UNIVERSIDAD DE GRANADA
FACULTAD DE PSICOLOGIA
DOCTORADO EN PSICOLOGÍA
MENCIÓN EN PERSONALIDAD,
EVALUACIÓN PSICOLÓGICA Y PSICOPATOLOGÍA

CARACTERIZACIÓN POBLACIONAL Y FONOAUDIOLÓGICA DE NIÑOS RESIDENTES EN LA ISLA ROBINSON CRUSOE (CHILE), CON ALTA PREVALENCIA DE TEL. CONSECUENCIA DE UN EFECTO FUNDADOR

PÍA VILLANUEVA BIANCHINI

TRABAJO DE INVESTIGACIÓN
REQUISITO PARA OPTAR AL
GRADO DE DOCTOR EN PSICOLOGÍA

DIRECTORAS

Prof. Elvira Mendoza
Prof. Mª Dolores Fresneda
Prof. Dra. Lilian Jara Sosa

Granada - España
2015

Editor: Universidad de Granada. Tesis Doctorales
Autora: Pía Villanueva Bianchini
ISBN: 978-84-9125-857-5
URI: http://hdl.handle.net/10481/43685

“Oh! my beloved island!
I wish I had never left thee!”
1710

,

Exclamación de Alejandro Selkirk. Marino escocés, quien después de 4 años 4 meses de abandono en esta isla fue rescatado y trasladado a su tierra natal e inspirara a Daniel Defoe en su novela “...Robinson Crusoe...” (1719).
Citado por Benjamín Vicuña Mackenna (1883): en Juan Fernández: Historia verdadera de la isla Robinson Crusoe

.....a mis preciosos hijos
Felipe y Agus

... Angelina, Enrique, Stéfano

… a Zule, Angélica, Andrea, Paula y Adrián

AGRADECIMIENTOS

Sincera gratitud a todos los habitantes de la isla Robinson Crusoe por su valiosa y desinteresada colaboración, a su Alcalde Sr. Felipe Paredes, al personal de la Municipalidad, Registro Civil, Escuela Rural, y Posta Rural, en especial a la Dra. Mónica Quevedo por su constante dedicación.

Especiales agradecimientos a mis directoras, Prof. Dra. Elvira Mendoza, Prof. Lola Fresneda, Prof. Dra. Lilian Jara Soza, por su confianza al haber aceptado dirigir esta tesis y por su inestimable ayuda en la formulación del proyecto de investigación y en la tarea posterior de aplicación del método científico.

En forma especial a Prof. Dra. Dianne Newbury, Prof. Flga. Zulema de Barbieri, Flga. Maria Angélica Fernández, Prof. Dra. Carmen Julia Coloma, Prof. Dr. Marcelo Cano y a todos los colegas y amigos, en todos los ámbitos de trabajo, quienes ya sea en forma general o específica han contribuido al desarrollo y finalización de la presente tesis.

A todos los que apasionadamente conforman el Laboratorio de Genética Humana, Biología Molecular y Genética del Instituto de Ciencias Biomédicas de la Universidad de Chile; el Wellcome Trust Centre of Human Genetics de la Universidad de Oxford en el Reino Unido; el Max Planck Institute for Psycholinguistics y el Donders Institute for Brain, Cognition and Behaviour en Nijmegen Holanda, ya que hicieron posible que comprendiera y disfrutara el mundo de la biología molecular.

Agradezco también a la Armada de Chile, en particular a la dotación de los buques Aquiles, Valdivia y Chacabuco, por el apoyo logístico brindado a esta investigación científica.

Un especial y cariñoso agradecimiento a mis hijos y mi gran familia por su apoyo, compañía y comprensión incondicional en el desarrollo de este trabajo.

LISTA DE PUBLICACIONES

La presente memoria de tesis está compuesta por los siguientes artículos científicos:

1. Villanueva P, Fernández MA, DE Barbieri Z, Palomino H. Consanguinity on Robinson Crusoe Island an isolated Chilean population. *Journal of Biosocial Science*. 2014 Aug 12:1- 10.

Indexada 0.883 .

Q 3 Demography (15 de 25 revistas),

Q3 Social Sciences (27 de 37 revistas)

2. Villanueva P, Nudel R, Hoischen A, Fernández M A, Simpson N H, Gilissen C, Reader R R, Jara L,. Echeverry M M, Francks C, Baird G, Conti-Ramsden G, O'Hare A, Bolton P F, Hennessey E R, The SLI Consortium, Palomino H, Carvajal-Carmona L, Veltman J, Cazier J-B, De Barbieri Z, Fisher S E., Newbury D F. Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for Specific Language Impairment. *PLoS Genetics* 2015, 11: e1004925. doi:10.1371/journal.pgen.1004925.

Indexada: 8.167,

Q1. Genetics and Heredity (14 de 164 revistas)

3. Villanueva P, Quevedo M, De Barbieri Z, Piñeiro S, Herrero C, Fernández MA, Palomino H.Dental morphological markers as a proxy for ethnicity in

Robinson Crusoe islanders. *International Journal of Morpholgy*. Aceptado Marzo 2015.

Indexada: 0.200

Q 3 Anatomy and Morphology (14 de 20 revistas)

INTRODUCCIÓN

El archipiélago de Juan Fernández es territorio chileno; está conformado por las islas Robinson Crusoe, Santa Clara, Alejandro Selkirk y un conjunto de pequeños islotes. Se encuentra a 667 kilómetros frente a las costas centrales de Chile continental y pertenece administrativamente a la Región de Valparaíso. La Isla de Robinson Crusoe es la única habitada, con 633 habitantes según el último CENSO nacional autorizado (INE 2002) en el pueblo San Juan Evangelista, ubicado en la Bahía Cumberland.

El Archipiélago fue descubierto por Juan Fernández en 1574. En 1750 empezó la colonización al fundarse el pueblo de San Juan Evangelista. La isla tuvo varios intentos de colonización entre 1750 y 1850 cuando fue despoblada. A fines del siglo XIX se establece la última colonización, con su arrendamiento al Barón suizo Alfred Von Rodt, quien con un grupo de 8 familias repobló la isla. (Arana, 2010; Brescia de Val, 1979; Ruh, 1975; Takahashi, 2002; Vicuña Mackenna, 1883).

Estas familias estaban compuestas por 37 adultos (europeos y chilenos) y diez niños (Brinck, 2005; Orellana, 1975; Schnyder, 2005). El registro civil e identificación de Chile fue abierto en la isla en el año 1902, lo que coincide con el nacimiento de la primera generación isleña. En el año 1940, se incluye por primera vez a la población de la isla en un CENSO nacional y se reportan 434

habitantes. Los valores de densidad poblacional de la isla se han mantenido estables en los últimos 60 años. Mientras que la densidad poblacional del país aumenta gradualmente, la densidad de la isla en términos generales se mantiene, mostrando un leve aumento en los últimos 10 años (INE, 2002). Es necesario destacar que la isla es Parque Nacional por lo que sólo se puede destinar a habitación una mínima parte de su superficie (CONAF, 1976, 2009).

