene) with skin pigmentation (P-value 1.58×10^{-12}). We observe a significant correlation between self-perception and genetic estimates of ancestry and find evidence for significant effects of pigmentation and social factors in biasing self-perceived relative to genetically estimated ancestry.

The history of the countries in the region collectively known as “Latin America” has involved the coming together of individuals with diverse geographic origins of which the demographically most prominent encounter has been that occurring between the American natives and people arriving mostly from the Iberian peninsula and West (Sub-Saharan) Africa starting in 1492(1-3). In the past several years there has been an increasingly detailed genetic characterization of the resulting history of genetic admixture in Latin America(4-9). Amongst the most prominent conclusions from these studies is the observation of an extensive variation in genetic ancestry between Latin American individuals and regions, and that admixture involved predominantly immigrant men and native women. These findings generally agree with information on pre-Columbian native population density across the Americas and the historical patterns of overseas immigration to specific geographic regions of the continent(2). The sex-bias in admixture is also consistent with documentary evidence showing that it was mostly men that migrated to the New World during the colonial period, particularly in the first century of settlement, when admixture was likely to have been most extensive(1, 10).

Studies relating the genetic diversity of Latin Americans to phenotypic variation have so far been scant(11-13). Such genotype-phenotype correlation analyses are of considerable biological importance as they would contribute to elucidating the genetic basis of variation in human physical appearance, particularly of traits differing in frequency between the major populations involved in Latin American admixture. The identification of genes influencing these traits is of wide-ranging relevance. Firstly, human physical attributes are used extensively in evolutionary analyses, ranging from human origins (i.e. “humans” are defined as “anatomically modern”) to studies on potential environmental adaptations, with the implicit assumption that such attributes reflect a specific genetic make-up. However, the poorly defined genetic basis of these traits limits many of the inferences made in such

analyses. Secondly, identifying genes for variation in human physical appearance is of biomedical relevance as such genes would constitute candidates potentially involved in pathological variation in those traits. It has also been shown that certain physical features are associated with a variable disease risk, but a biological explanation for these associations is often unclear. For instance, pigmentation has been associated with risk of Parkinson's disease(14) and skin cancer(15), male pattern baldness with heart disease(16) and prostate cancer(17), and facial features with general health history(18, 19). Thirdly, identification of genetic variants influencing human physical features promises to be useful in the prediction of an individual's appearance from genetic data, an application of considerable relevance in forensics(20), although one which is not without challenges and problematic social implications. Similarly, characterizing the genetic basis of variation in human physical appearance, particularly that occurring between individuals with different continental origins, has cultural ramifications outside of forensics and biomedicine. Variation in physical appearance has long been used in attempts to define "human races" and historically many of these traits have been highly charged socially(21). Evaluating factors influencing individual perception of variation in these traits, their association with popular notions of ancestry and identity and how these are informed or transformed by the genetic ancestry information, is thus of considerable relevance to the social sciences and for society at large(22-24).

Here we examine the correlation between genetically estimated African/European/Native American ancestry, variation in individual physical appearance and self-perception of ancestry in 6,938 individuals ascertained in Brazil, Chile, Colombia, Mexico and Perú. We also evaluated the possibility of a direct effect of the genetic markers used for ancestry estimation in variation in physical appearance and found novel associations of skin pigmentation with *ARHGEF1* and of macroscopic hair type with *EDAR*.

MATERIALS AND METHODS

Study subjects

Recruitment took place mainly in five cities: Mexico City (México), Medellín (Colombia), Lima (Perú), Arica (Chile) and Porto Alegre (State of Rio Grande do Sul, Brazil). With the exception of Chile, most subjects recruited in these cities were students and staff from the Universities participating in this research. In Chile about 2/3 of the subjects recruited were professional soldiers from a battalion in Arica. In Brazil some samples were also collected in smaller towns of Rio Grande do Sul, as well as in towns of the states of Bahia and Rondonia . Adult subjects of both sexes were invited to participate mainly through public lectures and media presentations. We obtained ethics approval from: Escuela Nacional de Antropología e Historia (México), Universidad de Antioquia (Colombia), Universidad Peruana Cayetano Heredia (Perú), Universidad de Tarapacá (Chile), Universidad Federal do Rio Grande do Sul (Brazil) and University College London (UK). Each participant provided written informed consent. Blood samples were collected by a certified phlebotomist and DNA extracted following standard laboratory procedures.

Phenotypic data

A physical examination of each volunteer was carried out by the local research team using the same protocol and instruments at all recruitment sites. We recorded: height, weight, head, hip and waist circumference, cheilion-cheilion width and sellion-gnation height. All measurements were obtained in duplicate and the mean of the two measurements retained for further analyses. Using visual inspection we recorded eye colour into five categories (1-blue/grey, 2-honey, 3-green, 4-light brown, 5-dark brown/black), and natural hair colour into four categories (1-red/reddish, 2-blond, 3-dark blond/light brown or 4-brown/black)(25). Balding in males was recorded using a modified Hamilton scale(26-29) as: 0) no hair loss, 1) frontal baldness only, 2) frontal hair loss with mild vertex baldness, 3) frontal hair loss with

moderate vertex baldness, and 4) frontal hair loss with severe vertex baldness. Similarly, greying was recorded along a five point scale: 0) for no greying, 1) for predominant non-greying, 2) for ~50% greying, 3) for predominant greying and 5) for totally white hair(29, 30). Macroscopic hair type was categorized by visual inspection as straight, wavy, curly or frizzy (31). A quantitative measure of constitutive skin pigmentation (the Melanin Index, MI) was obtained using the DermaSpectrometer DSMII reflectometer (Cortex Technology, Hadsund, Denmark)(32, 33). Measurements were obtained on both inner arms and the mean of the two readings used in the analyses.

Five digital photographs of the face: left side (-90°), left angle (-45°), frontal (0°), right angle (45°), right side (90°) were taken from ~1.5 meters at eye level using a Nikon D90 camera fitted with a Nikkor 50mm fixed focal length lens. The photographs were annotated manually with 36 anatomical landmarks and 3D landmark coordinates extracted using the software Photomodeler (<http://www.photomodeler.com/> Eos Systems Inc, Vancouver, Canada) (Supplementary Figure S1). Landmark configurations were superimposed by Generalized Procrustes Analysis(34) and the nuisance effects of BMI, sex, and age removed by multiple regression of the shape coordinates. Principal Components (PCs) of the residual landmark coordinates were obtained using the software MorphoJ(34). To ease visualization of the 3D shape changes associated with principal components, we generated a file with spatial information for the i -th PC score produced by MorphoJ and obtained deformation surfaces via thin a plate spline algorithm using the Landmark software (35).

Socioeconomic position and self-perception of ancestry

A structured questionnaire was applied to each volunteer so as to obtain information on socioeconomic position (SEP) and on self-perception of ancestry. We obtained two indicators of SEP. The first is maximum educational attainment (EA), categorized as:

none/primary (1), technical (2), secondary (3), university (4) and post-graduate (5). The second is a wealth index (WI) obtained from a list of items taken as indicators of standard of living. These items were: home ownership, number of bathrooms at the place of residence, ownership of cars, bicycles, fridge/freezer/dishwasher, TVs, radios, CD/DVD players, vacuum cleaner, washing machine and availability of domestic service. We used polychoric principal component analysis to examine the variability of the sample and retained the first principal component as a WI.

Volunteers were asked for their self-perception of African, European, and Native American ancestry. This was understood as a rough estimation of the proportion of their ancestors that could have a particular continental origin. We proposed a five point scale, expressed as 20% per cent brackets (and in words): 1) 0-20% (none or very low), 2) 20-40% (low), 3) 40-60% (moderate), 4) 60-80% (high) and 5) 80-100% (very high or total). Allowance was provided for recording ancestries outside Africa, Europe and America. The questionnaire also recorded information on: the place of birth of the volunteer and (when known) of their ancestors (up to grandparents), various racial/ethnicity categories (such as those used colloquially and by the census in each country), as well as on individual's attitudes towards human genetics research, ancestry studies and racial discrimination. Analyses of the data from the full socioeconomic questionnaire will be published separately.

Genetic admixture estimation

In order to select 30 markers highly informative for estimating African/European/Native American ancestry, we started from the list of 5,000 markers, highly informative for world-wide continental ancestry estimation, identified by Paschou et al (2010)(36) using the approach of Rosenberg et al. (2003)(37) based on the worldwide CEPH-HGDP cell panel genotyped with Illumina's 610 chip (including data for about 600,000

SNPs(38). The full list of these 5,000 markers is at:
http://www.cs.rpi.edu/~drinep/HGDPAIMS/WORLD_5000_INFAIMs.txt. Of these, allele genotype data is available in Native Americans for 3,848 markers(39), of which 2,392 markers have been placed on subsequent Illumina chip products. This subset of markers was retained for selection of those to be typed here in order to facilitate subsequent data comparison and integration. We ranked the 2,392 markers based on allele frequency differences in European-Native American or European-African samples. Amongst markers with the highest inter-continental allele-frequency differences we selected those with lowest heterozygosity in Native Americans (so as to reduce the effect of variable allele frequencies between Native Americans on ancestry estimation). Of the final set of 30 markers retained, 13 are monomorphic in 408 Native Americans (from 47 populations Mexico Southwards), the rest have minor allele frequencies ranging from 0.01 to 0.15 (median 0.06) in that group of populations. This set of 30 markers produces individual ancestry estimates in a sample of Colombians recently included in a genome-wide association study that used Illumina's 610 chip(40) with correlations of ~0.7 compared with estimates obtained from the chip data, and identical mean population estimates. The list of markers typed is provided in Supplementary Table S1. Genotyping was carried out by KBiosciences (Hoddesdon, Herts, UK; <http://www.kbioscience.co.uk/>). Proportions of African, European and Native American ancestry (and their standard errors) were estimated using the ADMIXTURE software(41), using supervised runs (unsupervised runs produced very similar estimates) and standard errors obtained by bootstrap using the program default parameters. Data from a total of 876 individuals sampled in putative parental populations were used in ancestry estimation. These were selected from HAPMAP, the CEPH-HDGP cell panel(38) and from published Native American data(39) as follows: 169 Africans (from 5 populations from Sub-Saharan West Africa), 299 Europeans (from 7 West and South European populations) and 408 Native

Americans (from 47 populations from Mexico Southwards). The full list of the putative parental population samples (and their sizes) is provided in Supplementary Table S2.

Geographic analyses

Information on birthplaces for volunteers, their parents and grandparents was consolidated into a unique list of locations (for all countries except Peru, due to a large number of missing data in that country). These placenames were organized into three fields: city/municipality, region/state and country. Data consistency was checked by comparing field names, e.g. name of a state into the list of city/states for each country. The list of placenames was converted via the Google Maps Geocoding API (<https://developers.google.com/maps/documentation/geocoding/>) into geographic coordinates. Query results were checked for consistency to verify that the geocoding process has identified all places properly. The GeodesiX software (<http://www.calvert.ch/geodesix/>) was used for the geocoding query. The final birthplaces datafile containing coordinates was represented into a searchable map using Google Fusion Tables (<http://www.google.com/drive/apps.html#fusiontables>).

Geographic maps displaying individual admixture estimates were obtained using the software ArcGis 9.3 (<http://www.esri.com/software/arcgis>). The cartographic database was geo-referenced to the SIRGAS geodesic system (Geocentric Reference System for the Americas <http://www.ibge.gov.br/home/geociencias/geodesia/sirgasing/index.html>) using a Universal Transverse Mercator projection. Krirging interpolation was used to adjust for uneven sampling across geographic regions. Corel-DRAW X3 (Corel Corporation, Ottawa, Canada) was used to edit the map images.

Statistical analyses

Linear or logistic regression was used to examine the effect of genetic ancestry or marker genotypes, and selected covariates on phenotypic variation. Analyses were performed using the R package (www.r-project.org)(42).

RESULTS

Genetic ancestry

Documentary information indicates that the geographic origin of the individuals contributing to the current gene pool of Latin America varied greatly. It is also well established that there is genetic substructure in the regions where these individuals originated. Here we use the terms “African(s)”, “European(s)” and “Native American(s)” to refer broadly to different geographic origins without assuming a biological homogeneity, or strict distinctiveness, of these regions. Our main aim here is to obtain ancestry estimates consistently across samples as an heuristic allowing us to examine patterns of variation in ancestry and their correlation with variation in phenotype and self-perception. We did not attempt to perform a refined examination of the recent geographic origin for the ancestors of individuals. Since our focus is not on fine-grained individual ancestry but on the correlation of broad continental ancestry with other variables we estimated it with a relatively small number of (highly informative) markers in a large number of individuals. We also do not attempt to draw a comprehensive picture of the admixture landscape of Latin America (the individuals studied here are not meant to be representative, in a statistical sense, of the countries where the samples were collected). In what follows we use the country names simply as labels indicating where recruitment took place.

The sample collected in Brazil has a highly predominant European ancestry (median 87%) and relatively low Native American and African ancestries (both medians <10%) (Table 1). The samples collected in Chile, Mexico and Perú have highest median Native

American ancestry (46%, 58% and 66% respectively) and lowest median African ancestry (1%, 3% and 1% respectively). The sample collected in Colombia is the most diverse in showing a predominantly European ancestry (median 62%) but substantial Native American admixture (median 28%) and the highest African component (median 7%). There is considerable heterogeneity across ascertainment sites in the distribution of individual ancestry estimates (Figure 1). The Mexican sample is characterized by the widest spread of individual ancestry along the Native American-European axis. By contrast the Chilean sample, shows a greater concentration around the median. The Brazilian, and particularly the Colombian samples, show a greater spread of ancestry along the African axis, with some individuals having a high (>70%) estimated African ancestry. Although most individuals studied here were born near the main ascertainment sites, there is some geographic spread of their birthplaces. When examining the geographic distribution of individual ancestry estimates within each country we observe considerable regional variation but with some broad trends being apparent (Figure 2). In Brazil (Figure 2A) European ancestry is highest in the south, African ancestry in the North East and Native American ancestry in the west. In Chile (Figure 2B) highest European ancestry is seen around the main urban areas of the north and centre of the country with Native ancestry predominating elsewhere and a north to south gradient of decreasing African ancestry. In Colombia (Figure 2C), African ancestry is highest in the coastal regions (particularly on the Pacific), Native ancestry being highest in the South West and in the East of the country, European ancestry being highest in central areas. In Mexico (Figure 2D) Native American ancestry is particularly high in the south, the north showing the highest proportion of European Ancestry (Figure 2b). Other than geographic variation, individual genetic ancestry also shows significant variation with SEP, European ancestry increasing with both higher educational attainment and wealth. This trend is

significant in the global sample ($r= 0.32$ P-value $<2.2 \cdot 10^{-16}$ and $r=0.06$ P-value = $3.05 \cdot 10^{-05}$ for EA and WI respectively) and at each of the sampling sites (Supplementary Note S1).

Phenotypic diversity and genetic ancestry

Table 2 and Figure 3 summarize the distribution of quantitative and categorical phenotypes (respectively) across samples. As is the case for the ancestry estimates, we find a significant heterogeneity for most of the phenotypes examined. For traits differing in frequency between continental populations, we usually observe a variation in frequency across samples consistent with the ancestry estimates (Table1). For instance, the (highly predominantly European) Brazilian sample shows lighter pigmentation (of skin, hair and eyes) and greater height than the samples from the other countries, which show higher Native American ancestry. When comparing eye color groups we observe an increase in European ancestry for the sequence of eye colors: black, brown, green, honey and blue (Figure 4). Amongst hair color groups (Figure 4) there is an increase in European ancestry from black to brown to blond, the red hair group showing a European ancestry between the black and brown hair color groups. Similarly, there is an increase in European ancestry for the groups with straight, wavy and curly hair, with the frizzy hair group showing the lowest European ancestry (Figure 4). Relative to the overall sample the frizzy hair group has markedly higher African ancestry (median of 0.42 v. 0.04).

Regression of phenotypic variation on genetic ancestry demonstrates a significant effect (P-value $< 10^{-3}$, using a Bonferroni multiple testing correction) for most of the traits examined (Table 3 and Figure 5). Accounting for sex, sampling site, age, and SEP, higher European ancestry is associated with increased height, lower body mass index (BMI) and central obesity (ICO), lighter pigmentation (of hair, skin and eyes), greater hair curliness and male pattern baldness (with hair greying approaching significance). Similarly, higher African ancestry is associated with increased height, greater skin pigmentation and greater hair

curliness type. The proportion of phenotypic variance explained by ancestry is highest for skin pigmentation (19%) and is also substantial for hair shape (10%) and color of eyes and hair (4% and 5 %, respectively) but considerably smaller for the other phenotypes. We also observe highly significant effects of EA (P-value $3.87 \cdot 10^{-13}$) and age (P-value $< 2 \cdot 10^{-16}$) on height with height increasing for younger individuals at a rate ~ 1 cm for every 10 years age difference (Supplementary Note S2). Accounting for height and BMI (as well as for the previous covariates) ancestry also has significant effects on facial variation both in terms of size and shape. Higher European ancestry is associated with an overall smaller face size (as estimated by centroid size). We also examined the effect of ancestry on all pair-wise inter-landmark distances (427, after averaging the 203 distances between symmetric landmarks on both sides of the face), which combine shape and size information. Of these, a total of 133 and 2 show significant effects of European and Africa ancestry, respectively (P-values $< 10^{-6}$, assuming a conservative Bonferroni multiple testing correction; Supplementary Table S3). The most significant effects of European ancestry (P $< 10^{-10}$) mainly involve distances between landmarks placed on the lips and nose. Face shape variation, independent of size, was assessed via Principal Component Analysis of procrustes coordinates. Significant effects of European ancestry were detected for PCs 1 and 3-5, while African ancestry impacts on PCs 1, 2 and 4 (Table 3, Supplementary Note S3 and Supplementary Figure S2). These 5 PCs account for $\sim 55\%$ of the variation in face shape captured by the 36 landmarks placed on the facial photographs, with ancestry explaining up to 5% of the variance in PC scores (for PC4). Examination of the correlation between inter-landmark distances and facial PCs (results not shown), indicates that the highest correlation of distances between landmarks of the lips and nose is seen with PC4, consistent with this PC showing the largest proportion of variance explained by ancestry (Table 3). Regression of individual shape scores (providing a single value for shape change) obtained from landmark coordinates as proposed by Drake and

Klingenberg(43), agree with a significant effect of ancestry on overall face shape (Table 3 and Figure 3).