La población de la isla tiene en la actualidad seis generaciones nacidas desde la última colonización, además de habitantes no isleños que son especialmente empleados de gobierno, o de las fuerzas armadas y de orden que son designados a la isla por periodos relativos (2 a 6 años). La población reporta poca migración, cuenta con transporte marítimo y aéreo muy limitados. Un barco al mes lleva los víveres y provisiones a sus habitantes. También, cuenta con un aeródromo con capacidad para pequeñas avionetas, aunque debe cerrar muchas veces en el año, debido a las condiciones climáticas poco favorables (Figueroa 2006).

La economía de la isla se basa en la pesca artesanal y en la captura de la langosta (Arana 2000), así como recientes esfuerzos por consolidar el turismo. Los habitantes de la isla hablan español, idioma oficial de Chile.

Actualmente, el 77% de la población infantil total, tiene al menos uno de los apellidos del grupo colonizador, lo que podría ser indicativo de un mayor

grado de consanguinidad poblacional, concordante con una población joven, pequeña y semiaislada (Villanueva, 2008).

En poblaciones pequeñas existen elevadas frecuencias de alelos deletéreos, conocido como deriva génica, que es la fluctuación por azar de la frecuencia de los alelos que ocurre en poblaciones de tamaño pequeño. El efecto fundador es una de las formas de la deriva genética y se refiere a la existencia de un alelo infrecuente en uno de los fundadores de la población y que produciría una frecuencia superior de ese alelo a la que existe en la población de la que proviene. (Cavalli Sforza, 1971).

En estudios anteriores hemos reportado una alta prevalencia de Trastorno Específico de Lenguaje (TEL) en la población infantil de la Isla de Robinson Crusoe, pudiendo ser ésta atribuida a un efecto fundador (Villanueva, 2008)

El Trastorno Específico de Lenguaje (TEL), en inglés Specific Language Impairment (SLI) (MIM 602081) se define como dificultades de desarrollo del lenguaje oral que no puede ser explicada por déficit en umbrales auditivos, ni por retraso intelectual, trastornos psicopatológicos o desigualdad sociocultural (Aguado, 2000; Bishop, 2011; Narbona, 1999).

El TEL es uno de los trastornos de la comunicación oral más frecuente en los niños preescolares, con prevalencias estimadas entre el 2 y 7% de la población pre-escolar de habla inglesa (Tomblin, 1997ª; Law, 1998). La

American Psychiatric Association (1994) indica una prevalencia de TEL entre un 6 y 8 % (Shriberg, 1999). Estudios en nuestro país indicarían la presencia de TEL en el 4% de los niños entre 3 y 7 años. (De Barbieri, 1999). Se presenta con mayor frecuencia en varones (De Barbieri, 1999; Shriberg, 1999).

El Trastorno Específico del Lenguaje se caracteriza por presentar una semiología lingüística variada donde se pueden afectar los distintos niveles del lenguaje. Estas dificultades van desde problemas puntuales en la producción de las palabras hasta formas más severas que dificultan una buena interacción comunicativa. Este Trastorno Específico se puede clasificar en dos grandes grupos: a)Trastorno expresivo: que se caracteriza por presentar dificultades en la formulación del lenguaje, en que se afecta la dimensión fonológica y/o déficit fonológico-sintácticas y en menor grado semánticas; y b) Trastorno comprensivo–expresivo: es un trastorno mixto donde se ven afectados todos los niveles del lenguaje (fonológico, semántico, morfosintáctico y pragmático) de forma variable, en las dimensiones expresivas y comprensivas del lenguaje oral. (Aguado, 2000; Conti – Ramsden, 1997; Falcaro, 2008; Mendoza, 2001)

Si este problema no es detectado y tratado a tiempo puede afectar otras áreas de aprendizaje del niño, especialmente la lectoescritura. Los menores con este tipo de cuadro generalmente presentan un inicio más tardío de la adquisición de las primeras palabras, lo que significaría tener dificultades en la memoria de trabajo y en habilidades metalingüísticas. Sin embargo, otras áreas del desarrollo relacionadas con aspectos no verbales se encontrarían dentro de

parámetros normales. Esto determina que el déficit del lenguaje sería un aspecto nuclear en el desarrollo psicomotor de estos niños. (Aguado, 2000; De Barbieri, 2004; Martínez, 2002; Orellana, 1996)

Para identificar un niño con TEL es fundamental la historia clínica, donde no se evidencie déficit auditivo, retraso psicomotor, déficit cognitivo, ni trastornos de la personalidad. Se ha descrito que frecuentemente hay antecedentes de déficit de lenguaje o aprendizaje en algún familiar directo del propósito afectado (Narbona, 1999).

Así, distintas evidencias indican la presencia de mecanismos genéticos involucrados en la susceptibilidad a los Trastornos Específicos del Lenguaje. Diferentes estudios han demostrado una fuerte agregación familiar. Es así como se ha observado que el 46% de los familiares de los probandos están también afectados, comparado con un 18% en los del grupo control (Stromswold, 1998, 2001). Tomblin y Buckwalter (1998) encontraron un 22% de parientes en primer grado afectados comparados con el 4 a 6% de la población general, por lo que la heredabilidad estimada es entre un 45 y un 77%. Estudios en mellizos encuentran una concordancia que fluctúa ente 70 y 96% en monocigóticos y entre 46 y 69% en dicigóticos (Bishop, 1995; Tomblin, 1998).

Tomblin y cols (2000) indicaron en base a análisis de segregación compleja que el modelo que mejor se adecua a los datos obtenidos de genealogías de familias de propósitos afectados es la presencia de un gen

mayor de naturaleza dominante, aunque no descartan transmisión poligénica, señalando que existen factores genéticos con un modo complejo de transmisión. Una familia extensa de tres generaciones estudiada en Inglaterra, en la cual más de la mitad de sus miembros estaban afectados por alteraciones severas del habla y del lenguaje, permitió postular a esta patología con un patrón de herencia simple de naturaleza autosómica dominante (Hurst, 1990).

Lai y cols (2000, 2001) detectaron que el gen FOXP2, que codifica para un factor de transcripción, ubicado en el cromosoma 7q31, estaría relacionado con los procesos del desarrollo que culminan en el habla y el lenguaje (Fisher, 1998). El SLI Consortium, (institución inglesa de investigación en TEL) identificó otros 2 loci relacionados con TEL, ubicados en el cromosoma 16q (SLI1) y 19q (SLI2) que podrían tener importancia como factor de riesgo para las alteraciones de lenguaje (The SLI Consortium 2002). Agrupando la información de otros estudios y amplificando la muestra, confirman un fuerte ligamiento del TEL al locus SLI1 en el cromosoma 16q. (The SLI Consortium 2004). Para luego mediante análisis multivariado de ligamiento comprobar cuáles eran los mejores marcadores fenotípicos para cada área de ligamiento descrita (Monaco, 2007). Además, se ha encontrado en una población canadiense de familias de ancestros célticos evidencia de ligamiento entre TEL y el locus 13q21 y posiblemente al cromosoma 2p (Barlett, 2002).

Múltiples zonas de ligamiento se han descrito en diferentes cromosomas, en algunos de ellos se han comparado los resultados con otras patologías como

dislexia, autismo y déficit atencional (Ver Newbury, 2010); sin embargo, todos los estudios se reportan en familias de habla inglesa.

Debido a nuestro interés en la población infantil de la isla de Robinson Crusoe, que tiene una excepcional alta prevalencia de TEL (35%), hemos comenzado los análisis genético-moleculares, obteniendo los primeros resultados moleculares en niños de habla hispana, describiendo asociación de TEL a un marcador microsatélite ubicado en el cromosoma 16 (Villanueva, 2010); dicho marcador fue previamente descrito en la población inglesa por el grupo de SLI Consortium (2002, 2004).

Mediante un trabajo colaborativo internacional con renombrados investigadores de la Universidad de Oxford y SLI Consortium, hemos realizado el Genome Wide Scann (GWS) de esta población y hemos identificado cinco interesantes regiones cromosómicas que se presentan con mayor frecuencia en los individuos afectados que lo esperado por sólo azar (Villanueva, 2011). Se hace necesario entonces continuar con los análisis genéticos para tentar la identificación de causas genéticas subyacentes.

Por lo anterior, el primer propósito de este estudio es:

a) realizar un estudio de asociación genético en relación al TEL, en sujetos de la población infantil de la isla Robinson Crusoe, mediante estudios de secuenciación exómica.

Debido a que los estudios genéticos realizados hasta ahora concuerdan con la literatura internacional, resulta necesario entonces, caracterizar a la población estudiada con el fin de poder comparar nuestros resultados con poblaciones europeas y norteamericanas, considerando los componentes étnicos respectivos y describir las características fenotípicas máxilofaciales de estos niños con el fin de contar con posibles marcadores morfológicos que orienten futuros estudios.

Por los que los siguientes propósitos de este estudio son:

b) establecer el grado de consanguinidad en la población de la isla Robinson Crusoe

c) establecer el grado de miscegenación indígena en la población de la isla Robinson Crusoe

HIPÓTESIS

Los niños isleños con TEL presentan características genéticas que difieren de las presentes en los niños isleños con desarrollo típico del lenguaje.

La población isleña presenta un mayor grado de consanguinidad que lo reportado en poblaciones geográficamente no aisladas.

La población actual de la isla, proveniente de una colonización mayoritariamente europea, presentando un bajo grado de miscegenación.

OBJETIVOS GENERALES

Identificar variantes de código mediante técnicas de genética molecular en pacientes afectados con TEL del Archipiélago de Juan Fernández.

Establecer el grado de consanguinidad y el coeficiente de endogamia en la población de la isla de Robinson Crusoe.

Establecer el grado de miscegenación de la población de la isla de Robinson Crusoe.

OBJETIVOS ESPECÍFICOS

1.1 Identificar las variantes de código en individuos isleños con y sin TEL, mediante secuenciación exómica.

1.2 Realizar análisis de asociación de la variante de código en la población isleña general.

1.3 Identificar cambios en proteínas, en individuos isleños con y sin TEL.

1.4 Comparar dichos códigos en individuos ingleses con diagnóstico de TEL.

2.1 Determinar el grado de consanguinidad de la población de la Isla Robinson Crusoe.

2.2 Determinar el coeficiente de entrecruzamiento de la población de la isla de Robinson Crusoe.

2.3 Determinar el tipo de matrimonios consanguíneos presentes en la población de la isla de Robinson Crusoe.

3.1 Determinar el grado de miscegenación de la población de la isla de Robinson Crusoe, mediante marcadores morfológicos dentarios.

J. Biosoc. Sci., page 1 of 10, © Cambridge University Press, 2013
doi:10.1017/S0021932013000436

CONSANGUINITY ON ROBINSON CRUSOE ISLAND, AN ISOLATED CHILEAN POPULATION

PIA VILLANUEVA*†‡§[1], MAŔA A. FERNÁNDEZ‡, ZULEMA DE BARBIERI‡ AND HERNÁN PALOMINO*

**Institute of Biomedical Sciences, University of Chile, Santiago, Chile, †Department of Pediatric Dentistry and Dentomaxillary Orthopedics, University of Chile, Santiago, Chile, ‡School of Speech and Hearing Sciences, University of Chile, Santiago, Chile and §Doctoral Program of Psychology, Graduate School, University of Granada, Spain*

Summary. The population of Robinson Crusoe Island is estimated at 633 inhabitants. The current population has a common origin from the first eight families who colonized the island at the end of the 19th century. The objective of this study was to determine the rates of consanguinity, the average coefficients of inbreeding, the types of consanguineous marriages and the inbreeding evolution between 1900 and 2000 on the island. All marriages registered on the island, from the last colonization until 2000 (417 in total), were included in the analysis. In addition, extended genealogies were obtained. The consanguinity rate was 14.9% and the average coefficient of inbreeding (α) 54.05×10^{-4}. The most frequent type of consanguineous marriages was between second cousins, followed by first cousins. The average value of the first/second cousin ratio was 1.11. The population of Robinson Crusoe Island has a high rate of inbreeding. The unique characteristic of the island – its small current population, originating from just a few families, with small rate of gene flow – could explain the observed high and increasing consanguinity.

Introduction

Consanguinity is an important mechanism of differentiation in the human population. It alters genotypic frequencies and therefore the structure of the population (Cavalli-Sforza & Bodmer, 1971). There are geographical areas in the world where the average consanguinity of the population is still high. Modern consanguinity usually arises from sociological-related issues such as religion, social systems based on castes, population size, migration rates, geographical isolation and the number of generations since the population was founded (Wang *et al*., 2002; Cohen *et al*., 2004; Patton, 2005; Blanco Villegas & Fuster, 2006; El-Kheshen & Saadat, 2013). All these factors can mean that a certain population has little gene exchange with other populations (Serre & Babron, 1985; Rothhammer & Llop, 2004). High levels of consanguinity were found in the

[1] Corresponding author. Email: piavilla@u.uchile.cl

1960s in some isolated populations in mainland Chile. For example Caleu, an isolated geographical area whose population has Spanish origins had an average coefficient of inbreeding (α) of 60.4×10^{-4} (Blanco & Covarrubias, 1971) and Pehuenches, a Chilean native ethnical group, exhibited a high level of consanguinity ($\alpha = 82.9 \times 10^{-4}$), mainly due to land inheritance (Blanco & Chakraborty, 1975). In contrast, due to an ancient Inca regulation, consanguineous marriages were unacceptable among the northern native ethnical groups of Chile such as the Aymaras ($\alpha = 1.5 \times 10^{-4}$) and Atacameños ($\alpha = 4.1 \times 10^{-4}$) in the same decade (Rothhammer & Llop, 2004).