Since the markers used for ancestry estimation have allele frequencies that are highly differentiated between Africans, Europeans and Native Americans, they constitute candidate variants potentially impacting on traits with variable frequencies between these populations. We therefore tested for association between the 30 markers typed and the phenotypes showing ancestry effects (with genome-wide ancestry, age, sampling site, SEP and sex as covariates). Marker rs260690, located in the first intron of the EDAR gene, was found to have a highly significant effect on hair type ($p=10^{-16}$). Examination of HapMap data for Mexican Americans (MEX) indicates high linkage disequilibrium ($D'=0.98$) between this SNP and rs3827760 an A to G variant coding for a p.Val370Ala substitution in EDAR. The derived G allele at rs3827760 has a high frequency in CHB+JPT (0.95) but is rare in CEU+TSI (frequency 0.05) and this allele has recently been associated with increased cross-sectional hair thickness in East Asians as well as in mouse models(44-46). Marker rs356652, 8 Mb from marker rs260690, also shows significant association with hair type (P-value = 1.94×10^{-07}). These two marker are in moderate linkage disequilibrium ($D'=0.52$ in the full sample) and accounting for rs260690 in the regression model markedly attenuates the evidence for association of hair type with rs356652 (P-value 0.004), confirming that this signal of association stems from LD with variants in the EDAR gene. We also found a highly significant association (P-value = 1.58×10^{-12}) between marker rs717225, located in the third intron of the ARHGEF1 gene, and skin pigmentation. In the HAPMAP dataset, the ancestral C allele at rs717225 is found at a frequency of 90% in YRI and 0.01 in CEU and is absent from available Native American data (Supplementary Table S1).

Genetic ancestry and self-perception

We observe a moderate, highly significant, correlation of self-perception of continental ancestry with the genetic estimates (ranked into five bands at 20% increments): America: $r=0.48$ $P < 2.2 \times 10^{-6}$, Europe: $r=0.48$ $P < 2.2 \times 10^{-6}$, Africa: $r=0.32$ $P < 2.2 \times 10^{-6}$. However, there is a trend at increasing Native American and African ancestry for self-perception to exceed the genetic estimates (Figure 6). Correspondingly, at decreasing European ancestry there is a trend for the genetic estimates to exceed self-perception (Figure 6). Regression analysis of the difference between individual self-perceived and genetically estimated ancestry (Table 4) indicates a significant effect of skin pigmentation on this difference: individuals with lower skin pigmentation tend to overestimate their European ancestry while individuals with higher pigmentation overestimate their Native American and African ancestries. Similarly, lighter eye color is associated with a significant overestimation of European ancestry and underestimation of Native American ancestry (but not African ancestry). Overall, men overestimate their European ancestry and underestimate their African ancestry, relative to women. There is also a significant variation between countries in that individuals from the Chilean, Colombian, and Mexican samples tend to overestimate their Native American ancestry and underestimate their European and African ancestries relative to the Brazilian sample.

DISCUSSION

The paleo-anthropological record shows that the physical appearance of humans has undergone considerable changes throughout evolution, presumably reflecting the interplay of various environmental and genetic factors. Previous studies have also shown that current human phenotypic diversity shows some clinal geographically structure, usually to a similar degree as seen with “neutral” genetic markers. For instance, it has been estimated that about 90% of both craniofacial variation and genetic diversity is present within continents and only

about 10% of this variation is accounted for by differentiation between continents(47). For certain phenotypes however (such as skin pigmentation), there often is greater differentiation between than within populations and this pattern has been interpreted as consistent with the action of selection on these phenotypes. In the case of skin pigmentation possibly due to variable levels of sunlight exposure, which could have a range of physiological effects potentially affecting survival(48, 49). Consistent with this scenario, recent genome-wide analyses aimed at detecting selection at the molecular level have often implicated genes involved in pigment synthesis(12, 50, 51).

Recently admixed populations, such as Latin Americans, offer an opportunity to characterize the genetic basis of traits showing geographic structure between the populations that contributed to admixture in the region; in this case, mostly Native Americans and individuals arriving from Africa and Europe. The analyses performed here illustrate the great diversity of Latin Americans, with an extensive variation in genetic ancestry and physical appearance. Although not designed to be statistically representative of the populations sampled, the geographic distribution of the ancestry estimates obtained here are consistent with the demographic history of the regions examined. The relatively low global estimates of African ancestry are consistent with the sampling sites not being located in areas of major historical African immigration. The relatively high Native American ancestry seen in the Peruvian, Mexican and Chilean samples, is consistent with the high pre-Columbian Native population density in Meso-America and the Andes, while the Colombian and Brazilian samples come from regions with comparatively lower pre-Columbian Native population densities(2, 3). In addition, the Brazilian sample is predominantly from a region (Rio Grande do Sul) that has received a large influx of European immigrants in the last two centuries(2, 3, 52). Although the sampling sites examined are not located in areas of major African immigration, the samples from Colombia and Brazil show the largest estimates of African

ancestry, consistent with the greater African immigration to those two countries, relative to Perú, México and Chile and these concentrate particularly in the coastal regions of Colombia and the Brazilian North East(2, 53-55). Within each country the geographic distribution of individual ancestry estimates is also consistent with regional demographic history. In Mexico most of the current Native population concentrates to the South of the country, European immigration having concentrated mostly in the centre and North of the country. In Colombia Native populations occupy predominantly peripheral areas, the central, highland areas having received most European immigration, while Africans predominantly settled in coastal areas and major inland river valleys(55, 56). In Brazil, most African settlement occurred in the North East, the South having received a major European population influx in the last two centuries, while most Native populations currently concentrate to the West(2, 3, 52).

The pattern of variation we observe between phenotypes and estimated genetic ancestry is generally consistent with available information regarding the variation in frequency of these phenotypes in Natives from America, Europe and Africa, and the generally complex genetic basis of these traits. Europeans usually have lighter skin pigmentation than Native Americans or Africans and also show frequent variation in eye and hair color, which is rare elsewhere. A number of genes influencing these traits have been recently identified in European(25, 57-60). Consistent with this geographic pattern, increased European ancestry is associated with lighter skin, hair and eye pigmentation and higher African ancestry with greater skin pigmentation (Table 3 and Figure3). An unexpected pattern was found for red hair. This trait is essentially absent in Native Americans and sub-Saharan Africans, and found at non-negligible frequencies in Europeans where it is associated with very low skin pigmentation, poor tanning response and high sun sensitivity. However, the Latin American individuals with red hair examined here were found to have lower European ancestry than individuals with blond or brown hair (Figure 3), showing

ancestry values closer to the population median. This observation could relate to red hair being determined by variants at relatively few genes (most prominently the melanocortin 1 receptor *MC1R*)(61), thus making this trait a particularly unreliable indicator of individual genome-wide ancestry. Of the SNPs examined here, rs717225, located in the third intron of *ARHGEF1* showed a strong association with skin pigmentation. The C allele of rs717225 is seen at high frequency in Africans and at very low frequency in Europeans and has not been reported in Native Americans (Supplementary Table S1) and this allele is associated with increased pigmentation in the Latin American sample examined here. Rho GTPases play a key role in cellular processes mediated by G-protein coupled receptors and recent genome-wide RNA-mediated interference experiments have implicated *ARHGEF1* in melanogenesis(62). These experiments have shown a direct impact of this gene on the expression of tyrosinase and of *MITF* (the Microphthalmia-associated transcription factor, a master regulator of melanocyte development)(62). *MC1R* is a G-protein coupled receptor whose association with red hair (as mentioned above) is well established(58) but this is the first time that variants in *ARHGEF1* are associated with human pigmentation variation.

Macroscopic hair type (curliness v. straightness) is a highly heritable trait showing considerable variation between human populations(57, 63). The hair of East Asians and Native Americans has been described as commonly coarse in texture, with individual fibres that are straight, thick and circular in cross-section(64). We found a strong effect of ancestry on hair type, with straight hair frequency increasing in individuals with greater Native American ancestry. This observation was further refined by the finding of a highly significant association of hair type with marker rs260290 in *EDAR*. A recent study found evidence of association in Asians of the p.370Ala substitution in *EDAR* (rs3827760>G) with hair cross-sectional thickness(44, 45). No genotype data is currently available for marker rs3827760 in Native Americans. However, there is strong LD across the *EDAR* region and an imputation

analysis of available high-density chip data for Native Americans(39) indicates that these populations are essentially fixed for the derived rs3827760 G allele (results not shown). Position 370 in EDAR is located in an intracellular domain required for the interaction with the EDAR-binding death domain adapter protein EDARADD in a pathway involved in formation of the hair placode(65) and experimental evidence suggests that the p.370Ala variant has higher activity than the ancestral p.370Val variant(46). This variant is seen at high frequency in East Asians but is absent from Europeans and Africans. The population differentiation and haplotypic structure of the EDAR region has been interpreted as evidence of recent strong selection on this gene(50, 66, 67). The absence of the p.370Ala variant in Europeans indicates that hair straightness has evolved more than once in human populations and consistent with this, variants in the trichohyalin gene have recently been shown to be associated with hair type in Europeans(68). The nature of the selection pressure however remains unclear although it has long been argued that hair type could represent an adaptation to variable levels of heat and UV radiation or that it could have been influenced by sexual selection(57). Further studies should clarify the role of EDAR variants in the development of the hair follicle and developmentally related structures such as sweat glands and teeth, helping to shed light on the selective forces that might have shaped diversity at this gene.

Hair greying and androgenetic/senescent alopecia are amongst the most obvious manifestations of aging in humans, yet their molecular basis is poorly understood. Both processes result from a weathering of the hair shaft and the aging of the hair follicle, leading to a decrease in melanocyte function and in hair production. Experimental evidence indicates that stem cell death and oxidative stress play a key role in this process(69, 70) and recent studies emphasize the importance of Notch2 signalling in the maintenance of melanocyte stem cells and hair pigmentation(71). Changes in the hair follicle associated with hair greying and loss, therefore represent a model system in which to study “intrinsic” aging (72) and

where to test anti-aging therapies. Various lines of evidence testify to the importance of genetic factors in hair loss and greying, the most extreme being Mendelian disorders associated with hair loss or premature aging(29, 73-76). A twin study estimated the heritability of androgenetic alopecia amongst Europeans at about 79%(29) and genome-wide studies in Europeans have now identified several loci associated with this trait(77, 78). It has been estimated that androgenetic alopecia affects ~45% of Caucasian men by the age of 49 and 70% by 79. Similarly, by 50 years of age, about 50% of Europeans have ~50% grey hair(75, 76, 79). Although no similar studies have been carried out in Amerindians, classical physical anthropology studies indicate that hair greying and androgenetic alopecia are rarer, less severe and of later onset in Amerinds than in Caucasians(80, 81). In the sample examined here we do find that greater European ancestry correlates with significantly higher rates of male balding and also see an increase of European ancestry with greater hair greying but only approaching statistical significance. It will be interesting to follow-up these results as elucidating the genetic basis of these traits could help increase our understanding of the aging of the hair follicle and the health risks associated by male pattern baldness.

The effect of genetic ancestry on height and related anthropometric phenotypes (BMI, ICO) is consistent with the polygenic basis of these traits documented by recent genome-wide association studies in Europeans and the existence of trait loci with allele frequencies varying between continental populations(82-84). Our results indicate that about 1% of the variation in height in the Latin American individuals examined results from genetic ancestry variation. Height is also under strong environmental influence(18) and, in line with this, we detect a significant effect of SEP in our sample. That the two SEP indicators examined here underestimates the amount of environmental variation is suggested by the significant effect of date of birth on height, with individuals born more recently tending to be taller than older individuals. The increase in height for individuals born more recently (~0.1 cm/year)

estimated here is similar to that estimated from large longitudinal surveys in Latin America (~1 cm per decade in the last century), an observation that has been interpreted as resulting from the historical improvement in living standards across the region(18, 85, 86).

Vertebrate craniofacial morphology shows great variation within species and diversification between species, where changes in craniofacial features closely match adaptive radiations to particular ecological niches. However, the genetic basis of this variation is poorly understood. In humans, twin and family studies have estimated a heritability of 60-90% for a range of facial traits(87) and anthropological studies have documented a geographic structure for facial features, which is of a similar degree to that seen with neutral genetic data(47, 88, 89). The characterization of single gene mutations in experimental organisms and in patients with dysmorphologies, has identified common signalling cascades during vertebrate craniofacial development(87, 90-92). Work in experimental organisms has often focused in cases where there is a main direction of shape change, for instance in the offspring of populations with some differentiated anatomical structures(93-100). A somewhat related situation applies here. We observe that ancestry impacts broadly on facial features described previously as varying between Africans, Europeans and Native Americans(101-103): mainly width and height of the face, facial flatness, position of the glabella and frontotemporal points and the relative size and position of lips (a fuller description of the main aspects of face shape variation associated with each PC is presented in the Supplementary Note S3 and Supplementary Figure S2). Two recent genome-wide association studies of face shape variation in Europeans have identified a few loci associated with aspects of face shape variation(104, 105) and both studies indicate that variants in the PAX3 gene influence human face shape. These analyses are consistent with the view that facial phenotypes are influenced by a large number of genetic variants of relatively small effect size (106, 107). The ancestry effect that we detect for face shape and

size but not for cranium size agrees with the notion of a greater developmental and evolutionary constraint on neurocranium than on facial variation, as suggested by several morphological analyses (108-113). These observations are in line with proposals that facial features of anatomically modern humans include environmental adaptations related to regionally differing selection pressures associated with the late Pleistocene dispersal of our species (114-117).

The correlation of individual ancestry with the various phenotypes discussed above and the specific effect of variants in *EDAR* and *ARHGGEF1* in hair type and skin pigmentation, illustrate the informativeness of Latin Americans for characterizing the genetic basis of traits geographically structured across the continents contributing to admixture in the region. Latin America also offers a special opportunity to examine the relationship between biological diversity and certain social variables. Our analyses emphasize the continuing correlation of education and wealth with genetic ancestry in the region(118, 119). Although admixture in Latin America has taken place since the beginning of the European colonial expansion (some 25 generations ago), this has occurred within the context of strong socioeconomic stratification, with European immigrants and their descendants usually having a higher socioeconomic position than people with greater Native and African ancestry(1, 120, 121). As an indicator of associated environmental variation and wellbeing (i.e. stress, nutrition, healthcare, medical history) SEP is thus an important confounder that needs to be taken into account when evaluating ancestry effects on phenotype, including disease, in these populations, as we see here for height and have reported previously in relation to type 2 diabetes(118, 119). The effect of age on height detected (consistent with historical surveys(18, 85, 86) suggests that the two indicators of SEP used here are likely to capture only part of the environmental variation affecting this trait and it will be important to develop more specific indicators of environmental variation for each phenotype that capture relevant

factors. Our data also underline the major role played by physical appearance in self-perception of ancestry. As individual genealogical information on ancestry is usually limited to 2-3 generations and since admixture in Latin America can go back over 20 generations it is likely that physical appearance will often influence self-perception of ancestry. Based on the geographic structure of genetic and phenotypic diversity it is to be expected that self-perception and genetic ancestry correlate, as we observe. However, since certain visible traits show a higher level of geographic structure than the genome average, perception of ancestry based on these traits is biased relative to genetically estimated ancestry. This is illustrated for instance, by the different effect that we see for skin pigmentation and African self-perception or for eye/hair color and European self-perception (Table 4). Our findings are thus consistent with the notion that self-perception is “skin deep” in that it is heavily influenced by certain visible attributes and biased relative to genetic ancestry. We also find evidence that this bias is influenced by social factors possibly related to a variable perception regarding the extent of geographic structure of visible phenotypes, as suggested by the significant differences that we find between countries and sexes. These observations are in line with social phenomena such as the so-called “one drop rule” in the US, according to which an individual with some African ancestry is identified as African, thus constituting an extreme form of perception bias.

As the product of biological evolution our species shows complex patterns of phenotypic and genetic variation, including some geographic structure. However, perception of the geographic distribution of human genetic and phenotypic diversity impinges upon the sensitive subjects of race and racism. Both of these notions are social constructs associated with grossly inaccurate interpretations of human biological variation. It is our belief that a fuller understanding of this variation, and of its social correlates, will help undermine the

social prejudices upon which racism is based. We trust that this study makes a contribution in that direction.

ACKNOWLEDGMENTS

We are extremely grateful to the volunteers for their enthusiastic support of this research.

This work was funded by grants from the Leverhulme Trust (F/07 134/DF to A.R-L) and BBSRC (BB/I021213/1 to A.R-L). We thank Ken Weiss and Peter Wade for comments on the manuscript and Lucía Castillo and Ana Carolina Orozco for technical assistance.

REFERENCES

1. Morner M (1967) *Race Mixture in the History of Latin America* (Little Brown & Company).
2. Sanchez-Albornoz N (1974) *The population of Latin America: a history* (University of California Press, Berkeley).
3. Salzano FM & Bortolini MC (2002) *The Evolution and Genetics of Latin American Populations* (Cambridge University Press).
4. Carvajal-Carmona LG, *et al.* (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* 67(5):1287-1295.
5. Carvalho-Silva DR, Santos FR, Rocha J, & Pena SD (2001) The phylogeography of Brazilian Y-chromosome lineages 1. *Am.J.Hum.Genet.* 68(1):281-286.
6. Alves-Silva J, *et al.* (2000) The Ancestry of Brazilian mtDNA Lineages. *Am.J.Hum.Genet.* 67(2):444-461.
7. Bedoya G, *et al.* (2006) Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc. Natl. Acad. Sci. U. S. A.* 103(19):7234-7239.
8. Wang S, *et al.* (2008) Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet* 4(3):e1000037.
9. Bryc K, *et al.* (2010) Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci. U. S. A.* 107 Suppl 2:8954-8961.
10. Boyd-Bowman P (1976) Patterns of Spanish Emigration to the Indies until 1600. *Hispanic American Historical Review* 66:580-604.
11. Ventura Santos R, *et al.* (2009) Color, Race, and Genomic Ancestry in Brazil: Dialogues between Anthropology and Genetics. *Curr. Anthropol.* 50(6):787-819.
12. Myles S, Somel M, Tang K, Kelso J, & Stoneking M (2007) Identifying genes underlying skin pigmentation differences among human populations. *Hum. Genet.* 120(5):613-621.
13. Klimentidis YC, Miller GF, & Shriver MD (2009) Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am. J. Phys. Anthropol.* 138(4):375-383.
14. Gao X, Simon KC, Han J, Schwarzschild MA, & Ascherio A (2009) Genetic determinants of hair color and Parkinson's disease risk. *Ann. Neurol.* 65(1):76-82.

15. Duffy DL, *et al.* (2010) Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. *J. Invest. Dermatol.* 130(2):520-528.
16. Yamada T, Hara K, Umematsu H, & Kadowaki T (2013) Male pattern baldness and its association with coronary heart disease: a meta-analysis. *BMJ open* 3(4).
17. Wright JL, Page ST, Lin DW, & Stanford JL (2010) Male pattern baldness and prostate cancer risk in a population-based case-control study. *Cancer epidemiology* 34(2):131-135.
18. McEvoy BP & Visscher PM (2009) Genetics of human height. *Economics and human biology* 7(3):294-306.
19. Friedlander AH & Scully C (2010) Diagonal ear lobe crease and atherosclerosis: a review of the medical literature and oral and maxillofacial implications. *J. Oral Maxillofac. Surg.* 68(12):3043-3050.
20. Kayser M & de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature reviews. Genetics* 12(3):179-192.
21. Gilman SL (1985) *Difference and pathology : stereotypes of sexuality, race, and madness* (Cornell University Press, Ithaca) p 292 p.
22. Cho MK & Sankar P (2004) Forensic genetics and ethical, legal and social implications beyond the clinic. *Nat. Genet.* 36(11 Suppl):S8-12.
23. Ossorio PN (2006) About face: forensic genetic testing for race and visible traits. *J. Law. Med. Ethics* 34(2):277-292.
24. M'Charek A (2008) Silent witness, articulate collective: DNA evidence and the inference of visible traits. *Bioethics* 22(9):519-528.
25. Sulem P, *et al.* (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* 39(12):1443-1452.
26. Hamilton JB (1951) Patterned loss of hair in man; types and incidence. *Ann. N. Y. Acad. Sci.* 53(3):708-728.
27. Norwood OT (1975) Male pattern baldness: classification and incidence. *South. Med. J.* 68(11):1359-1365.
28. Lotufo PA, Chae CU, Ajani UA, Hennekens CH, & Manson JE (2000) Male pattern baldness and coronary heart disease: the Physicians' Health Study. *Arch. Intern. Med.* 160(2):165-171.
29. Rexbye H, *et al.* (2005) Hair loss among elderly men: etiology and impact on perceived age. *J. Gerontol. A. Biol. Sci. Med. Sci.* 60(8):1077-1082.
30. Morton DJ, Kritz-Silverstein D, Riley DJ, Barrett-Connor EL, & Wingard DL (2007) Premature graying, balding, and low bone mineral density in older women and men: the Rancho Bernardo study. *J. Aging Health* 19(2):275-285.
31. Comas J (1984) *Manual de Antropología Física* (Universidad Nacional Autónoma de México, México).
32. Clarys P, Alewaeters K, Lambrecht R, & Barel AO (2000) Skin color measurements: comparison between three instruments: the Chromameter(R), the DermaSpectrometer(R) and the Mexameter(R). *Skin Res Technol* 6(4):230-238.
33. Van den Kerckhove E, Staes F, Flour M, Stappaerts K, & Boeckx W (2001) Reproducibility of repeated measurements on healthy skin with Minolta Chromameter CR-300. *Skin Res Technol* 7(1):56-59.
34. Klingenberg CP (2008) MorphoJ (Faculty of Life Sciences, University of Manchester, UK. http://www.flywings.org.uk/MorphoJ_page.htm).
35. Wiley DF, N. Amenta, D. A. Alcantara, D. Ghosh, Y. J. Kil, E. Delson, W. Harcourt-Smith, F. J. Rohlf, K. St. John, and B. Hamann. 2005. (2005) Evolutionary morphing. *Proceedings of the IEEE Visualization 2005* (Proceedings of the IEEE Visualization 2005).
36. Paschou P, Lewis J, Javed A, & Drineas P (2010) Ancestry informative markers for fine-scale individual assignment to worldwide populations. *J. Med. Genet.* 47(12):835-847.
37. Rosenberg NA, Li LM, Ward R, & Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 73(6):1402-1422.

38. Li JZ, *et al.* (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866):1100-1104.
39. Reich D, *et al.* (2012) Reconstructing Native American population history. *Nature* 488(7411):370-374.
40. Scharf JM, *et al.* (2012) Genome-wide association study of Tourette's syndrome. *Mol. Psychiatry*.
41. Alexander DH, Novembre J, & Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655-1664.
42. R_Development_Core_Team (2010) R: A language and environment for statistical computing (Vienna).
43. Drake AG & Klingenberg CP (2008) The pace of morphological change: historical transformation of skull shape in St Bernard dogs. *Proceedings. Biological sciences / The Royal Society* 275(1630):71-76.
44. Fujimoto A, *et al.* (2008) A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet.* 17(6):835-843.
45. Fujimoto A, *et al.* (2008) A replication study confirmed the EDAR gene to be a major contributor to population differentiation regarding head hair thickness in Asia. *Hum. Genet.* 124(2):179-185.
46. Mou C, *et al.* (2008) Enhanced ectodysplasin-A receptor (EDAR) signaling alters multiple fiber characteristics to produce the East Asian hair form. *Hum. Mutat.* 29(12):1405-1411.
47. Relethford JH (2002) Apportionment of global human genetic diversity based on craniometrics and skin color. *Am. J. Phys. Anthropol.* 118(4):393-398.
48. Jablonski NG (2012) Human skin pigmentation as an example of adaptive evolution. *Proc Am Philos Soc* 156(1):45-57.
49. Jablonski NG & Chaplin G (2012) Human skin pigmentation, migration and disease susceptibility. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367(1590):785-792.
50. Sabeti PC, *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449(7164):913-918.
51. Williamson SH, *et al.* (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3(6):e90.
52. (IBGE) IBdGeE (2000) *Brasil: 500 anos de povoamento*. (IBGE, Rio de Janeiro).
53. Curtin PD (1969) *The Atlantic slave trade: a census* (University of Wisconsin Press, Madison).
54. Thomas H (1997) *The Slave Trade* (Simon and Schuster).
55. Safford F & Palacios M (2001) *Colombia: Fragmented Land, Divided Society (Latin American Histories)* (Oxford University Press).
56. Codazzi IGA (2002) *Atlas de Colombia* (IGAC, Bogota, Colombia).
57. Jablonski NG (2006) *Skin: A Natural History* (University of California Press, Berkeley).
58. Sturm RA (2009) Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.* 18(R1):R9-17.
59. Kayser M, *et al.* (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet* 82(2):411-423.
60. Liu F, *et al.* (2010) Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet* 6:e1000934.
61. Valverde P, Healy E, Jackson I, Rees JL, & Thody AJ (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat. Genet.* 11(3):328-330.
62. Ganesan AK, *et al.* (2008) Genome-wide siRNA-based functional genomics of pigmentation identifies novel genes and pathways that impact melanogenesis in human cells. *PLoS Genet* 4(12):e1000298.

63. Medland SE, Zhu G, & Martin NG (2009) Estimating the heritability of hair curliness in twins of European ancestry. *Twin research and human genetics : the official journal of the International Society for Twin Studies* 12(5):514-518.
64. Berardesca E, Leveque J-L, & Maibach HI (2007) *Ethnic skin and hair* (Informa Healthcare, New York) pp xii, 269 p.
65. Schmidt-Ullrich R & Paus R (2005) Molecular principles of hair follicle induction and morphogenesis. *Bioessays* 27(3):247-261.
66. Carlson CS, *et al.* (2005) Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* 15(11):1553-1565.
67. Williamson SH, *et al.* (2005) Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 102(22):7882-7887.
68. Medland SE, *et al.* (2009) Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 85(5):750-755.
69. Nishimura EK, Granter SR, & Fisher DE (2005) Mechanisms of hair graying: incomplete melanocyte stem cell maintenance in the niche. *Science* 307(5710):720-724.
70. Arck PC, *et al.* (2006) Towards a "free radical theory of graying": melanocyte apoptosis in the aging human hair follicle is an indicator of oxidative stress induced tissue damage. *FASEB J.* 20(9):1567-1569.
71. Schouwey K & Beermann F (2008) The Notch pathway: hair graying and pigment cell homeostasis. *Histol. Histopathol.* 23(5):609-619.
72. Vijg J & Campisi J (2008) Puzzles, promises and a cure for ageing. *Nature* 454(7208):1065-1071.
73. Ali G, Chishti MS, Raza SI, John P, & Ahmad W (2007) A mutation in the lipase H (LIPH) gene underlie autosomal recessive hypotrichosis. *Hum. Genet.* 121(3-4):319-325.
74. Azeem Z, *et al.* (2008) Novel mutations in G protein-coupled receptor gene (P2RY5) in families with autosomal recessive hypotrichosis (LAH3). *Hum. Genet.* 123(5):515-519.
75. Tobin DJ & Paus R (2001) Graying: gerontobiology of the hair follicle pigmentary unit. *Exp. Gerontol.* 36(1):29-54.
76. Van Neste D & Tobin DJ (2004) Hair cycle and hair pigmentation: dynamic interactions and changes associated with aging. *Micron* 35(3):193-200.
77. Hillmer AM, *et al.* (2008) Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nat. Genet.* 40(11):1279-1281.
78. Li R, *et al.* (2012) Six novel susceptibility loci for early-onset androgenetic alopecia and their unexpected association with common diseases. *PLoS Genet* 8(5):e1002746.
79. Keogh EV & Walsh RJ (1965) Rate of greying of human hair. *Nature* 207(999):877-878.
80. Cavalli-Sforza LL, Menozzi P, & Piazza A (1994) *The History and Geography of Human Genes* (Princeton U.P., Princeton).
81. Gray J (2000) *Human hair diversity* (Blackwell Science, Abingdon, Oxon) p 70 p.
82. Yang J, *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42(7):565-569.
83. Lango Allen H, *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317):832-838.
84. McQuillan R, *et al.* (2012) Evidence of inbreeding depression on human height. *PLoS Genet* 8(7):e1002655.
85. Meisel A & Vega M (2007) The biological standard of living (and its convergence) in Colombia, 1870-2003. A tropical success story. *Economics and human biology* 5(1):100-122.
86. Salvatore RD, Coatsworth JH, & Challú AIE (2010) *Living standards in Latin American history : height, welfare, and development, 1750-2000* (Harvard University David Rockefeller Center for Latin American Studies : Distributed by Harvard University Press, Cambridge, Mass.) pp iii, 313 p.

87. Carson EA (2006) Maximum likelihood estimation of human craniometric heritabilities. *Am. J. Phys. Anthropol.* 131(2):169-180.
88. Hennessy RJ & Stringer CB (2002) Geometric morphometric study of the regional variation of modern human craniofacial form. *Am. J. Phys. Anthropol.* 117(1):37-48.
89. Manica A, Amos W, Balloux F, & Hanihara T (2007) The effect of ancient population bottlenecks on human phenotypic variation. *Nature* 448(7151):346-348.
90. Tapadia MD, Cordero DR, & Helms JA (2005) It's all in your head: new insights into craniofacial development and deformation. *J. Anat.* 207(5):461-477.
91. Tassabehji M, *et al.* (2005) GTF2IRD1 in craniofacial development of humans and mice. *Science* 310(5751):1184-1187.
92. Tobin JL, *et al.* (2008) Inhibition of neural crest migration underlies craniofacial dysmorphology and Hirschsprung's disease in Bardet-Biedl syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 105(18):6714-6719.
93. Liu J, *et al.* (1996) Genetic Analysis of a Morphological Shape Difference in the Male Genitalia of *Drosophila simulans* and *D. mauritiana*. *Genetics*. 142:1129-1145.
94. Laurie CC, True JR, Liu J, & Mercer JM (1997) An introgression analysis of quantitative trait loci that contribute to a morphological difference between *Drosophila simulans* and *D. mauritiana*. *Genetics* 145(2):339-348.
95. Zeng ZB, *et al.* (2000) Genetic architecture of a morphological shape difference between two *Drosophila* species. *Genetics* 154(1):299-310.
96. Zimmerman E, Palsson A, & Gibson G (2000) Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. *Genetics* 155(2):671-683.
97. Klingenberg CP, Leamy LJ, Routman EJ, & Cheverud JM (2001) Genetic architecture of mandible shape in mice: Effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics* 157(2):785-802.
98. Klingenberg CP, Leamy LJ, & Cheverud JM (2004) Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. *Genetics* 166(4):1909-1921.
99. Kimura T, *et al.* (2007) Genetic analysis of craniofacial traits in the medaka. *Genetics* 177(4):2379-2388.
100. Kenney-Hunt JP, *et al.* (2008) Pleiotropic patterns of quantitative trait loci for 70 murine skeletal traits. *Genetics* 178(4):2275-2288.
101. Martinez-Abadias N, *et al.* (2006) Phenotypic evolution of human craniofacial morphology after admixture: a geometric morphometrics approach. *Am. J. Phys. Anthropol.* 129(3):387-398.
102. Sardi ML, Ramirez RF, Gonzalez-Jose R, & Pucciarelli HM (2005) South Amerindian craniofacial morphology: diversity and implications for Amerindian evolution. *Am. J. Phys. Anthropol.* 128(4):747-756.
103. Hanihara T (2000) Frontal and facial flatness of major human populations. *Am. J. Phys. Anthropol.* 111(1):105-134.
104. Paternoster L, *et al.* (2012) Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am J Hum Genet* 90(3):478-485.
105. Liu F, *et al.* (2012) A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet* 8(9):e1002932.
106. Martinez-Abadias N, *et al.* (2009) Heritability of human cranial dimensions: comparing the evolvability of different cranial regions. *J. Anat.* 214(1):19-35.
107. Martinez-Abadias N, *et al.* (2011) FGF/FGFR signaling coordinates skull development by modulating magnitude of morphological integration: evidence from Apert syndrome mouse models. *PloS one* 6(10):e26425.
108. Enlow DH (1990) *Facial growth* (Saunders, Philadelphia).

109. Ackermann RR, Cheverud JM, Pigliucci M, & Preston K (2004) Morphological integration in primate evolution. *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes*, (Oxford University Press, Oxford), pp 302-319.
110. Lieberman DE, McBratney BM, & Krovitz G (2002) The evolution and development of cranial form in *Homo sapiens*. *Proc.Natl.Acad.Sci.* 99(3):1134-1139.
111. Lieberman DE, Krovitz GE, & McBratney-Owen B (2004) Testing hypotheses about tinkering in the fossil record: the case of the human skull. *J.Exp.Zoolog.B Mol.Dev.Evol.* 302(3):284-301.
112. Bastir M, O'Higgins P, & Rosas A (2007) Facial ontogeny in Neanderthals and modern humans. *Proc.Biol.Sci.* 274(1614):1125-1132.
113. Martinez-Abadias N, Paschetta, C., de Azevedo, S., Esparza, M., Gonzalez-Jose, R. (2009) Developmental and genetic constraints on neurocranial globularity: Insights from analyses of deformed skulls and quantitative genetics. *Evolutionary Biology* 36:37-56.
114. Roseman CC (2004) Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proc.Natl.Acad.Sci.U.S.A* 101(35):12824-12829.
115. Roseman CC & Weaver TD (2004) Multivariate apportionment of global human craniometric diversity. *Am. J. Phys. Anthropol.* 125(3):257-263.
116. Gonzalez-Jose RG, *et al.* (2005) Functional-cranial approach to the influence of economic strategy on skull morphology. *Am. J. Phys. Anthropol.* 128(4):757-771.
117. Harvati K & Weaver TD (2006) Human cranial anatomy and the differential preservation of population history and climate signatures. *The anatomical record. Part A, Discoveries in molecular, cellular, and evolutionary biology* 288(12):1225-1233.
118. Florez JC, *et al.* (2009) Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes. *Diabetologia* 52(8):1528-1536.
119. Campbell DD, *et al.* (2012) Amerind ancestry, socioeconomic status and the genetics of type 2 diabetes in a Colombian population. *PLoS one* 7(4):e33570.
120. Mörner M, Cornell University. Latin American Studies Program., & Columbia University. Institute of Latin American Studies. (1970) *Race and class in Latin America* (Columbia University Press, New York,) pp viii, 309 p.
121. Burkholder MA & Johnson LL (2003) *Colonial Latin America* (Oxford University Press).