In an inbred population, the gene pool is likely to reflect a founder effect. This is because of the change in the genotypic frequencies from one generation to another and the increase in the frequency of homozygotes. In turn this can potentiate the existence of uncommon genetic pathologies in the population, and also can influence reproductive and survival capabilities (Machado *et al.*, 2013a, b).

The present research was carried out in Robinson Crusoe Island, an isolated Chilean population, where 77% of the current infant population exhibits at least one founder surname. A high prevalence of Specific Language Impairment (SLI) has been described among islander children (Villanueva *et al.*, 2008). This is the failure to normally develop language when the appropriate environment is provided (Law *et al.*, 2000). The prevalence of SLI among islander children (35%) is higher than that reported for children from other countries and from the mainland (4–8%) (Tomblin *et al.*, 1997a, b; Shriberg *et al.*, 1999; De Barbieri *et al.*, 1999) and also for those children not originally from Robinson Crusoe Island but who lived in the island (4%) (Villanueva *et al.*, 2008, 2010, 2011).

Robinson Crusoe Island is part of the Juan Fernandez Archipelago and is the only inhabited island in the archipelago. The archipelago is located 361 miles from the central coast of Chile (33° 37′ 56″ S; 78° 49′ 45″ W) (Brinck, 2005) and belongs governmentally to the Valparaiso region of Chile (see Fig. 1). The island was discovered in 1574 by the Spanish sailor Juan Fernandez. The first colonization took place in 1750. The island was later depopulated around 1850 by an order of the Chilean government. The current population comes directly from the last colonization, which occurred in 1881 with only eight families. These families comprised 37 European and Caucasian Chilean founders including ten children. Two of these male founders were siblings; the other six families were not related (Vicuña Mackenna, 1883; Orellana *et al.*, 1975; Schnyder, 2004).

In 1902 the General Register Office was created. The 1940 Chilean census is the first official record providing information about the total population of the island (434 inhabitants). According to the most recent Chilean census (INE, 2002) the island's population was estimated at 633 inhabitants. The number of women was lower than the number of men, there being no evidence to explain this difference. This ratio has been increasing over the years (see Fig. 2). Following the initial population growth, the average population density on Robinson Crusoe Island remained very similar from 1940 to 1992, and lower than the average of the mainland region to which the island belongs governmentally (Valparaiso region) the region in Chile with the second highest density. While the total Chilean population density gradually increased over the study period (1940–2002), the population of the island decreased. However, it increased slightly over the last decade studied (1990–2000) (INE, 2002) (see Fig. 3).

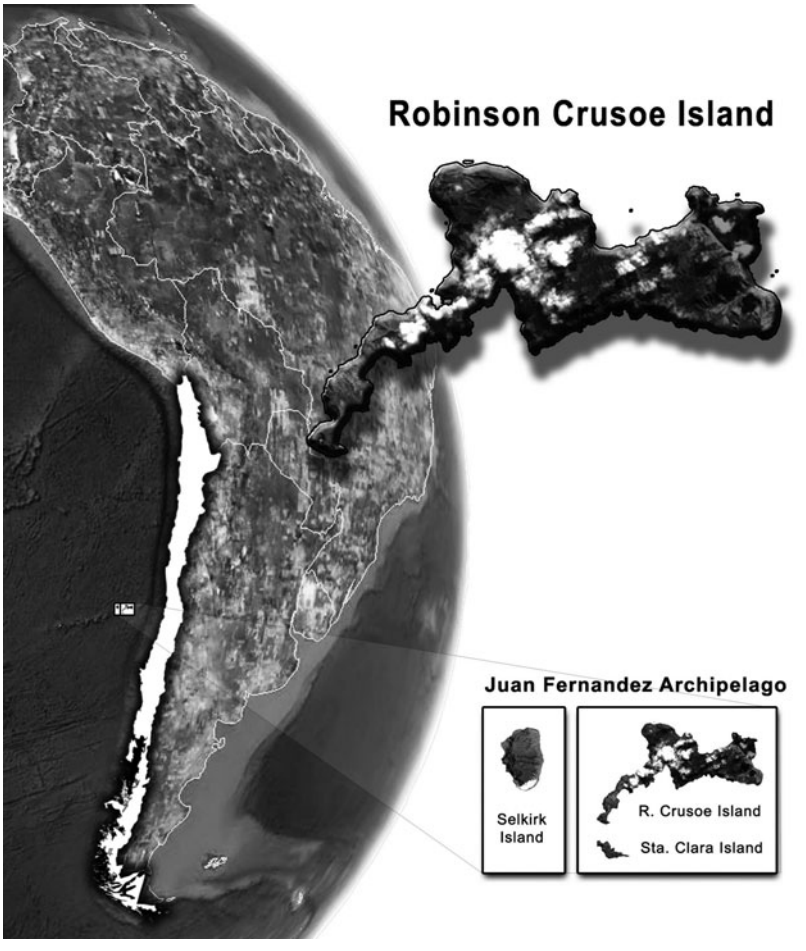


Fig. 1. Geographic area studied: Robinson Crusoe Island, Juan Fernandez Archipelago, Chile.

The economy of the island is based almost exclusively on fishing, especially lobsters, endemic fish and golden crabs. In the last 15 years studied (1985–2000) there has been an increase in tourism-related businesses, which may explain the population growth in the decade 1990–2000. The people of Robinson Crusoe Island speak Spanish, the only official language in Chile. Both sea and air transport are very limited on the island. Only one vessel per month brings food and other basic goods. Recently an aerodrome for small aeroplanes was built which, due to climate conditions, has to close several times throughout the year.

Given the geographic isolation of this population and the documented founder effect on language impairment prevalence (Villanueva *et al.*, 2010), the aim of the present study was to determine rates of consanguinity, the average coefficients of inbreeding (α), the types of consanguineous marriages, and inbreeding between 1900 and 2000 on Robinson Crusoe Island.

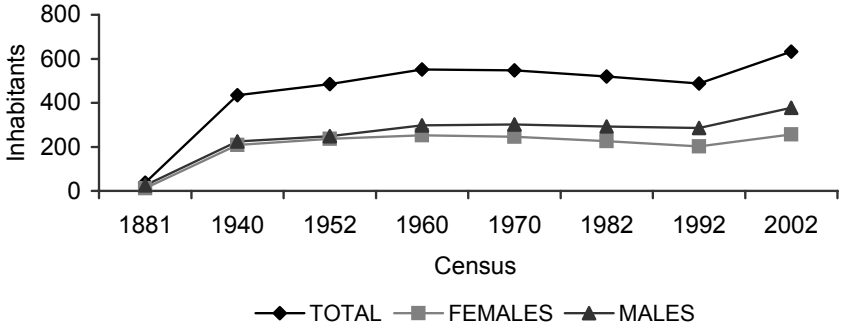


Fig. 2. Population development on Robinson Crusoe Island from 1881 to 2002.

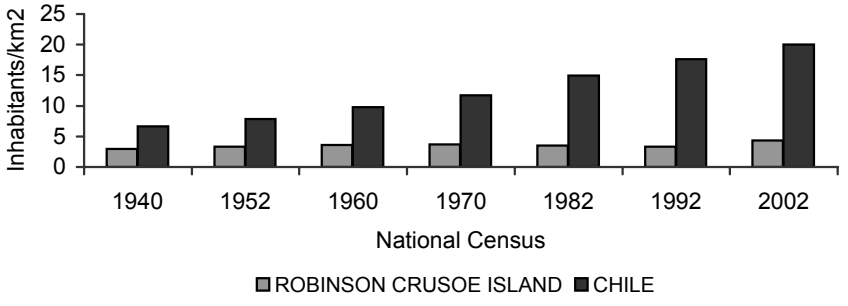


Fig. 3. Population density of Robinson Crusoe Island and Chile from 1940 to 2002.

Methods

The study used a retrospective research design to assess the total number of marriages registered on Robinson Crusoe Island over the ten decades from 1900 to 2000. The data were taken from a village genealogy compiled by the research group from birth records going back to the last colonization. The genealogical records, which are organized into surname groups, were supplemented with any additional data that could be found in civil and family records of births, marriages and deaths. The records of individuals from 1900 to 2000 were classified into the eight founder families with four to six generations since the foundation date.

Rates of consanguinity, average inbreeding coefficient (α coefficient or Bernstein coefficient) and the percentage of different types of consanguineous marriages were calculated (Cavalli-Sforza & Bodmer, 1971). Uncle–niece and/or aunt–nephew marriages (M12, $F = 1/8$), and those between first cousins (M22, $F = 1/16$), first cousins once removed (M23, $F = 1/32$), half first cousins ($M22^{1/2}$, $F = 1/32$), second cousins (M33, $F = 1/64$), second cousins once removed (M34, $F = 1/128$), half second cousins ($M33^{1/2}$, $F = 1/128$), and even third cousins (M44, $F = 1/256$) were considered for the purposes of this research, because extensive genealogies were available over six generations (Cavalli-Sforza & Bodmer, 1971; Varela *et al.*, 2001).

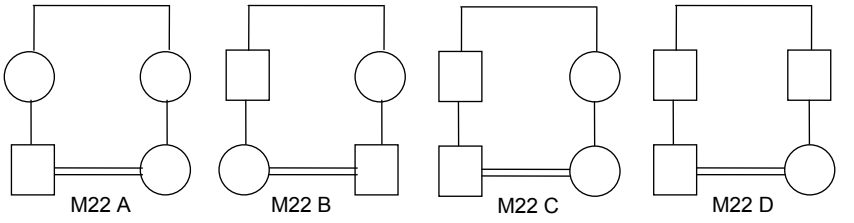


Fig. 4. Types of first cousin consanguineous marriages. Squares, male; circles, female. Reproduced from Haldane & Moshinsky (1939).

According to the type of analysis, the calculations were performed every 20 years for the period between 1900 and 2000. The ratio of marriages between first cousins and second cousins (M22/M33) was also calculated, as this is considered to be an important indicator of evolution of the structure of a population (Varela *et al.*, 2001). The percentages of different types of first cousin marriages (M22A, M22B, M22C, M22D) were calculated to estimate non-random factors (Haldane & Moshinsky, 1939) (see Fig. 4).

The protocol of the study was approved by the Ethics Board of the School of Medicine, University of Chile.

Results

Today the founder families of Robinson Crusoe Island have between four and five generations born on the island. Between 1900 and 2000, 417 marriages were registered. The average consanguinity rate was 14.9% and the average coefficient of inbreeding (α) was 54.05×10^{-4}. During the period analysed 62 marriages were consanguineous, and of seven different types. The absolute and relative frequencies of the different types of consanguineous marriages are given in Table 1. The most frequent consanguineous marriages were between second cousins (M33: 5.04% of the total marriages). The least frequent marriages were between half second cousins ($M33^{1/2}$: 0.48%). Temporal variation of the proportion of the different types of consanguineous marriages every 20 years is also shown in Table 1. It is possible to see a trend of increased consanguinity on the island over time. There was a clear increase in the relative proportion of M33 and M44 marriages, at the expense of a decrease in M22 marriages.

The relative frequencies of $M33^{1/2}$ (half second cousins) marriages decreased throughout the period studied. There were no marriages between uncles and nieces or aunts and nephews (M12) in the 100 years analysed. In consanguineous marriages between first cousins ($n = 17$), the highest frequency was between cousins who were the offspring of two brothers (M22D: 47%). The other relation values were M22A: 5.9%, M22B: 35.3%, M22C: 11.8%.

The average first cousin/second cousin (M22/M33) ratio from 1941 to 2000 was 1.11. There were no second cousin marriages between 1900 and 1940. The M22/M33 ratio for the next period (1941–1960) was 2.67. From 1961 to 1980 the ratio was 0.57. For the last 20-year period (1981–2000) the ratio was 0.09. The variation of the inbreeding coefficient between 1900 and 2000 in 20-year periods is shown in Table 2. This shows a gradual increase until reaching maximum values in the decennia 1981–2000 (values above 96.84×10^{-4}).

Table 1. Absolute and relative frequencies of the different types of consanguineous marriages, Robinson Crusoe Island, 1901–2000

	M22		$M22^{1/2}$		M23		M33		$M33^{1/2}$		M34		M44		Consanguineous		Not consanguineous		
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	Total
1901–1920	1	2.08	0	0	0	0	0	0	0	0	0	0	0	0	1	2.08	47	97.92	48
1921–1940	3	5	0	0	0	0	0	0	0	0	0	0	0	0	3	5	57	95	60
1941–1960	8	6.4	1	0.8	1	0.8	3	2.4	0	0	0	0	0	0	13	10.4	112	89.6	125
1961–1980	4	4.76	2	2.38	0	0	7	8.33	2	2.38	1	1.19	0	0	16	19.05	68	80.95	84
1981–2000	1	1	2	2	3	3	11	11	0	0	4	4	8	8	29	29	71	71	100
Total	17	4.07	5	1.2	4	0.96	21	5.04	2	0.48	5	1.20	8	1.92	62	14.9	355	85.1	417

Table 2. Variation of the coefficient of inbreeding (α) every 20 years, Robinson Crusoe Island, 1901–2000

	M22		M22 1/2		M23		M33		M33 1/2		M34		M44		Total
	n	α[a]	*n*	α	*n*	α	*n*	α	*n*	α	*n*	α	*n*	α	α
1901–1920	1	13	0	0	0	0	0	0	0	0	0	0	0	0	13
1921–1940	3	31.25	0	0	0	0	0	0	0	0	0	0	0	0	31.25
1941–1960	8	40	1	2.5	1	2.5	3	3.75	0	0	0	0	0	0	48.75
1961–1980	4	29	2	7.45	0	0	7	13.02	2	1.86	1	0.93	0	0	52.26
1981–2000	1	6.25	2	6.26	3	9.39	11	42.14	0	0	4	6.25	8	26.55	96.84

[a] α ($\times 10^{-4}$).