Table 1. Sample size, proportion of women, age (median) and estimated genetic ancestry proportion (median) for the sample of each country and in the total sample.

| | Brazil | Chile | Colombia | Mexico | Peru | Total |
|---------|--------|-------|----------|--------|------|-------|
| N | 1,600 | 1,561 | 1,671 | 1,652 | 454 | 6,938 |
| Women | 0.67 | 0.34 | 0.56 | 0.60 | 0.61 | 0.55 |
| Age | 25 | 24 | 23 | 25 | 20 | 24 |
| America | 0.09 | 0.48 | 0.29 | 0.56 | 0.64 | 0.37 |
| Africa | 0.09 | 0.04 | 0.11 | 0.05 | 0.04 | 0.07 |
| Europe | 0.81 | 0.48 | 0.61 | 0.39 | 0.32 | 0.56 |

Table 2. Median of anthropometric features and skin pigmentation in females and males (F/M) of the samples studied

| | Brazil | Chile | Colombia | Mexico | Peru | Total |
|---------------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| Weight (kg) | 60/76 | 61/76 | 56/70 | 60/74 | 56/71 | 59/74 |
| Height (cm) | 162/175 | 159/172 | 160/173 | 158/172 | 158/171 | 160/172 |
| BMI (kg/cm ²) | 23/25 | 25/26 | 22/23 | 24/25 | 23/24 | 23/25 |
| Hip circum. (cm) | 98/100 | 100/102 | 94/95 | 95/97 | 96/99 | 97/99 |
| Waist circum. (cm) | 75/87 | 77/90 | 77/81 | 81/87 | 80/88 | 78/87 |
| ICO | 0.46/0.50 | 0.48/0.52 | 0.48/0.47 | 0.51/0.51 | 0.51/0.52 | 0.49/0.51 |
| Head circum. (cm) | 55/57 | 56/57 | 54/56 | 55/57 | 55/57 | 55/57 |
| MI | 32/32 | 36/35 | 34/33 | 36/35 | 37/37 | 34/35 |

Note: BMI = Body Mass Index; ICO = Index of Central Obesity; MI = Melanin Index

Table 3. Regression of phenotypes on genetic estimates of European and African ancestry.

| Trait | European ancestry | | African ancestry | | R ² | Diff. R ² |
|----------------------|-------------------|-----------------|------------------|-----------------|----------------|----------------------|
| | Coef. | P-value | Coef. | P-value | | |
| Balding | 0.16 | 1.99E-04 | 0.17 | 0.04 | 0.19 | 0.00 |
| Greying | 0.10 | 8.82E-03 | 0.09 | 0.19 | 0.53 | 0.00 |
| Hair shape | 0.49 | 5.29E-15 | 2.43 | 2.00E-16 | 0.13 | 0.10 |
| Hair color | -0.58 | 2.00E-16 | -0.03 | 0.67 | 0.11 | 0.05 |
| Eye color | -1.11 | 2.00E-16 | 0.25 | 0.14 | 0.10 | 0.04 |
| MI | -9.18 | 2.00E-16 | 12.88 | 2.00E-16 | 0.25 | 0.19 |
| Weight | 0.15 | 0.88 | 4.58 | 0.02 | 0.33 | 0.01 |
| Height | 6.77 | 2.00E-16 | 8.03 | 5.05E-16 | 0.54 | 0.01 |
| BMI | -1.92 | 5.43E-09 | -0.61 | 0.32 | 0.18 | 0.01 |
| Waist circum. | -4.22 | 2.05E-06 | -3.20 | 0.05 | 0.27 | 0.01 |
| Hip circum. | -0.20 | 0.78 | 2.56 | 0.06 | 0.12 | 0.01 |
| ICO | -0.05 | 2.00E-16 | -0.04 | 9.63E-06 | 0.23 | 0.02 |
| Head circum. | -0.26 | 0.28 | 0.45 | 0.33 | 0.26 | 0.01 |
| CS | -12.98 | 7.74E-05 | 2.94 | 0.62 | 0.48 | 0.00 |
| PC1(19%) | 0.02 | 1.42E-11 | 0.04 | 1.24E-12 | 0.15 | 0.02 |
| PC2(12%) | -0.01 | 1.00E-03 | -0.02 | 7.80E-06 | 0.02 | 0.00 |
| PC3(10%) | -0.01 | 8.85E-13 | 0.00 | 0.84 | 0.21 | 0.03 |
| PC4(7%) | 0.01 | 2.34E-08 | -0.02 | 6.14E-08 | 0.18 | 0.05 |
| PC5(7%) | 0.01 | 3.16E-05 | 0.00 | 0.60 | 0.05 | 0.00 |
| European shape score | 0.02 | 2.00E-16 | 0.01 | 7.06E-04 | 0.34 | 0.26 |

Note: Being proportions, European, African and American ancestries sum up to 1 and since in this sample African ancestry is very low (median of 7%), we use Native American ancestry as a baseline. All regressions account for age, sex, sampling site and SEP (EA and WI). Regressions for facial traits also account for BMI and height. P-Values < 10⁻³ are shown in bold italic. % Ancestry refers to variation explained by incorporating European and African ancestry in the model (with values greater than 1% highlighted in bold). BMI = Body Mass Index; ICO = Index of Central Obesity; MI = Melanin Index; PC = Principal Component (% in parenthesis refers to phenotypic variance explained by that PC); Diff. R² refers to the difference in R² for the full regression model and a model without European and African ancestry as explanatory variables.

Table 4. Regression of the difference between individual self-perception and genetic ancestry bands.

| | Δ AMERICA | | Δ EUROPE | | Δ AFRICA | |
|---------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | Coef. | P-value | Coef. | P-value | Coef. | P-value |
| Age | 0.00 | 0.93 | 0.00 | 0.87 | 0.00 | 0.16 |
| Sex(M) | -0.05 | 0.13 | 0.26 | 4.25E-15 | -0.12 | 1.87E-07 |
| Chile | 0.18 | 0.01 | -0.61 | 2.00E-16 | -0.27 | 4.16E-09 |
| Colombia | 0.57 | 2.00E-16 | -0.88 | 2.00E-16 | -0.09 | 0.03 |
| Mexico | 0.47 | 1.29E-11 | -0.51 | 3.55E-14 | -0.26 | 7.52E-08 |
| Peru | -0.03 | 0.78 | -0.80 | 2.27E-11 | -0.35 | 2.43E-05 |
| European anc. | 0.62 | 2.00E-16 | -0.75 | 2.00E-16 | -0.02 | 0.15 |
| Africa anc. | 0.46 | 2.00E-16 | 0.06 | 0.22 | -0.60 | 2.00E-16 |
| EA | 0.04 | 0.04 | 0.05 | 0.01 | -0.02 | 0.18 |
| MI | 0.02 | 1.68E-07 | -0.05 | 2.00E-16 | 0.03 | 2.00E-16 |
| Hair color | 0.10 | 4.84E-03 | -0.17 | 2.86E-06 | 0.05 | 0.05 |
| Eye color | 0.08 | 2.52E-06 | -0.11 | 4.55E-10 | 0.01 | 0.27 |
| Hair shape | -0.06 | 0.01 | 0.01 | 0.66 | 0.15 | 2.00E-16 |

NOTE: Δ refers to self-perception (bands 1 to 5) minus continental ancestry (grouped into 1 to 5 bands at 20 intervals) estimated from the genetic data. MI= Melanin index, EA = Educational Attainment. Sex and Sampling site (here with country names) were incorporated in the analyses as factors while the other variables were treated as quantitative.

Figure 1. Frequency distribution of individual ancestry values in each country.

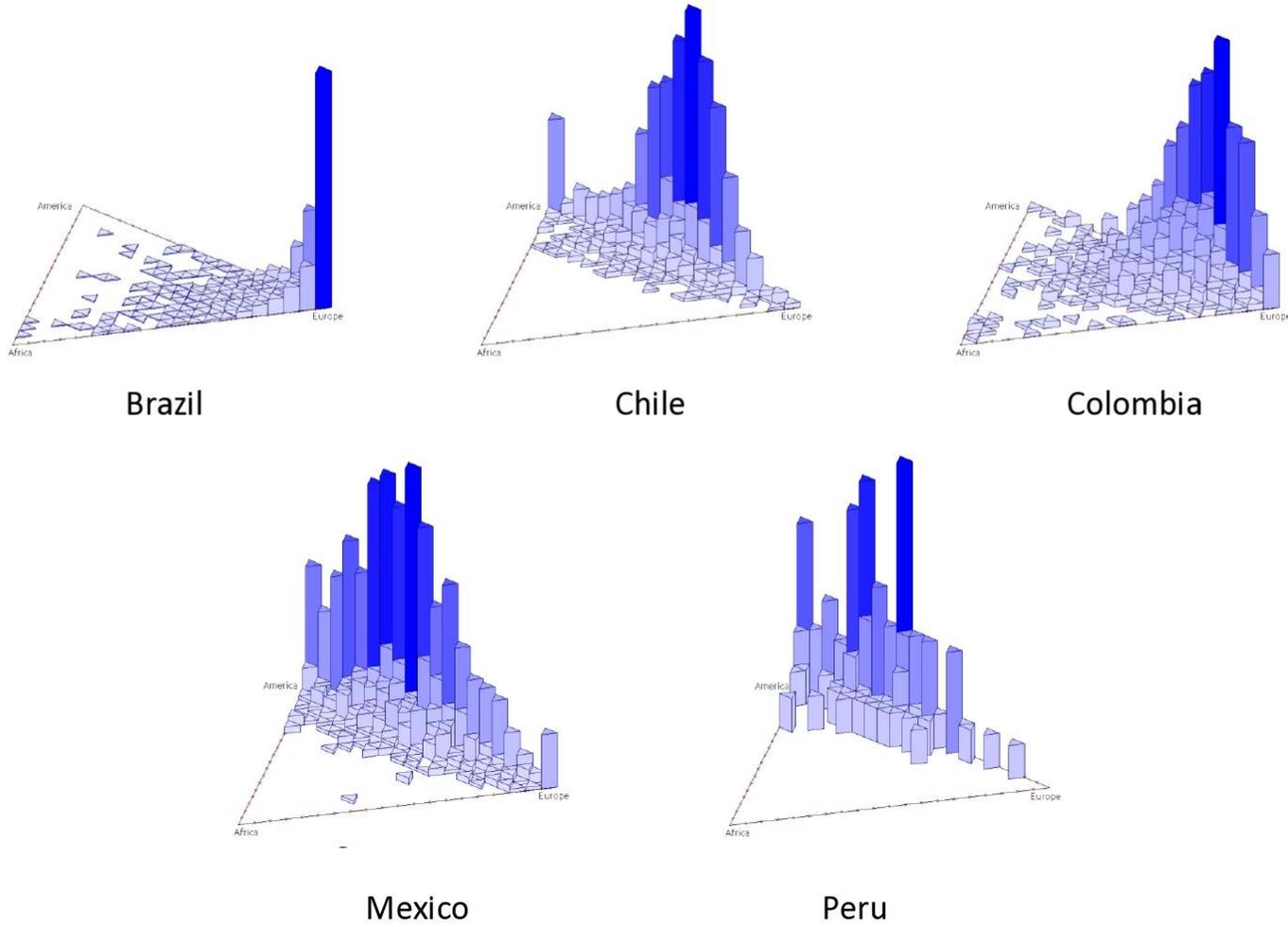


Figure 2. Geographic distribution of individual ancestry values in each country.

Fig 2A

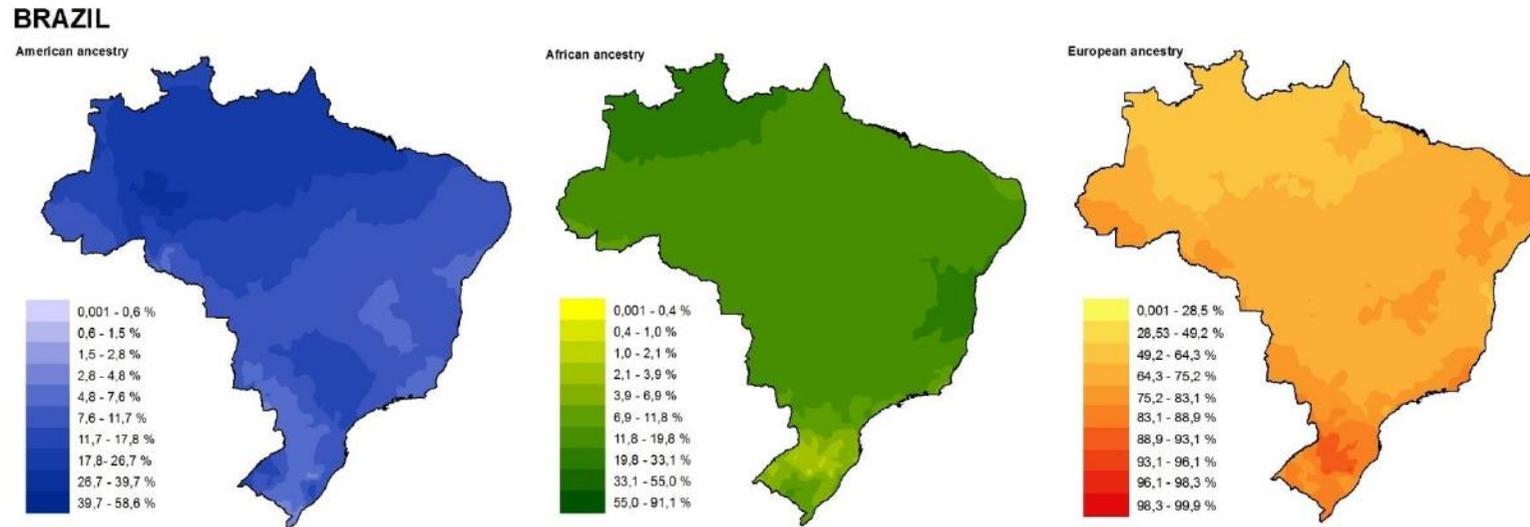


Fig 2B

CHILE

American ancestry



African ancestry



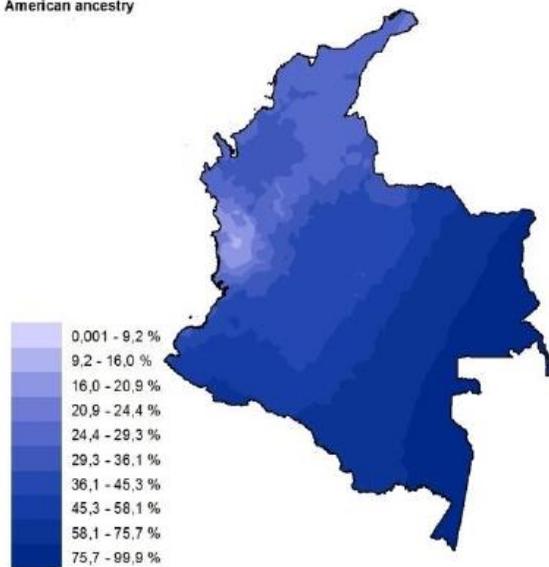
European ancestry



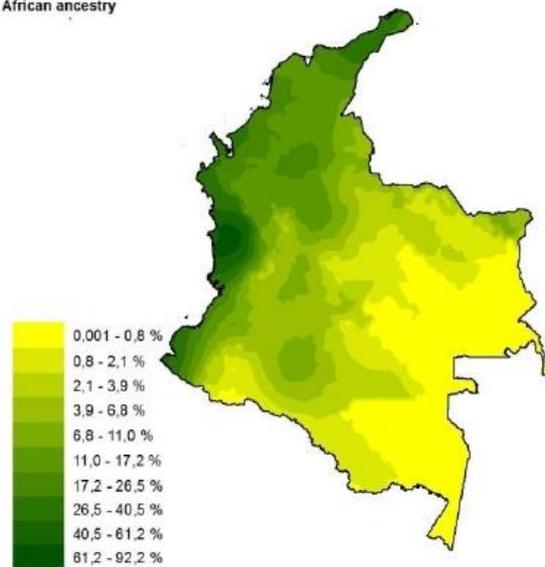
Fig 2C

COLOMBIA

American ancestry



African ancestry



European ancestry

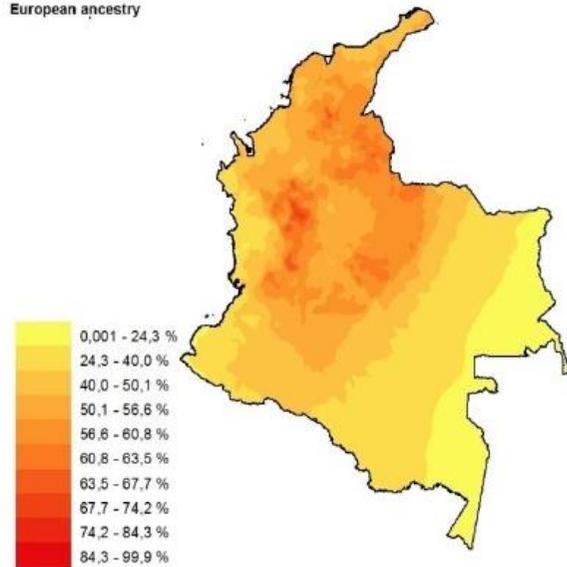
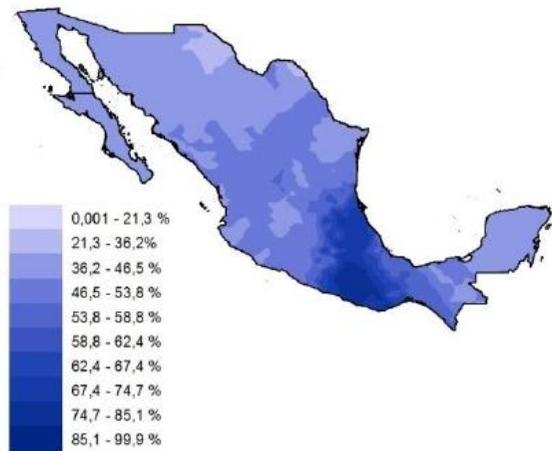