Discussion

In Chile, consanguinity rates have been analysed in several native populations (Mapuches, $\alpha = 57.9 \times 10^{-4;}$ Pehuenche, $\alpha = 82.9 \times 10^{-4}$), non-native Chilean populations living in urban areas (Viña del Mar, $\alpha = 3.3 \times 10^{-4}$; La Serena, $\alpha = 5.43 \times 10^{-4}$), non-native Chilean populations living in rural areas (Olmue, $\alpha = 37 \times 10^{-4}$) and in the population living on Easter Island, where the α value was 5.6×10^{-4} (Quezada & Barrantes, 1973; Champin *et al*., 1976; Villarroel *et al*., 1978; Zuñiga, 1978; Cruz Coke, 1989). All of the above populations show a tendency for α values to decrease over the years, which is as expected for expanding populations with no restriction on marriages. In contrast, the average consanguinity on Robinson Crusoe Island constantly increased from 1900 to 2000. At the beginning of the XXI century, 77% of the population has at least one parent who was descended from the founder families. This may be due to the low rate of immigration to the island and the stability of the size of the population, which has grown slowly with slight differences in density throughout the study decades (see Figs 2 and 3).

The number of consanguineous marriages among the islanders has increased due to a growing number of second cousin and third cousins marriages. The period from 1941 to 2000 showed the highest number of consanguineous marriages. This reflects the fact that during this period the population grew to comprise at least two generations born on the island. Before 1941 there were only four marriages between first cousins.

On the other hand, during the same period the native population of mainland Chile had opposite social behaviours. North Chilean Amerindians (Aymaras [$\alpha = 1.5 \times 10^{-4}$] and Atacameños [$\alpha = 4.1 \times 10^{-4}$]) have always been characterized by low degrees of consanguinity because of their social structure, where consanguineous marriage is considered a taboo. In contrast, southern Chilean Amerindian (Mapuches [$\alpha = 57.9 \times 10^{-4}$] and Pehuenches [$\alpha = 82.9 \times 10^{-4}$]) encouraged consanguineous marriages. Currently, among these latter groups there is still a high tendency to encourage certain types of first cousin mating due to their matri- or patrilineal social system (see Fig. 4) (Rothhammer & Llop, 2004). There is no evidence that on Robinson Crusoe Island there was a preferred mating. However, the present study observed a higher frequency of matings between the offspring of brother pairs. This may primarily be due to the sex distribution on the island (Fig. 2) as there are no known cultural, religious or land-ownership factors involved in the choice of spouses.

The most two common types of consanguineous marriages in the world population are those between (a) first cousins and (b) second cousins. In the population of Robinson Crusoe Island first cousin mating and second cousin mating together represent over the 60% of the total consanguineous marriages. These types of marriages involve spouses with the highest degree of consanguinity. Thus they contribute with a high load to homozygosity as the spouses are most likely to share ancestral chromosomes.

Nevertheless, the number of matings between individuals with shared relatives more than two generations back increased noticeably from 1940 to 2000. This explains the growing tendency of the observed consanguinity on the island and may be, in part, due to the growing size of the population.

The average values of the consanguinity rate (14.9%) and the coefficient of inbreeding (54.05×10^{-4}) on Robinson Crusoe Island during the period 1900–2000 were

higher than those registered in other non-native small areas of Chile over the same period of time (Rothhammer & Llop, 2004). A large part of the observed consanguinity can be explained by the geographical isolation of the island and its poor transport links. Certainly, improved transport between the island and the mainland would be expected to break down this isolation. However, if this hypothetic improvement did happen, there would be other factors to take into consideration. First, the island is small in size with a difficult geography to live in. It has a volcanic origin with almost no flat surfaces or plateaus. Second, it has been declared a World Heritage Site by UNESCO and thus most of the areas on the island cannot be used for housing. These two issues restrict immigration to the island. Finally, because of the lack of tertiary education on the island, young adults have to leave to go to the mainland for further education opportunities.

Despite the suggestions from some scientists in the early 1960s about the uniqueness of the population structure of the island (Orellana *et al*., 1975), no systematic studies have been carried out in this population, as they have in other Chilean islands such as Easter Island (Cruz-Coke, 1989). This study found a high level of consanguinity, primarily accounted for by first and second cousin marriages. The unique characteristic of the island – its small current population, originating from just a few families, with small rate of gene flow – could explain the observed high and increasing consanguinity. The population structure of Robinson Crusoe Island offers an unparalleled opportunity to study the underlying causes of genetic pathologies on the island.

Acknowledgments

The authors are extremely grateful to the inhabitants of Robinson Crusoe Island who agreed to participate in this study. They would like to thank the Mayor of the Ilustre Municipalidad de Juan Fernandez for his infinite assistance in the development of this research and the authorities of Armada de Chile for allowing travel to Robinson Crusoe Island. Monica Garcia is thanked for her work on the genealogical records, and also Dr Dianne Newbury and Dr Adrian Fuente for their comments on the original manuscript, and Professor A. Bakr Rabie for his encouragement of this project. The work was specifically funded by Vicerrectoria de Investigacion, Universidad de Chile (www.uchile.cl), UCHILE DID TNAC 01-02/01, UCHILE DI MULT 05-05/02 grants. The authors declare no conflict of interest.

References

Blanco, R. & Chakraborty, R. (1975) Consanguinity and demography in some Chilean Population. *Human Heredity* **25**, 477–487.

Blanco, R. & Covarrubias, E. (1971) Estudio genético en Caleu. *Revista Médica de Chile* **99**, 139–144.

Blanco Villegas, M. J. & Fuster, V. (2006) Reproductive pattern in consanguineous and non-consanguineous marriages in La Cabrera, Spain. *Annals of Human Biology* **33**, 330–341.

Brinck, G. (2005) *Memorias insulares. Archipiélago Juan Fernández*. Azapa.

Cavalli-Sforza, L. & Bodmer, W. F. (1971) *The Genetics of Human Populations*. Freeman and Co., San Francisco.

Champin, J., Rothhammer, F., Blanco, R. & Palomino, H. (1976) Consanguinidad en poblaciones chilenas. *Revista Médica de Chile* **104**, 202–206.

Cohen, T., Vardi-Saliternik, R. & Friedlander, Y. (2004) Consanguinity, intracommunity and intercommunity marriages in a population sample of Israel Jews. *Annals of Human Biology* **31**, 38–48.

Cruz-Coke, M. (1989) The genes of Easter Islanders. *Revista Médica de Chile* **117**, 685–694.

De Barbieri, Z., Maggiolo, M. & Alfaro, S. (1999) Trastornos de la comunicación oral en niños que asisten a control de salud en un consultorio de atención primaria. *Revista Chilena de Pediatría* **70**, 36–40.

El-Kheshen, G. & Saadat, M. (2013) Prevalence of consanguineous marriages among Shi'a populations of Lebanon. *Journal of Biosocial Science* **10**, 1–8.

Haldane, J. B. S. & Moshinsky, P. (1939) Inbreeding in Mendelian populations with special reference to human cousin marriages. *Annals of Eugenics* **9**, 321–340.

INE Instituto Nacional de estadísticas (2002) *Censo 2002*. REDATAM. URL: www.INE.cl.

Law, J., Boyle, J., Harris, F., Harkness, A. & Nye, C. (2000) Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *Developmental Medicine and Child Neurology* **42**,190–200.

Machado, T. M., Bomfim, T. F., Souza, L. V., Soares, N., Santos, F. L., Acosta, A. X. & Abe-Sandes, K. (2013a) Types of marriages, population structure and genetic disease. *Journal of Biosocial Science* **45**, 461–470.

Machado, T. M., Bomfim, T. F., Souza, L. V., Soares, N., Santos, F. L., Acosta, A. X. & Abe-Sandes, K. (2013b) Types of marriages, population structure and genetic disease – Erratum. *Journal of Biosocial Science* **45**, 575.

Orellana, M., Medina, A., Morel, T., Ruh, M., Hernández, R. & Monleón, J. (1975) *Las islas de Juan Fernández*. Depto de Ciencias Antropológicas y Arqueológicas, University of Chile, Santiago.

Patton, M. A. (2005) Genetics studies in the Amish community. *Annals of Human Biology* **32**, 163–167.

Quezada, M. & Barrantes, R. (1973) Análisis de la endogamia en una población chilena. *Revista Médica de Chile* **101**, 227–230.

Rothhammer, R. & Llop, E. (2004) *Poblaciones chilenas cuatro décadas de investigaciones bioantropológicas*. Editorial Universitaria, Santiago.

Schnyder, J. O. (2004) *Juan Fernández – El tesoro de Chile.* Editorial Cumberland, Santiago.

Serre, J. L. & Babron, M. C. (1985) Polymorphism and genetic evolution in an isolate in the Antilles: Saint-Barthelemy. *Annals of Human Biology* **12**, 413–419.

Shriberg, L. D., Tomblin, J. B. & McSweeny, J. L. (1999) Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language and Hearing Research* **42**, 1461–1481.

Tomblin, J. B., Records, N. L., Buckwalter, P. R., Zhang, X., Smith, E. & O'Brien, M. (1997a) Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language and Hearing Research* **40**, 1245–1260.

Tomblin, J. B., Smith, E. & Zhang, X. (1997b) Epidemiology of specific language impairment: prenatal and perinatal risk factors. *Journal of Communication Disorders* **30**, 325–344.

Varela, T. A., Ainsua, R. L. & Fariña, J. (2001) Evolution of consanguinity in Bishopric of Lugo (Spain) from 1900 to 1979. *Annals of Human Biology* **28**, 575–588.

Vicuña Mackenna, B. (1883) *Juan Fernández. Historia verdadera de la isla Robinson Crusoe*. Rafael Jover, Santiago.

Villanueva, P., De Barbieri, Z., Palomino, H. M. & Palomino, H. (2008) High prevalence of specific language impairment in Robinson Crusoe Island. A possible founder effect. *Revista Médica de Chile* **136**, 186–192.

Villanueva, P., Jara, L. & Palomino, H. (2010) Association of D16S515 microsatellite with specific language impairment on Robinson Crusoe Island, an isolated Chilean Population: a possible key to understanding language development. *Human Biology* **82**, 395–408.

Villarroel, H., Lazo, B., Campusano, C., Pinto-Cisternas, J. & Ballesteros, S. (1978) Consanguinity and migration in Viña del Mar, Chile 1888–1967. *Revista Médica de Chile* **108**, 407–441.

Villanueva, P., Newbury, D. F., Jara, L., De Barbieri, Z., Mira, G., Palomino, H. M., Fernández, M. A. *et al.* (2011) Genome wide analysis of genetic susceptibility to language impairment in a isolated Chilean population. *European Journal of Human Genetics* **19**, 687–695.

Wang, W., Cong Qian & Bittles, A. H. (2002) Consanguineous marriage in PR China: a study in rural Man (Manchu) communities. *Annals of Human Biology* **29**, 685–690.

Zuñiga, J. (1978) Inbreeding in the Elqui Valley's hydrographic basin (1900–1975). *Revista Médica de Chile* **106**, 303–312.

PLOS | GENETICS

RESEARCH ARTICLE

Exome Sequencing in an Admixed Isolated Population Indicates *NFXL1* Variants Confer a Risk for Specific Language Impairment

Pía Villanueva[1,2,3,4‡*], Ron Nudel[5‡], Alexander Hoischen[6], María Angélica Fernández[2], Nuala H. Simpson[5], Christian Gilissen[6], Rose H. Reader[5], Lillian Jara[1], Maria Magdalena Echeverry[7], Clyde Francks[8,9], Gillian Baird[10¶], Gina Conti-Ramsden[11¶], Anne O'Hare[12¶], Patrick F. Bolton[13¶], Elizabeth R. Hennessy[14¶], the SLI Consortium[¶], Hernán Palomino[3], Luis Carvajal-Carmona[7,15], Joris A. Veltman[6], Jean-Baptiste Cazier[16,17], Zulema De Barbieri[2], Simon E. Fisher[8,9☯¶], Dianne F. Newbury[5,18☯¶*]

1 Human Genetics Program, Institute of Biomedical Sciences (ICBM), Faculty of Medicine, University of Chile, Santiago, Chile, **2** School of Speech and Hearing Therapy, Faculty of Medicine, University of Chile, Santiago, Chile, **3** Department of Child and Dental Maxillary Orthopedics, Faculty of Dentistry, University of Chile, Santiago, Chile, **4** Doctoral Program of Psychology, Graduate School, University of Granada, Granada, Spain, **5** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **6** Department of Human Genetics, Radboud Institute for Molecular Life Sciences and Donders Centre for Neuroscience, Radboud University Medical Center, Nijmegen, the Netherlands, **7** Grupo de Citogenetica, Filogenia y Evolucion de las Poblaciones, Facultades de Ciencias y de Ciencias de la Salud, Universidad del Tolima, Ibague, Colombia, **8** Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, **9** Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands, **10** Newcomen Centre, the Evelina Children's Hospital, London, United Kingdom, **11** School of Psychological Sciences, University of Manchester, Manchester, United Kingdom, **12** Department of Reproductive and Developmental Sciences, University of Edinburgh, Edinburgh, United Kingdom, **13** Departments of Child & Adolescent Psychiatry & Social Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom, **14** University Child Health and DMDE, University of Aberdeen, Aberdeen, United Kingdom, **15** UC Davis Genome Center, Department of Biochemistry and Molecular Medicine, School of Medicine, University of California Davis, Davis, California, United States of America, **16** Department of Oncology, University of Oxford, Oxford, United Kingdom, **17** Centre for Computational Biology, University of Birmingham, Edgbaston, United Kingdom, **18** St Johns College, University of Oxford, Oxford, United Kingdom

☯ These authors contributed equally to this work.
‡ These authors contributed equally to this work.
¶ On behalf of the SLI Consortium. Membership of the SLI consortium is listed in acknowledgments.
* piavillahk@uchile.cl (PV, linguistic and population queries); dianne@well.ox.ac.uk (DFN, genetic queries)

Citation: Villanueva P, Nudel R, Hoischen A, Fernández MA, Simpson NH, Gilissen C, et al. (2015) Exome Sequencing in an Admixed Isolated Population Indicates *NFXL1* Variants Confer a Risk for Specific Language Impairment. PLoS Genet 11(3): e1004925. doi:10.1371/journal.pgen.1004925

Editor: Brett S. Abrahams, UCLA, United States of America

Received: June 3, 2014

Accepted: November 25, 2014

Published: March 17, 2015

Data Availability Statement: Data used in this paper come from a small and well-defined population. To protect the identity of individuals, these confidential data are not publically available. Data are available from the University of Chile Ethics Committee for researchers who meet the criteria for access to confidential data.