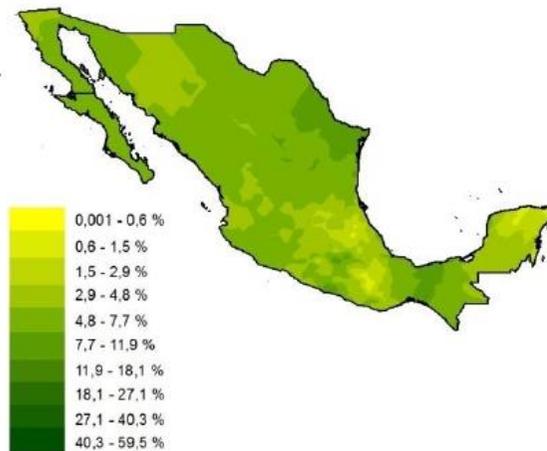
Fig 2D

MEXICO

American ancestry



African ancestry



European ancestry

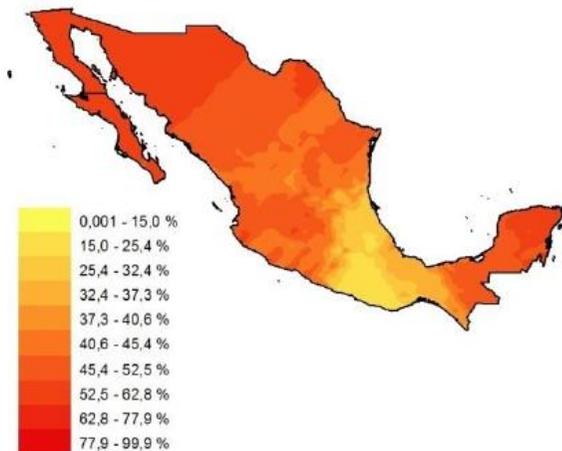


Figure 3. Distribution of eye color, hair color and hair type in each country.

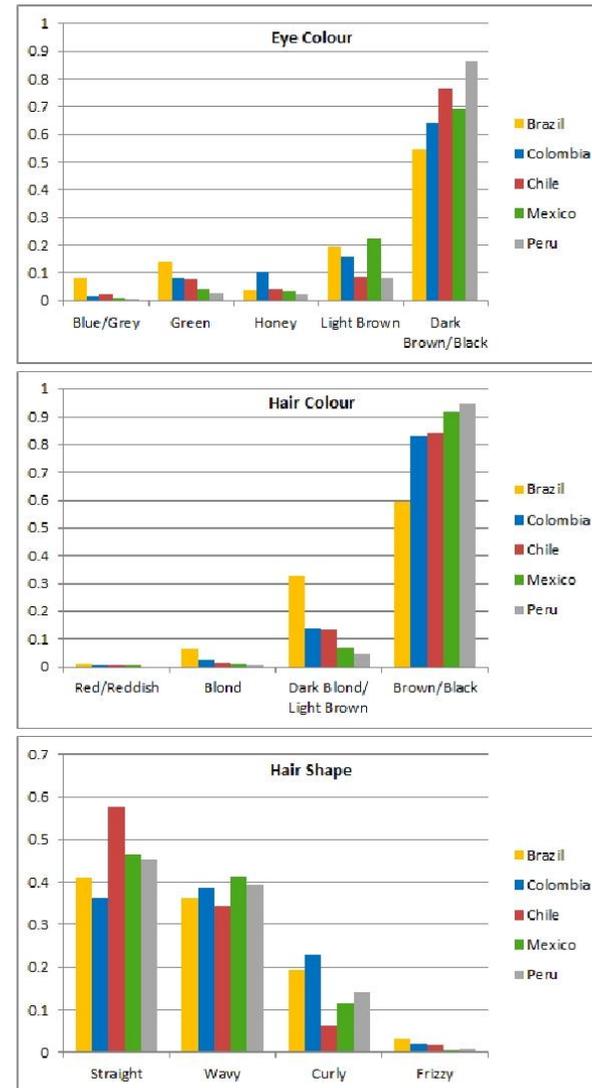


Figure 4. Box plots relating hair shape, hair color and eye color categories to European ancestry in the total sample.

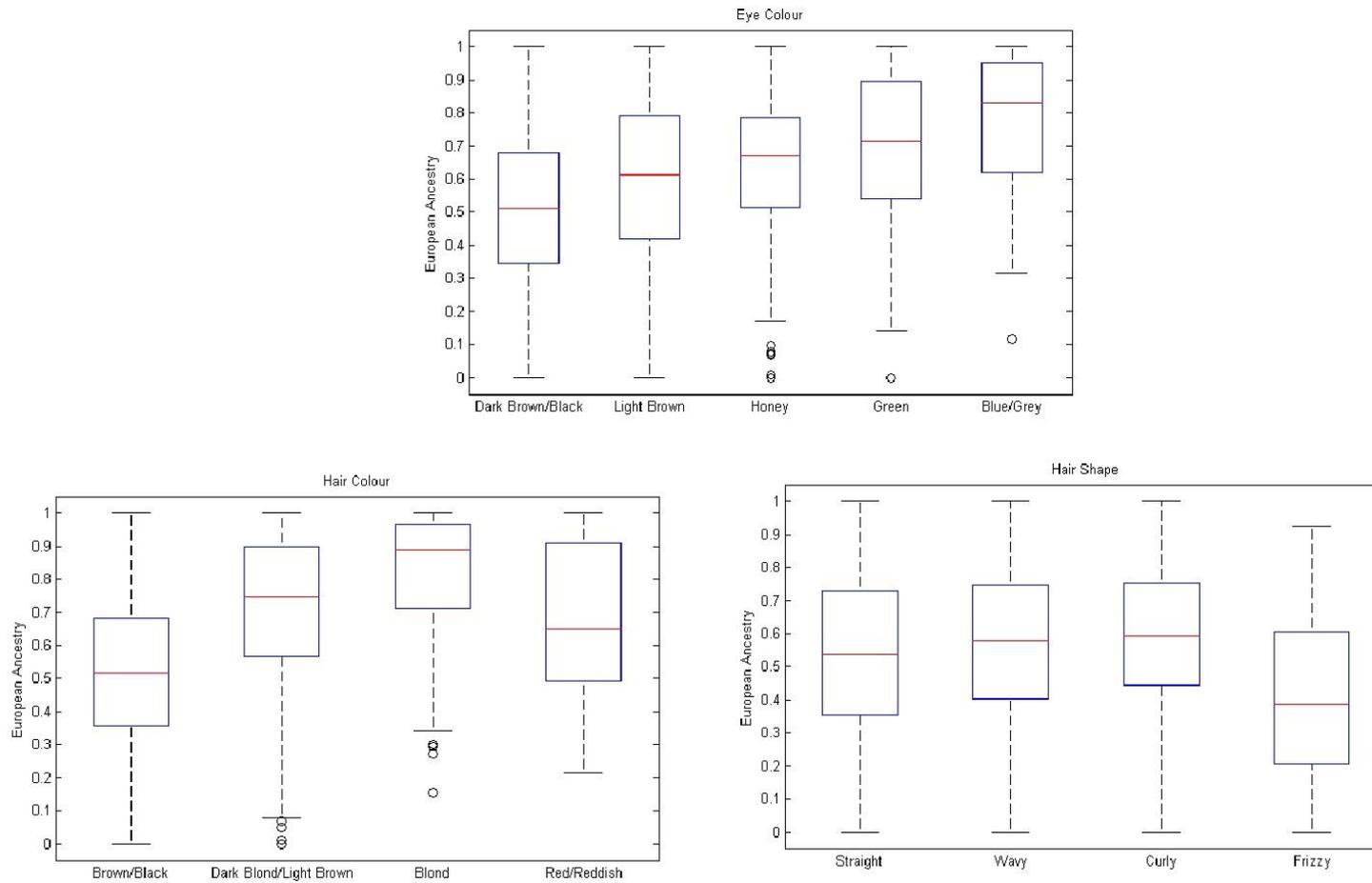


Figure 5. Scatter plots relating (A) skin pigmentation (melanin index MI) (B) height and (C) Shape score, and European ancestry in the total study sample.

Fig 5A

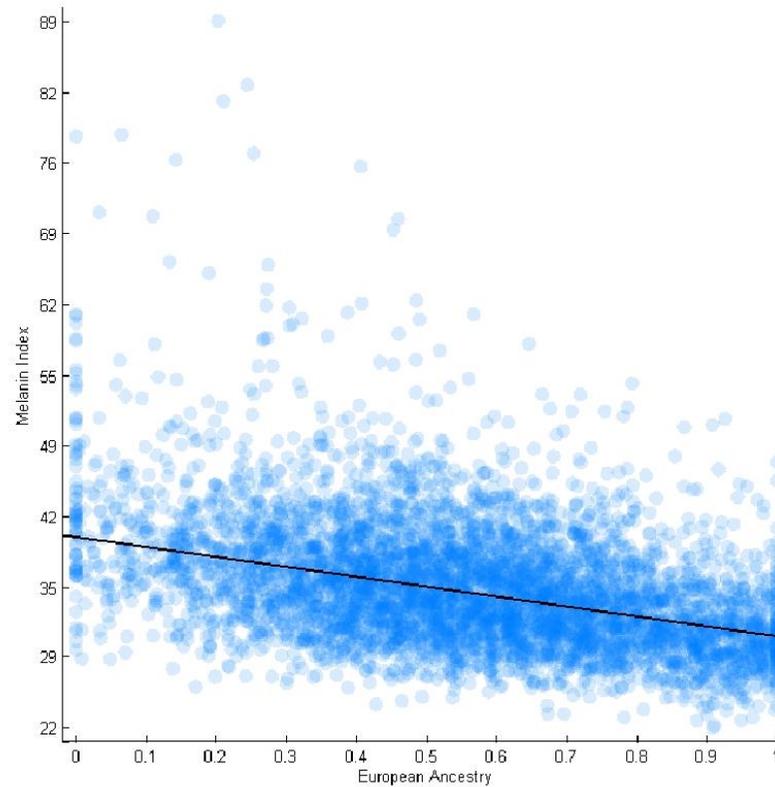


Fig5B

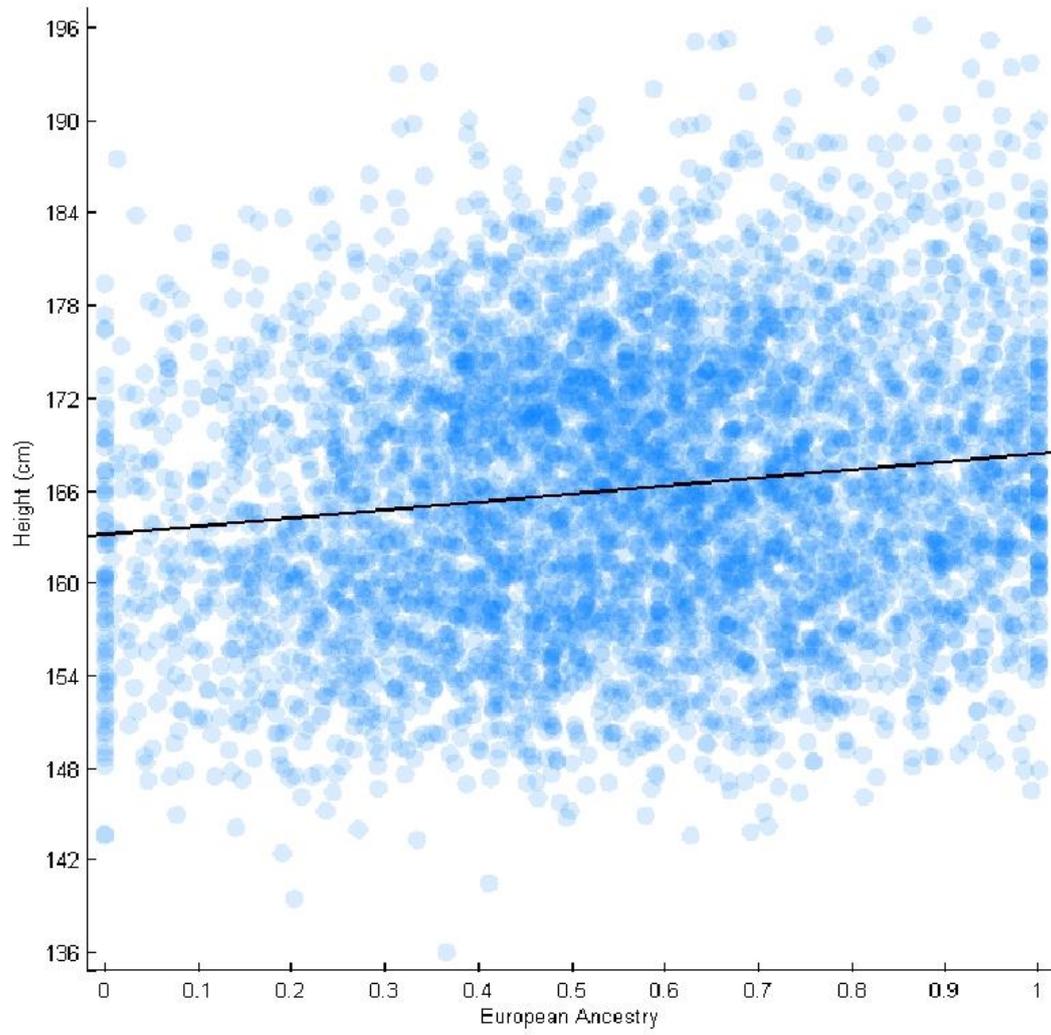


Fig 5C

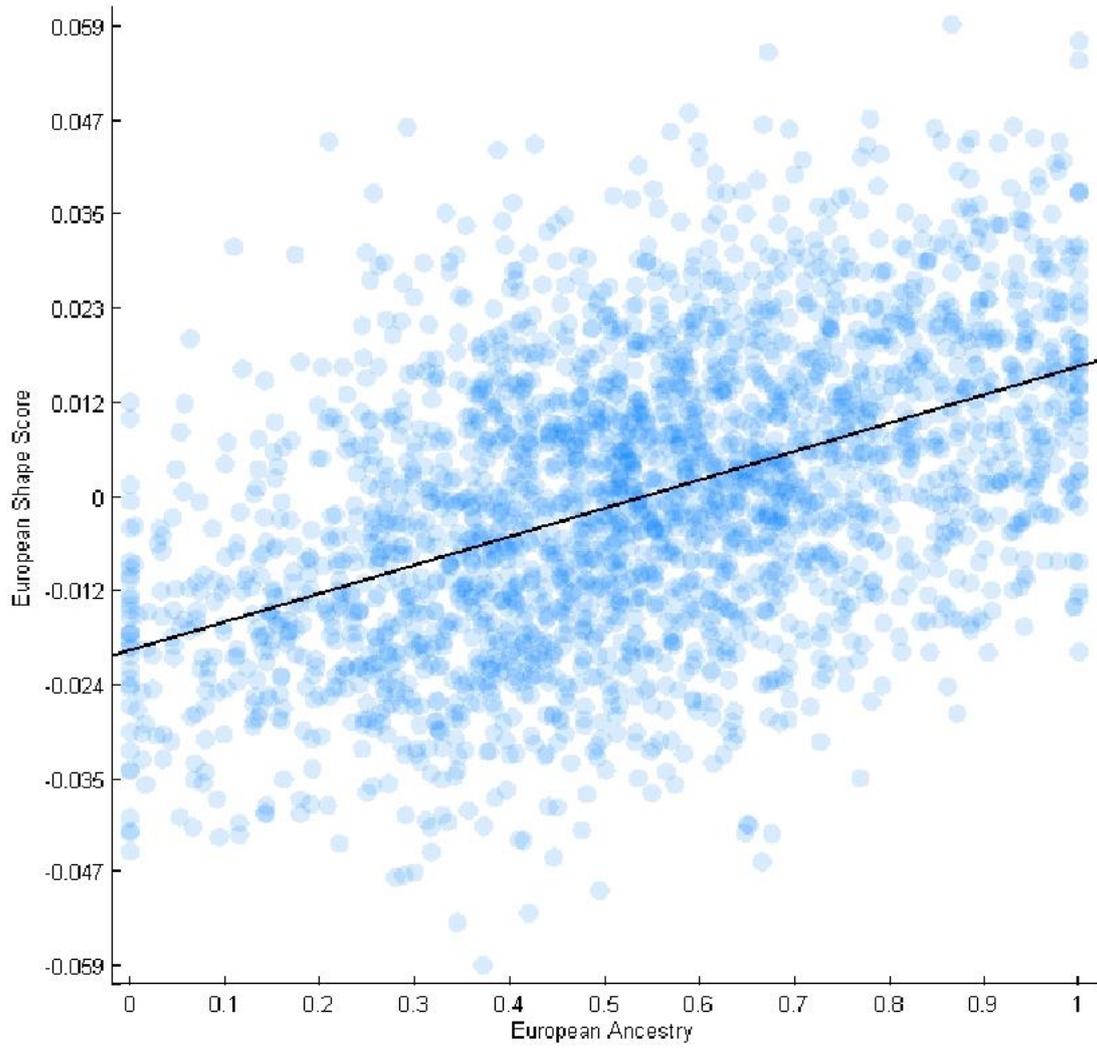
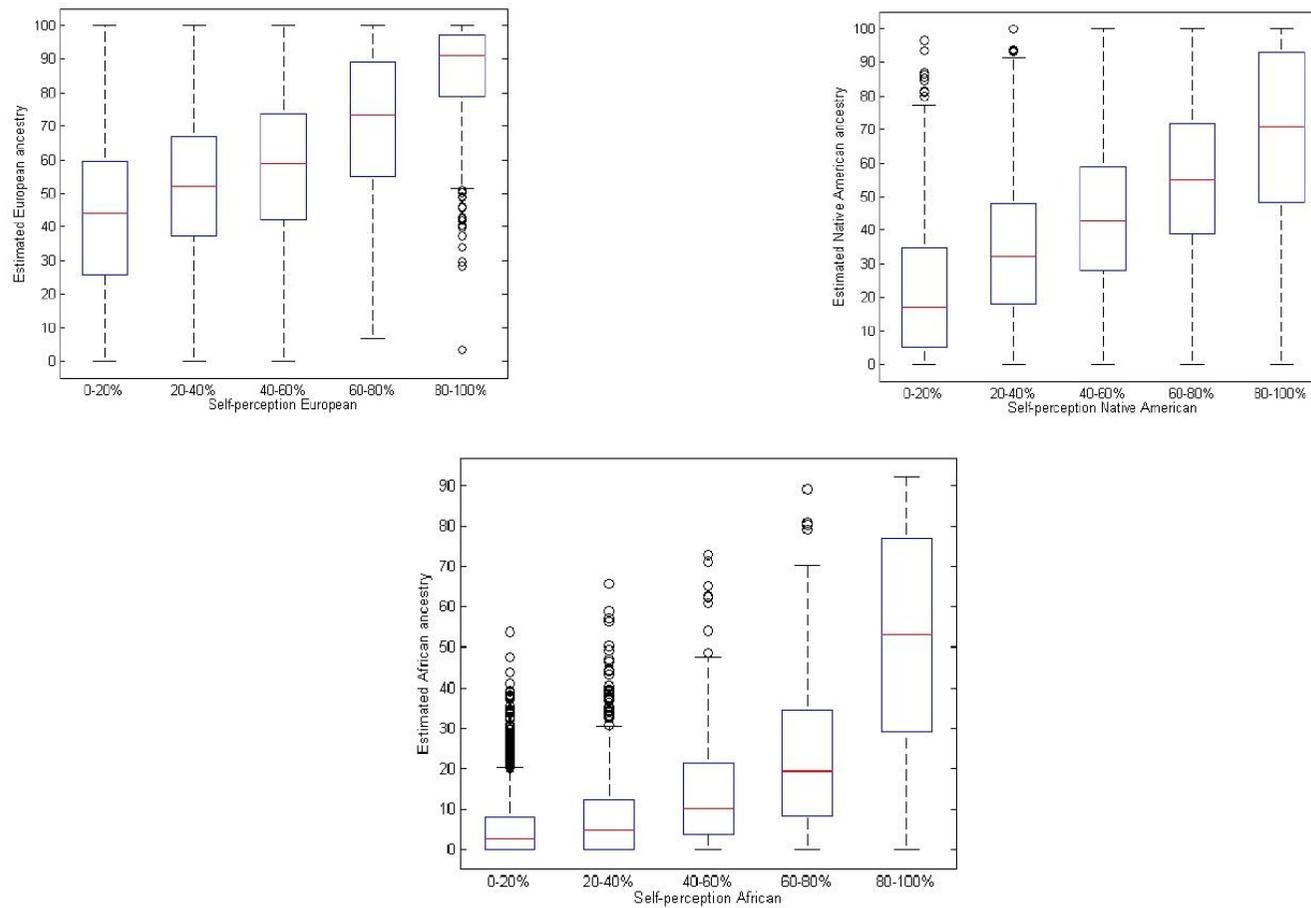


Figure 6. Box plots relating individual genetic ancestry estimates to the bar of self-perceived African, Native American and European ancestry.



Anexo 3 – Artigo produzido em colaboração - Twin Town in South Brazil: A Nazi's Experiment or a Genetic Founder Effect? (Plos One, 2011)

OPEN ACCESS Freely available online



Twin Town in South Brazil: A Nazi's Experiment or a Genetic Founder Effect?

Alice Tagliani-Ribeiro^{1,2}, Mariana Oliveira^{2,3}, Adriana K. Sassi^{2,3}, Maira R. Rodrigues⁴, Marcelo Zagonel-Oliveira^{1,2}, Gary Steinman⁶, Ursula Matte^{2,5}, Nelson J. R. Fagundes^{2,4}, Lavinia Schuler-Faccini^{1,2,5*}

1 Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, **2** National Institute of Science and Technology in Populational Medical Genetics (INAGEMP), Porto Alegre, Brazil, **3** Universidade Federal do Pampa, São Gabriel, Rio Grande do Sul, Brazil, **4** Departamento de Biologia Geral, Instituto de Biociências, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **5** Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil, **6** Department of Biochemistry, Touro College of Osteopathic Medicine, New York, New York, United States of America

Abstract

Cândido Godói (CG) is a small municipality in South Brazil with approximately 6,000 inhabitants. It is known as the "Twins' Town" due to its high rate of twin births. Recently it was claimed that such high frequency of twinning would be connected to experiments performed by the German Nazi doctor Joseph Mengele. It is known, however, that this town was founded by a small number of families and therefore a genetic founder effect may represent an alternatively explanation for the high twinning prevalence in CG. In this study, we tested specific predictions of the "Nazi's experiment" and of the "founder effect" hypotheses. We surveyed a total of 6,262 baptism records from 1959–2008 in CG catholic churches, and identified 91 twin pairs and one triplet. Contrary to the "Nazi's experiment hypothesis", there is no spurt in twinning between the years (1964–1968) when Mengele allegedly was in CG ($P=0.482$). Moreover, there is no temporal trend for a declining rate of twinning since the 1960s ($P=0.351$), and no difference in twinning among CG districts considering two different periods: 1927–1958 and 1959–2008 ($P=0.638$). On the other hand, the "founder effect hypothesis" is supported by an isonymy analysis that shows that women who gave birth to twins have a higher inbreeding coefficient when compared to women who never had twins (0.0148, 0.0081, respectively, $P=0.019$). In summary, our results show no evidence for the "Nazi's experiment hypothesis" and strongly suggest that the "founder effect hypothesis" is a much more likely alternative for explaining the high prevalence of twinning in CG. If this hypothesis is correct, then this community represents a valuable population where genetic factors linked to twinning may be identified.

Citation: Tagliani-Ribeiro A, Oliveira M, Sassi AK, Rodrigues MR, Zagonel-Oliveira M, et al. (2011) Twin Town in South Brazil: A Nazi's Experiment or a Genetic Founder Effect? PLoS ONE 6(6): e20328. doi:10.1371/journal.pone.0020328

Editor: Henry Harpending, University of Utah, United States of America

Received: January 22, 2011; **Accepted:** April 25, 2011; **Published:** June 8, 2011

Copyright: © 2011 Tagliani-Ribeiro et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This project was funded by the National Research Council (Brazil), grant number 573993/2008-4. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lavinia.faccini@ufrgs.br

Introduction

The etiology of twin births in humans is still largely unclear and is the subject of a number of investigations [1–4]. Traditionally, twins are classified as monozygotic (MZ) and dizygotic (DZ). MZ twins are developed when an embryo splits soon after fertilization. DZ twins occur when two separate oocytes, released during the same menstrual period, are fertilized by separate sperm cells [1]. MZ is rarer than DZ twinning except in Japan [5] and no maternal, genetic, or environmental factors have been associated to it so far [6]. On the other hand, DZ twinning is a phenomenon of complex etiology, in which both genetic predisposition and environmental factors play a role [7]. Although familial aggregation of DZ twins has been known for a long time, few genes have been associated with DZ twinning, including the receptor of FSH hormone (*FSHR*), growth differentiation factor 9 (*GDF9*), methylenetetrahydrofolate reductase (*MTHFR*), etc. [8–10]. Non-genetic factors such as advanced maternal age, increased parity, lactation, diet, higher maternal height body mass index and race are observed also as risk factors for DZ twinning too [11].

Twinning rates shows a wide geographical and temporal variation, being extremely rare in Asian populations (5–6 in

1,000 maternities) and more frequent in Sub-Saharan populations (23 in 1,000 maternities) [11]. Previous studies show that twinning rates in European countries started to decline around 1900, but have increased steadily from the 1970s onwards [12], possibly as a result of both increased maternal age and a more widespread use of assisted reproductive technology (ART) procedures [11,13,14]. Even in a single continent such as Europe variations on twinning rates between different countries are observed [15]. The variability on twinning rates has been assigned to the variation in DZ twinning rates, as MZ twinning has a constant prevalence around the world and time (4 in 1,000 maternities) [16].

Cândido Godói (CG) is a small town in South Brazil (27°57'07"S; 54°45'07"W) with approximately 6,000 inhabitants. It is known as the "Twins' Town" due to its high rate of twin births. According to the Brazilian Ministry of Health, between 1994 and 2006, around 2% of the live births in CG were twins, compared to an average of 1% for the whole country [17]. However, twinning may not be equally distributed throughout the municipality. In 1994, the twinning birth rate in Linha São Pedro (LSP), a small district of CG was estimated as 10% [18]. CG and LSP were both founded at the beginning of the Twentieth Century by a few families of German ancestry coming from other German-

founded towns in Rio Grande do Sul. Presently, the population of LSP is less than 600 inhabitants and most are catholic [19]. The reasons for the higher twinning rate in CG in general, and in LSP in particular, however, are still unclear.

Recently, a controversial theory was raised by an Argentinean journalist [20] who wrote a book alleging a possible link between the twinning phenomenon in CG and supposed experiments of the Nazi physician, Joseph Mengele. According to Camarasa, Mengele could have lived and worked as a physician in CG in the beginning of the 1960s, after living in Buenos Aires. Even though Camarasa's suppositions were not based on any actual historical records available [21], his story has caught attention of the international press, which created a worldwide "fuzz" around the "Brazilian Twin Town" in many countries such as the United Kingdom (<http://www.dailymail.co.uk/news/worldnews/article-1126504/The-Twins-Brazil-Did-Nazi-doctor-Mengele-Angel-Death-cause-twin-surge-South-American-town.html>), Brazil (<http://revistaepoca.globo.com/Revista/Epoca/0,EMI24803-15228,00-NAZISTA+JOSEF+MENGELE+CRIOU+CIDADE+DOS+GEMEOS+NO+RIO+GRANDE+DO+SUL+DIZ+LIVR.html>) and USA (<http://news.nationalgeographic.com/news/2009/11/091125-nazi-twins-brazil-mengele.html>).

In this study, we surveyed baptism records in CG and LSP to evaluate temporal trends to test the predictions of the "Nazi's experiment hypothesis". More specifically, we asked the following questions: 1. Is there any temporal and geographical variation in twinning in CG (and LSP)? 2. Is there any increase of the twinning rate around the 60 s or late 60 s, when Joseph Mengele was supposedly working as a physician there? We also searched for alternative explanations for the higher twinning rate in CG. Given the known history of CG, we suspected that a genetic founder effect may be involved in the increased twinning rate in CG (and LSP). To test this hypothesis, we asked, using isonymy methods, whether twin's mothers have a higher inbreeding coefficient as compared to women who never gave birth to twins. Our results clearly show that contrary to the "Nazi's experiment hypothesis" there is no peak on the twinning rate around the 60 s. On the other hand, the isonymy analysis supports the hypothesis that a founder effect is a much more likely explanation for the higher prevalence of twinning in LSP and in CG as a whole.

Materials and Methods

Ethics Statement

This study used only secondary public data for analysis, which was treated anonymously. This research project was approved by the Hospital de Clinicas de Porto Alegre Ethics Committee under the protocol number 09-359. Written informed consent was obtained for all participants before interviews were conducted.

Twinning patterns within CG

Live births were surveyed from baptism records available in CG Catholic churches. These records include the child first name, family names of both parents, sex, and the locality (district) where the family lives. The earliest available baptism records date from 1927, but complete and reliable data were available only from 1959 onwards. Since there is a suggestion that LSP may have an especially high twinning frequency [18], we used a chi-square test to assess whether LSP has a higher twinning rate as compared to the remaining CG districts (CG-LSP) considering 1959 to 2008. Because maternal age (MA) is a known factor affecting twinning [11] we compared MA distribution in LSP and CG-LSP using data obtained from interviews to test if women living in LSP have children later in life, thus increasing the chance of twin births. Mean maternal age between these groups was compared using a Student *t*-test adjusted for unequal variances.

Temporal tendencies in LSP and CG

According to Camarasa [20], Mengele would have arrived in CG around 1963 and visited the city until 1968, but the exact timing for his stay is uncertain. To evaluate the hypothesis that Mengele's stay in CG would have increased the twinning rate in the city, we used a chi-square test to compare the occurrence of twin births between the period of 1964–1968 and the remaining years.

To have a more detailed view on the temporal variations of twinning in CG, we used a G-test for trends to ask if there is any temporal trend over the frequency of twin births occurring after 1959, grouped by intervals of five years. We tested for temporal trends in twin births for CG as a whole, for LSP only, and for CG-LSP. The "Nazi's experiment hypothesis" predicts a peak of

Table 1. Distribution of twin baptism records in Cândido Godói from 1959 to 2008.

| District | Total baptisms | | Twin records | | Prevalence of twinning* |
|--|----------------|------------|--------------|------------|-------------------------|
| | N | % | N | % | % |
| Centro | 2,548 | 40.7 | 13 | 14.1 | 0.5 |
| Linha São Miguel | 512 | 8.2 | 6 | 6.5 | 1.2 |
| Linha São Pedro | 469 | 7.5 | 33 | 35.9 | 7.0 |
| Sede Boa Vista | 588 | 9.4 | 9 | 9.8 | 1.5 |
| Linha São Bonifácio | 398 | 6.4 | 1 | 1.1 | 0.3 |
| Linha Secção A | 406 | 6.5 | 5 | 5.4 | 1.2 |
| Linha Acre | 293 | 4.7 | 6 | 6.5 | 2.0 |
| Linha Paranaguá | 196 | 3.1 | 2 | 2.2 | 1.0 |
| Linha Natal | 338 | 5.4 | 5 | 5.4 | 1.5 |
| Linha São João, La Salle and Esquina União | 168 | 2.6 | 6 | 6.5 | 3.5 |
| Linha Godói Centro | 217 | 3.5 | 2 | 2.2 | 0.9 |
| Linha Silva Jardim | 129 | 2.1 | 4 | 4.3 | 3.1 |
| TOTAL | 6,262 | 100 | 92 | 100 | 1.5 |

*Calculated as twin records/total baptism records.
doi:10.1371/journal.pone.0020328.t001

twinning around the late 1960s followed by either a constant or abrupt decline on the twinning rate towards the average level in the Brazilian population, since the increase in twinning obtained by Mengele would be effective only on the short-term.

Finally we used a chi-square test to evaluate if the relative frequency of twinning between LSP and CG-LSP was held constant when two periods are considered: 1927–1958 and 1959–2008. Because the earliest records (up to 1959) are only reliable for

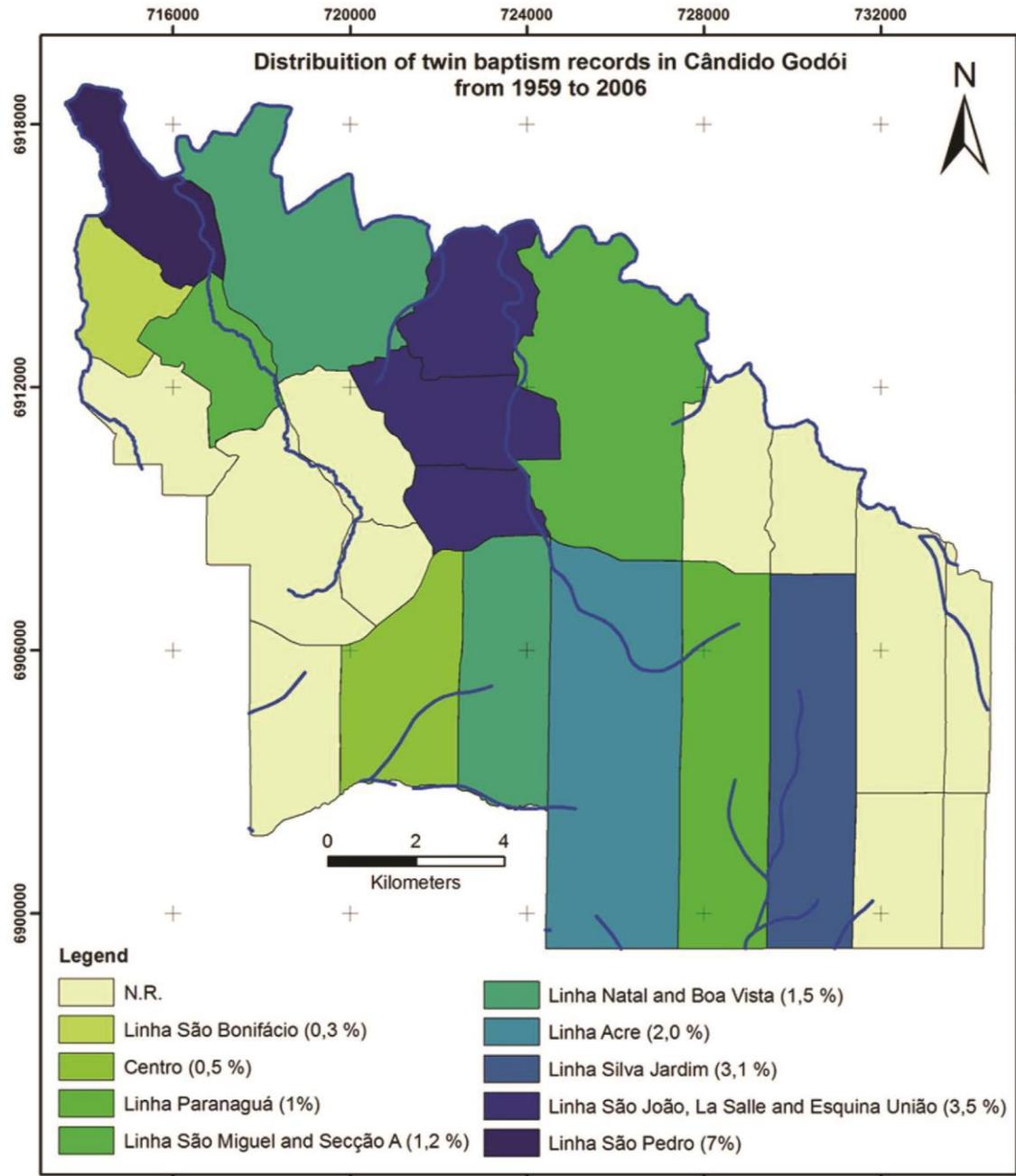


Figure 1. Districts of Cândido Godói with their twin births registration.
doi:10.1371/journal.pone.0020328.g001

live born twins – but not for overall births only the relative twinning frequency in different CG districts could be compared. All statistical tests, odds ratios, and their 95% confidence interval were computed using BioEstat 5.0 [22].

MZ/DZ ratio

The proportion of the MZ twinning rate (MZ_f) over the DZ twinning rate (DZ_f) was estimated using Weinberg’s differential method [23] in which DZ_f can be estimated by doubling the number of opposite sex twin pairs (OS) and dividing by the total number of maternities (N): $DZ_f = 2OS/N$, while MZ_f can be estimated by subtracting the number of OS twins from the number of same sex (SS) twins and dividing by N : $MZ_f = (SS - OS)/N$. Because we only have the total number of maternities for the period 1959–2008, this was the period considered for MZ_f and DZ_f estimation. The corrected variances for MZ_f and DZ_f ($Var(MZ_f)$ and $Var(DZ_f)$, respectively) was estimated as in [24]. We compared MZ_f and DZ_f between LSP and CG-LSP using a binomial test. Finally, we used the chi-square distribution to test if MZ_f and DZ_f are equally increased in LSP compared to CG-LSP using the absolute number of MZ and DZ twin pairs estimated by the numerator of Weinberg formula. As the total number of maternities is not necessary for this estimation we used the full birth records from 1927–2008 to increase statistical power.

Pedigree and isonymy

To have a more detailed picture of the pedigree structure for families where a high twinning rate is reported compared to families without familial history of twinning, we interviewed 42 women from independent households who gave birth to twins (cases) and 101 women from independent households who only had single births (controls). All 143 women were residents from CG. The aim of this questionnaire was to collect further data that would provide a better description of these women and to access other possible risk factors for twinning. More specifically, we asked familiar data, including surnames from both parents and history of twinning in their family, and questions about their lifestyle. For each household, we used the software Progeny 7.0® to draw pedigrees. We then used the surname data to calculate isonymy, a quantity that makes use of the frequency of surnames in a given population to measure deviations from panmixia [25]. In this sense, isonymy is related to the inbreeding coefficient F [26], which was estimated as in [27]. Analyses of population structure by isonymy methods have been carried out in many human populations [28–32]. The “founder effect hypothesis” predicts that the higher twinning rate in CG reflects, at least in part, a genetic founder effect which happened when CG was settled by few families of German descent in the beginning of the 20th century. Thus, if such founder effect affects the twinning rate in CG we

Table 2. Twin and single baptism records in Linha São Pedro (LSP) and in others Cândido Godói’s (CG-LSP) districts, from 1959–2008.

| | Twin births | Single births |
|------------|-------------|---------------|
| LSP | 33 (35.9%) | 436 (7.1%) |
| CG – LSP | 59 (64.1%) | 5,735 (92.9%) |
| CG (total) | 92 (100.0%) | 6,170(100.0%) |

Relative Risk for twinning in LSP = 7.3 (CI 95%: 4.75–11.38).
Chi-square test $P < 0.0001$.
doi:10.1371/journal.pone.0020328.t002

Table 3. Twin baptism records between 1964–1968 and in the remainders years in Linha Sao Pedro (LSP) and in others Cândido Godói’s (CG-LSP) districts.

| CG | Twins | No-twins | P value |
|--------------|-------|----------|----------|
| 1964–1968 | 9 | 778 | 0.4815 |
| Others years | 80 | 5395 | |
| CG-LSP | Twins | No-twins | |
| 1964–1968 | 4 | 718 | 0.2940* |
| Others years | 53 | 5018 | |
| LSP | Twins | No-twins | |
| 1964–1968 | 5 | 60 | 0.7724** |
| Others years | 27 | 377 | |

*Pyates.
**Pwilliams.
doi:10.1371/journal.pone.0020328.t003

expect a higher inbreeding coefficient for cases. We calculated F for cases (F_{cases}) and controls ($F_{controls}$), and the statistical significance of this difference was assessed by 1,000 permutations. The reported P -value thus represents the proportion of permutations resulting in a ratio $F_{cases}/F_{controls}$ equal or higher than the observed.

Results

From 1959 to 2008 there were a total of 6,262 baptisms in 14 CG districts, including 91 pairs of twins and one triplet. Table 1 presents the geographical distribution of baptism records in CG according to the district where the parents lived. Using total baptism records we estimated the frequency of twinning in these districts. LSP shows a frequency of live born twins of 7.0% compared to 1.5% in CG as a whole. No other district presented frequencies above 3.5% (Table 1, Figure 1). It is noteworthy that while LSP contributes only 7.5% to all baptism records, it contributes almost 1/3 (33/92) to all twin births registered in CG, representing an odds ratio of 7.3 (CI 95% 4.75–11.38) for LSP compared to the remaining districts in CG ($P < 0.0001$) (Table 2). There was no difference in maternal age between LSP and CG-LSP ($P = 0.3036$; $MA_{LSP} = 26.10 \pm 4.71$ years; $MA_{CG-LSP} = 26.87 \pm 5.88$ years (mean \pm SD), suggesting that differences in maternal age is unlikely to explain the higher twinning rate in LSP.

There is no increase on the twinning rate in CG between the period 1964–1968 and the remaining years (Table 3, $P = 0.482$). These results are consistent when only LSP or CG-LSP are analyzed ($P = 0.772$, $P = 0.294$, respectively). There is also no temporal variation on the relative frequency of twinning between LSP and CG-LSP, considering the periods of 1927–1958 and 1959–2008 (Table 4, $P = 0.638$), suggesting that LSP has been

Table 4. Twin baptism records before and after 1958 in Linha Sao Pedro (LSP) and in others Cândido Godói’s (CG-LSP) districts.

| | 1927–1958 | 1959–2008 |
|------------|-------------|---------------|
| LSP | 11 (31.4%) | 33 (35.9%) |
| CG – LSP | 24 (68.6%) | 59 (64.1%) |
| CG (total) | 92 (100.0%) | 6,170(100.0%) |

Qui-square test: $P = 0.638$.
doi:10.1371/journal.pone.0020328.t004

Table 5. Temporal distribution of twin baptism records from 1959 to 2008, in Linha São Pedro (LSP), in other districts of Cândido Godói (CG-LSP) and Cândido Godói as a whole (CG).

| Period | LSP | | CG-LSP | | CG | |
|-----------|---------------------|------|---------------------|-----|---------------------|-----|
| | Twins/single births | % | Twins/single births | % | Twins/single births | % |
| 1959–1963 | 4/84 | 4.8 | 12/965 | 1.2 | 16/1049 | 1.5 |
| 1964–1968 | 5/60 | 8.3 | 4/718 | 0.6 | 9/778 | 1.2 |
| 1969–1973 | 0/67 | 0.0 | 8/749 | 1.1 | 8/816 | 1.0 |
| 1974–1978 | 3/54 | 5.6 | 10/641 | 1.6 | 13/695 | 1.9 |
| 1979–1983 | 3/53 | 5.7 | 10/723 | 1.4 | 13/776 | 1.7 |
| 1984–1988 | 4/34 | 11.8 | 4/596 | 0.7 | 8/630 | 1.3 |
| 1989–1993 | 7/36 | 19.4 | 4/471 | 0.8 | 11/507 | 2.2 |
| 1994–1998 | 0/30 | 0 | 4/390 | 1.0 | 4/420 | 1.0 |
| 1999–2003 | 3/9 | 33.0 | 1/257 | 0.4 | 4/266 | 1.5 |
| 2004–2008 | 4/9 | 44.4 | 2/224 | 0.9 | 6/233 | 2.6 |

doi:10.1371/journal.pone.0020328.t005

historically the major contributor for the overall higher twinning rate in CG. The test for temporal trends in the frequency of twin births reveals an interesting pattern (Table 5). There is no evidence

for any temporal trend in either CG as a whole (Figure 2A; $P=0.351$) or CG-LSP (Figure 2B; $P=0.486$). In contrast, for LSP alone there is a tendency for an increase on the twinning rate

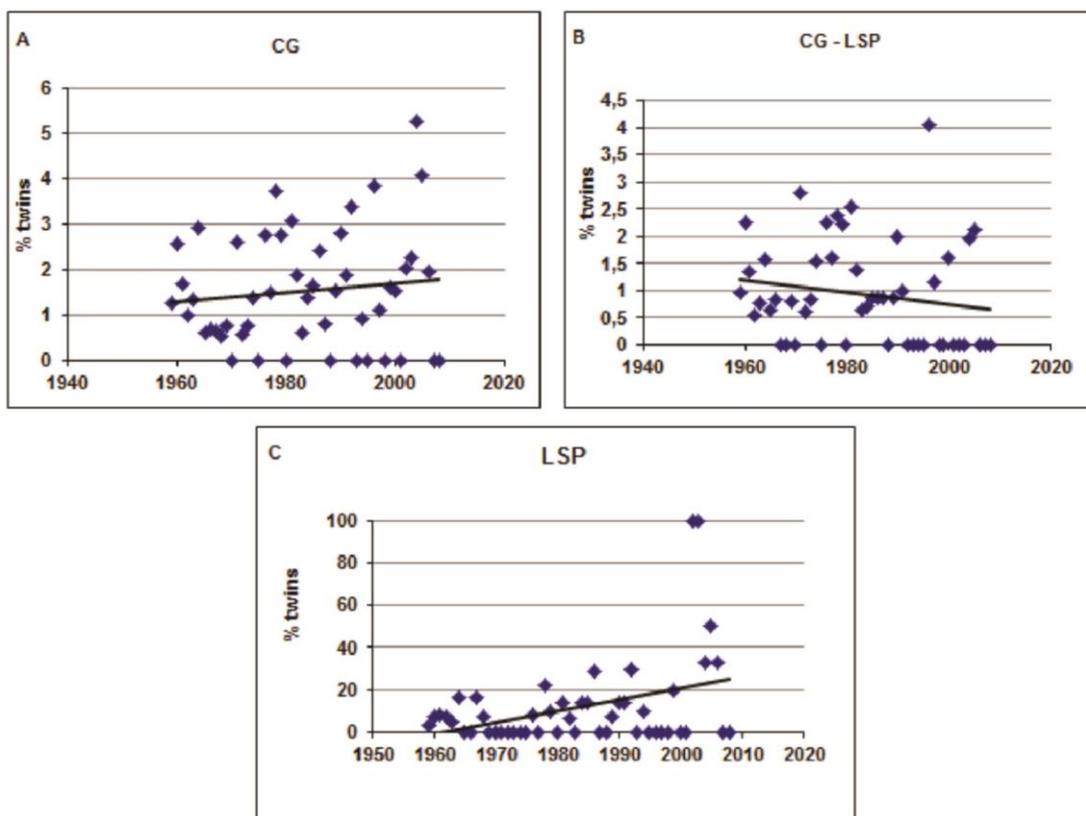


Figure 2. Temporal twinning tendencies. 2a: Temporal twinning tendencies in Cândido Godói; 2b: Temporal twinning tendencies in Cândido Godói, excluded Linha Sao Pedro; 2c: Temporal twinning tendencies in Linha Sao Pedro only.
doi:10.1371/journal.pone.0020328.g002

Table 6. Estimation of DZ twinning rate (DZr), MZ twinning rate (MZr) and the absolute number of DZ and MZ twin pairs in LSP and CG-LSP.

| | LSP | | | | CG-LSP | | | |
|---------------------|-----------|----|-----------|----|-----------|----|-----------|----|
| | 1927–1958 | | 1959–2008 | | 1927–1958 | | 1959–2008 | |
| | SS | OS | SS | OS | SS | OS | SS | OS |
| | 4 | 7 | 9 | 23 | 16 | 7 | 14 | 51 |
| Total births | Unknown | | 469 | | Unknown | | 5793 | |
| DZr* | 3.8380% | | | | 0.4833% | | | |
| Var(DZr)* | 0.0161% | | | | 0.0002% | | | |
| MZr* | 2.9851% | | | | 0.6387% | | | |
| Var(MZr)* | 0.0144% | | | | 0.0002% | | | |

SS, same sex twin pairs; OS, opposite sex twin pairs.

*estimated using data from 1959–2008 only.

doi:10.1371/journal.pone.0020328.t006

across time (Figure 2C; $P=0.001$). Importantly, this trend goes in the opposite direction of that expected according to the “Nazi’s experiment hypothesis”, whose predictions are also rejected in all other tests we performed.

Table 6 shows for both LSP and CG-LSP the numbers of SS and OS for the period between 1927–1958, and SS, OS and total births for the period between 1959–2008. For this last period, DZr equals 3,84% for LSP but only 0,48% in CG-LSP, while MZr equals 2,98% in LSP and only 0,64% in CG-LSP. This difference is highly significant for both DZr and MZr ($P<0,0001$ for both comparisons). To test if MZr and DZr are increased by the same amount in LSP we used the numerator of Weinberg formula to estimate that 32 DZ twin pairs and 11 MZ twin pairs were born in LSP between 1927–2008, while during the same period 42 DZ twin pairs and 46 MZ twin pairs were born in CG-LSP. This difference is significant ($P=0.004$). Thus, even though both DZr and MZr are increased in LSP, such increase seems to affect DZ twin births more strongly than MZ twin births. Since genetic factors have been only associated with DZ twinning, this result may be taken as a suggestion that genetic factors are responsible for the increased twinning rate seen in CG (for which LSP is the major contributor). Such notion is reinforced from the pedigrees resulting from the reported familiar history, which allows us to observe that twin birth is recurrent for several families living in CG

Table 7. Inbreeding coefficient by Isonomy (F) from cases and controls.

| Sample | Number of Individuals | Number of Surnames | F |
|----------|-----------------------|--------------------|--------|
| Cases | 42 | 85 | 0.0148 |
| Controls | 101 | 196 | 0.0081 |

$P=0.019$.

doi:10.1371/journal.pone.0020328.t007

(Figure 3). Twenty-eight out of 42 mothers of twins had female sisters, and in five out of these 28 kindreds (17,8%), recurrence of twin births was observed. The total number of sisters of index mothers was 85, five of them having twins, which means an overall recurrence chance of having twins in females by 5.9%.

Finally, by analyzing the distribution of surnames between cases and controls, we found that the estimated F for cases is statistically higher than the estimated F for controls ($P=0.019$; Table 7), suggesting that, as predicted by the founder effect hypothesis, women having twins in CG are more genetically related than women having single births.

Discussion

Our results provide a robust refutation to the hypothesis that the high twinning rate in CG is due to Nazi’s experiments in the 1960s, and strongly suggest that a genetic founder effect occurred during the settlement of CG is a much better alternative to explain this phenomenon. On other hand, we confirmed that the high prevalence of twins in CG is particularly concentrated in LSP and that this is a historical trend. Importantly, this increase on the twinning rate at LSP is not due to fertility treatments, because these techniques have been extremely rare in Brazil due to its high costs, especially in rural areas like CG.

One of the limitations of this study is that it is based on baptism registries and not on official birth certificates or hospital births records. However, birth certificates in Brazil provide reliable information for twinning only after 1990, and old hospital records were not available in CG. Moreover, in small communities such as CG was quite common in the first decades of the 20th century for women to have babies at home or to attend to hospitals in the neighbor cities. Baptism records, on the other hand, are carefully kept in the different churches located in the municipality, and these

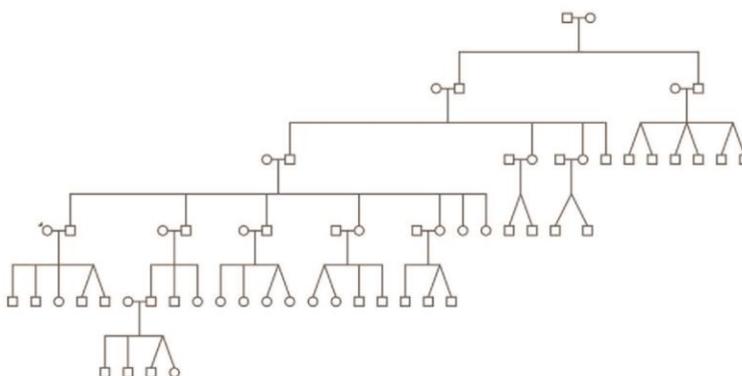


Figure 3. Illustrative pedigree of one family living in Linha São Pedro.

doi:10.1371/journal.pone.0020328.g003

records include the name and sex of the children, date of birth, parent's name and locality where parents were living. The fact that only catholic churches were included may also be a limitation, but the last demographic census in Brazil in 2007 showed that 76.6% of CG population is Catholic. Baptism records may suffer from another bias, in that only live births are registered. However, this would underestimate the actual twinning prevalence in CG, and since the overall pattern of twinning is consistent between LSP and CG-LSP considering the periods of 1927–1958 and 1959–2008, it is unlikely that this would bring a significant bias to our data.

Although the importance of environmental factors cannot be formally ruled out, some factors which are important for twinning etiology such as the use of oral contraceptives, or acid folic supplementation [33–35] are unlikely to explain the high twinning rates already displayed by this population during the 1960s, when these substances were not available in the Brazilian market. On the other hand, high levels of IGF in milk have been implicated in elevated twinning rates [4]. Since all CG districts are based in rural economy and most of the families consume dairy products, dietary factors may represent a possible adjuvant for the high prevalence of twinning. However, because all CG districts have a similar lifestyle, milk consumption itself hardly explains the difference in twinning rates between LSP and CG-LSP. Similarly, differences in maternal age between LSP and CG-LSP are also unlikely to explain the observed pattern, since there is no statistical difference in maternal age between LSP and CG-LSP.

When one analyses extreme twinning rates in small isolates one has to keep in mind that the random fluctuations are also extreme. In Europe, a phenomenon similar to Cândido Godói has been registered in the Åland Islands, where higher twinning rates were historically recorded compared to mainland areas of Finland and Sweden [36]. The colonization history of CG is suggestive that a genetic founder effect may have played a role in this process, leading to high shifts in allele frequency between the ancestral and the derived population. Because it is a random process, founder effects have unpredictable phenotypic effects, which, in this specific case, may be a high frequency of twinning. Our results provide strong evidence that corroborates this idea. Firstly because of the familiar aggregation of twin births and the high prevalence of DZ twins in LSP which are both indicative that genetic factors are involved in twinning in the CG population (Figure 3). Although causes for MZ twinning are still obscure, DZ twinning has a clear familial aggregation, which is taken as evidence for genetic predisposition [11]. LSP, despite being a small district with few baptisms, concentrates roughly one third of all CG twins. The increased DZ twinning proportions in LSP thus indicates that such contribution for overall twinning occurs because there are genetic factors

enhancing twinning rate in CG as a whole and in LSP in particular. Secondly, strong evidence for the hypothesis of a genetic founder effect comes from the higher inbreeding coefficient found for women who gave birth to twins compared to other women.

Currently, there are several reported examples of the importance of genetic founder effects for some specific phenotypes which are more frequently in some human populations. As a general rule, a relatively high frequency of an autosomal recessive disease in an isolated population suggests a founder effect [37]. For example, it was suggested that a subgroup of Native American Athabascan populations living in Arizona and New Mexico have an unusually high incidence of severe combined immunodeficiency probably due to founder effect [38]. More recently, a case for the importance of genetic founder effects has been made for explaining the high frequency of genes responsible for 'single-gene' disorders and disease predispositions in Ashkenazi Jews compared to Sephardi Jews and non-Jews [39]. Another example is the Quebec population, in Canada. This population was founded by 8500 French settlers and shows, for some genetic disease, a geographical distribution consistent with a story of serial founder effects occurred during the migration of those settlers and their descendants [40]. It was also recently proposed that for the Sardinian population the relatively high frequency of glycogen storage disease type Ib in this population may be related to a founder effect [41].

Finally, our results also suggest that CG in general and LSP in special represent excellent population isolates where specific genetic variants influencing twin birth in humans may be identified. For the citizens of CG, our results may also be relieving, since we could formally reject the possibility that CG twins are a result of Nazi experiments and ideology. In this sense, our study illustrates how knowledge of population history and the genetic consequences may be of direct interest for the populations under study.

Acknowledgments

We are indebted to the following people that helped us with this study: Selia Heck, Vanusa Dresch, and Vanessa Schardong (from the Municipality of Cândido Godói), for all their help in contacting families with twins and obtaining church records.

Author Contributions

Performed the experiments: AT-R MZ-O AKS MO. Analyzed the data: AT-R MZ-O MRR GS UM NJRF LS-F. Contributed reagents/materials/analysis tools: MRR NJRF UM LS-F. Wrote the paper: AT-R NJRF GS LS-F.

References

- Hall JG (1996) Twinning: mechanisms and genetic implications. *Curr Opin Genet Dev* 6: 343–347.
- Lummaa V, Haukioja E, Lemmetyinen R, Pikkola M (1998) Natural selection on human twinning. *Nature* 394: 533–534.
- Beemsterboer SN, Homburg R, Gorter NA, Schats R, Hompes PG, et al. (2006) The paradox of declining fertility but increasing twinning rates with advancing maternal age. *Hum Reprod* 21: 1531–1532.
- Steinman G (2006) Can the chance of having twins be modified by diet? *Lancet* 367: 1513–1519.
- Imaizumi Y, Nishida E (2007) The frequency of recurrent multiple maternities using two sets of census data in Japan: 1990 and 1995. *Twin Res Hum Genet* 10: 638–643.
- Steinman G (2000) The mechanism initiating and controlling monozygotic twinning in humans remains to be elucidated. *Twin Res* 3: 337.
- Steinman G, Verni C (2007) *Womb Mates: A modern guide to fertility and twinning*. New York: Baffin Books.
- Al-Hendy A, Moshynska O, Saxena A, Feyles V (2000) Association between mutations of the follicle-stimulating-hormone receptor and repeated twinning. *Lancet* 356: 914.
- Hasbargen U, Lohse P, Thaler CJ (2000) The number of dichorionic twin pregnancies is reduced by the common MTHFR 677C→T mutation. *Hum Reprod* 15: 2659–2662.
- Montgomery GW, Zhao ZZ, Marsh AJ, Mayne R, Treloar SA, et al. (2004) A deletion mutation in GDF9 in sisters with spontaneous DZ twins. *Twin Res* 7: 548–555.
- Hoekstra C, Zhao ZZ, Lambalk CB, Willemsen G, Martin NG, et al. (2008) Dizygotic twinning. *Hum Reprod Update* 14: 37–47.
- Imaizumi Y (1998) A comparative study of twinning and triplet rates in 17 countries, 1972–1996. *Acta Genet Med Gemellol (Roma)* 47: 101–114.
- Fausser BC, Devroey P, Macklon NS (2005) Multiple birth resulting from ovarian stimulation for subfertility treatment. *Lancet* 365: 1807–1816.
- Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Menacker F, et al. (2005) Births: final data for 2003. *Natl Vital Stat Rep* 54: 1–116.
- Hall JG (2003) Twinning. *Lancet* 362: 735–743.
- Tong S, Caddy D, Short RV (1997) Use of dizygotic to monozygotic twinning ratio as a measure of fertility. *Lancet* 349: 843–845.
- Datusus (2010) Available: <http://tabnet.datusus.gov.br/cgi/defohtm.exe?sinasc/cnv/nvrs.def>. Accessed 2010 Nov 28.

18. Matte U, Le Roux MG, Bénichou B, Moisan JP, Giugliani R (1996) Study on possible increase in twinning rate at a small village in south Brazil. *Acta Genet Med Gemellol* 45: 431–437.
19. IBGE (2010) Available: http://www.ibge.com.br/home/estatistica/populacao/censo2000/populacao/tabela_brasil.shtm. Accessed 2010 Nov 28.
20. Camarasa J (2008) Joseph Mengele: O Anjo da Morte da América do Sul. Buenos Aires: Grupo Editorial NORMA. 188 p.
21. Ware J, Posner G (1986) Mengele: The Complete Story. New York: McGraw-Hill. 364 p.
22. Ayres M, Ayres M, Jr., Ayres D, Santos AA (2007) BIOESTAT – Aplicações estatísticas nas áreas das ciências bio-médicas. Belém: Ong Mamirauá. 364 p.
23. Weinberg W (1934) Differenzmethode und Geburtenfolge bei Zwillingen. *Genetica* 16: 282–288.
24. Fellmann J, Eriksson AW (2006) Weinberg's differential rule reconsidered. *Hum Biol* 78: 253–275.
25. Crow JF, Mange AP (1965) Measurements of inbreeding from frequency of marriages between persons of the same surnames. *Eugen Q* 12: 199–203.
26. Wright S (1951) The genetical structure of populations. *Ann Eugen* 15: 323–354.
27. Rodríguez-Larralde A, Morales J, Barraí I (2000) Surname frequency and the isonymy structure of Venezuela. *Am J Hum Biol* 12: 352–362.
28. Bronberg RA, Dipierri JE, Alfaro EL, Barraí I, Rodríguez-Larralde A, et al. (2009) Isonymy structure of Buenos Aires city. *Hum Biol* 81: 447–461.
29. Barraí I, Rodríguez-Larralde A, Mamolini E, et al. (2000) Elements of the surname structure of Austria. *Ann Hum Biol* 27: 607–622.
30. Loyo MA, de Guerra DC, Izaguirre MH, Rodríguez-Larralde A (2004) Admixture estimates for Churuguara, a Venezuelan town in the State of Falcón. *Ann Hum Biol* 31: 669–680.
31. Scapoli C, Mamolini E, Carrieri A, et al. (2007) Surnames in Western Europe: A comparison of the subcontinental population through isonymy. *Theor Popul Biol* 71: 37–48.
32. Tarskaia L, El'chinova GI, Scapoli C, Mamolini E, Carrieri A, et al. (2009) Surnames in Siberia: a study of the population of Yakutia through isonymy. *Am J Phys Anthropol* 138: 190–198.
33. Ericson A, Kallen B, Aberg A (2001) Use of multivitamins and folic acid in early pregnancy and multiple births in Sweden. *Twin Res* 4: 63–66.
34. Lumley J, Watson L, Watson M, Bower C (2001) Modelling the potential impact of population-wide periconceptional folate/multivitamin supplementation on multiple births. *BJOG* 108: 937–942.
35. Li Z, Gindler J, Wang H, Berry RJ, Li S, et al. (2003) Folic acid supplements during early pregnancy and likelihood of multiple births: a population-based cohort study. *Lancet* 361: 380–384.
36. Eriksson AW (1973) Human twinning in and around Åland Islands. *Commentationes Biologicae*, 64: 1–159.
37. Zlotogora J (2007) Multiple mutations responsible for frequent genetic diseases in isolated populations. *Eur J Hum Genet* 3: 272–278.
38. Murphy S, Hayward A, Troup G, Devor EJ, Coons T (1980) Gene enrichment in an American Indian population: an excess of severe combined immunodeficiency disease. *Lancet* 8193: 502–505.
39. Charrow J (2004) Ashkenazi Jewish genetic disorders. *Fam Cancer* 4: 201–6.
40. Laberge AM, Michaud J, Richter A, Lemyre E, Lambert M, et al. (2005) Population history and its impact on medical genetics in Quebec. *Clin Genet* 4: 287–301.
41. Zappu A, Lilliu F, Podda RA, Loudianos G (2010) Molecular analysis of glycogen storage disease type Ib in Sardinian population: evidence for a founder effect. *Genet Test Mol Biomarkers* 3: 399–403.

Anexo 4 – Artigo produzido em colaboração - Li-Fraumeni and Li-Fraumeni-Like Syndrome among children diagnosed with pediatric cancer in Southern Brazil (Abstract, aceito em Cancer, 2013)

Giacomazzi J, Selistre SGA, Rossi C, Alemar B, Santos-Silva P, Pereira FS, Netto CB, Cossio SL, Roth DE, Brunetto AL, Zagonel-Oliveira M, Martel-Planche G, Goldim JR, Hainaut P, Camey SA, Ashton-Prolla P

ABSTRACT

Background: Pediatric cancers are a feature in Li-Fraumeni syndrome and its variant Li-Fraumeni-like syndrome (LFS/LFL). *TP53* germline mutations are currently the only molecular defect known to be associated with this disease. Recently, a specific germline mutation in this gene, p.R337H, has been reported at a high prevalence in Brazil. **Methods:** The prevalence of LFS/LFL was investigated in cancer-affected children diagnosed with tumors of LFS/LFL spectrum and in a small consecutive series of cancer-unaffected controls. Additionally, the prevalence of the germline p.R337H mutation was investigated in overall cancer-affected children and of other germline *TP53* mutation only in children fulfilling clinical criteria for LFS/LFL. **Results:** Among cancer-unaffected children (n=65), 1.5% had a family history (FH) of LFL. Among cancer-affected children (n=292), 25.3% had a FH of LFL (p<0.001). Screening for the p.R337H identified 11 (3.7%) carriers: 9 diagnosed with adrenocortical carcinomas (ACC) and 2 with choroid plexus carcinomas (CPC). One of the ACC probands was homozygous mutant. The Brazilian founder haplotype and loss of heterozygosity at the p.R337H locus was present in all carriers. In addition, direct sequencing of the entire *TP53* coding region and gene rearrangement analysis of probands fulfilling LFL criteria (Eeles 2, Birch and/or Chompret) and negative for the p.R337H mutation, revealed a DNA binding domain mutation, p.G245S, in one child. **Conclusion:** *TP53* p.R337H testing should be offered to all Brazilian children diagnosed with ACC and CPC. A significant proportion of children with cancer in Southern Brazil fulfills criteria for LFL and should be referred for genetic risk assessment.

Anexo 5 – Artigo produzido em colaboração - Prevalence of the *TP53* p.R337H mutation in breast cancer patients in Brazil (abstract, submetido ao Journal of Clinical Oncology)

Juliana Giacomazzi, Marcia S Graudenz, Cynthia ABT Osorio, Patricia Koehler-Santos, Edenir I Palmero, Marcelo Zagonel-Oliveira⁷, Rodrigo D Michelli, Cristovam Scapulatempo Neto, Gabriela C Fernandes, Maria Isabel WS Achatz, Ghyslaine Martel-Planche, Fernando A Soares, Maira Caleffi, José Roberto Goldim, Pierre Hainaut, Suzi A Camey, and Patricia Ashton-Prolla

ABSTRACT

Background: Germline *TP53* mutations predispose individuals to multiple cancers and are associated with Li-Fraumeni/Li-Fraumeni-Like Syndromes (LFS/LFL). Breast Cancer (BC) represents over 25% of all cancers diagnosed in *TP53* mutation carriers. The mutation *TP53* p.R337H has been detected in 0.3% of the general population in southern and southeastern Brazil. Although this mutation is common in LFS/LFL families, it has also been identified in subjects without LFS/LFL familial history (FH). Methods: This study was designed to assess the prevalence of p.R337H in two BC-affected women groups: (1) 59 women with a FH suggestive of a hereditary cancer syndrome but no LFS/LFL features; (2) 815 women unselected for a FH of cancer diagnosed with BC at or before age 45 or at age 55 or older. Results: Among group 1 and group 2 patients, 2/59 (3.4%) and 70/815 (8.6%), respectively, were p.R337H carriers in the germline. The prevalence of p.R337H was higher in women who were diagnosed with BC at or before age 45 (12.1%) than in women who were diagnosed at age 55 or older (5.1%, $p < 0.001$). The Brazilian founder p.R337H haplotype was detected in all of mutation carriers analysed. The analysis of breast tumour tissues revealed that the loss of heterozygosity (LOH) was not a frequent event in p.R337H carriers and that these tumours were more likely to be HER2-positive than age-matched tumours from non-p.R337H carriers. Conclusions and impact: These results suggest that the inheritance of this mutation significantly contributes to the high incidence of BC in Brazil.