Funding: DFN and the work of the Newbury lab are funded by an MRC Career Development Fellow and a Junior Research Fellow at St John's College, University of Oxford. The work of the Newbury lab is funded by the Medical Research Council [G1000569/

Abstract

Children affected by Specific Language Impairment (SLI) fail to acquire age appropriate language skills despite adequate intelligence and opportunity. SLI is highly heritable, but the understanding of underlying genetic mechanisms has proved challenging. In this study, we use molecular genetic techniques to investigate an admixed isolated founder population from the Robinson Crusoe Island (Chile), who are affected by a high incidence of SLI, increasing the power to discover contributory genetic factors. We utilize exome sequencing in selected individuals from this population to identify eight coding variants that are of putative significance. We then apply association analyses across the wider population to highlight a single rare coding variant (rs144169475, Minor Allele Frequency of 4.1% in admixed South American populations) in the *NFXL1* gene that confers a nonsynonymous change (N150K)

1]. The Robinson Crusoe project is funded by the Medical Research Council [MR/J003719/1]. The collection of DNA samples and characterisation of the Robinson Crusoe population was funded by Vicerrectoría de Investigación, Universidad de Chile (www.uchile.cl), UCHILE DID TNAC 01-02/01, UCHILE DI MULT 05-05/02 grants. RN is funded by a University of Oxford Nuffield Department of Medicine Prize Studentship. SEF and CF are supported by the Max Planck Society, who also funded the exome sequencing. LCC and MME receive funding from the European Union FP7 CHIBCHA Consortium, GSK Oncology (Ethnic Research Initiative), Colciencias, Cancer Research UK and Universidad del Tolima. LCC receives funding from the University of California Davis, The V Foundation for Cancer Research, and The National Institute On Aging (award number P30AG043097) and The National Cancer Institute (Award number K12CA138464) of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The collection of the SLIC samples was supported by the Wellcome Trust (060774and 076566). PFB is supported by a National Institute of Health Research (UK) Senior Investigator award and the Biomedical Research Centre in Mental Health at the South London & Maudsley NHS Trust Hospital, London. The High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics is funded by the Wellcome Trust [090532/Z/09/Z] and the MRC [G0900747 91070]. 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.

and is significantly associated with language impairment in the Robinson Crusoe population ($p = 2.04 \times 10–4$, 8 variants tested). Subsequent sequencing of NFXL1 in 117 UK SLI cases identified four individuals with heterozygous variants predicted to be of functional consequence. We conclude that coding variants within NFXL1 confer an increased risk of SLI within a complex genetic model.

Author Summary

Children affected by Specific Language Impairment (SLI) have unexpected problems learning to talk and understand language, despite developing normally in all other areas. This disorder runs in families but we do not understand how the genetic contributions work, or which genetic mechanisms might be important. In this paper, we study a Chilean population who are affected by a high incidence of SLI. Such populations may provide increased power to discover contributory genetic factors, under appropriate conditions. We identify a genetic change in the population that causes a change to a protein called NFXL1. This change is usually very rare but is found at a higher frequency than expected in our population, particularly in those people affected by SLI. We then looked at this gene in over 100 individuals from the UK affected by SLI and found four more changes that probably affect the protein. This is a higher number than we would expect by chance. We therefore propose that the *NFXL1* gene and the protein it encodes might be important in risk of SLI.

Introduction

Language deficits form a central feature of many developmental disorders and account for a high number of pediatric referrals and statements of special educational need [1]. These language impairments often represent a secondary clinical feature of a more pertinent developmental disability such as Down syndrome, Autistic Spectrum Disorder or intellectual disability. However, in a proportion of cases, the primary clinical concern is the language difficulties, which occur in the absence of any other developmental deficit or neurological impairment and in the presence of normal non-verbal IQ. In such cases, the diagnosis is Specific Language Impairment (SLI) [2].

SLI affects between 5 and 7% of children in the UK [3] and significantly more boys than girls [4]. The disorder is highly heritable [5] but genetic contributions are expected to be complex in nature with significant heterogeneity between individuals [6]. Common risk variants within *ATP2C2* (OMIM#613082), *CMIP* (OMIM#610112) [7], *ABCC13* (OMIM#608835) [8], *FLNC* (OMIM#102565), *RBFOX2* (OMIM#612149) [9] and *ROBO2* (OMIM#602431) [10] have been associated with quantitative measures of language skills. Genome-wide association studies of language-impaired probands have also highlighted potential risk variants in *NDST4* (OMIM#615039), *ZNF385D, COL4A2* (OMIM#120090) *[11]* and *NOP9 [12]*. Other studies implicate rare genetic events which may have higher penetrance [13,14]. However, it is clear that the contributions of these various genetic effects are complex. Some may be specific to individuals with certain forms of language deficits, others may contribute across the range of ability [7,8,11,15,16]. The functional impact of these candidate genes has yet to be elucidated and further candidates need to be identified before we can properly understand the molecular pathways underlying SLI.

Clearer links have been made between the presence of language deficits and disruption of the *FOXP2* gene (OMIM#605317), a forkhead/winged-helix transcription factor [17,18]. Reduced functional dosage of *FOXP2*, caused by mutation or chromosomal rearrangements, leads to characteristic deficits in coordinating sequences of orofacial movements, impairing speech, producing a disorder known as developmental verbal dyspraxia (DVD) or childhood apraxia of speech (CAS) [18–22]. Typically the DVD/CAS features of *FOXP2* mutation cases are accompanied by wide-ranging problems with spoken and written language [23]. Whilst *FOXP2* disruptions are rare and account for only a small proportion of DVD/CAS cases, the investigation of this gene, its expression patterns and interactions, have led to the elucidation of genetic networks that are important to language development and contribute to more common forms of language impairment [23–25]. One of the transcriptional targets of FOXP2 is *CNTNAP2* (OMIM#604569), a member of the neurexin family which mediates interactions between neurons and glia during nervous system development [26]. Genetic variation across *CNTNAP2* has been associated both with language deficits [15,27–29] and language ability in the general population [30–32]. Variations in, and disruptions of, this gene have also been implicated across a range of neurodevelopmental disorders such as autism, epilepsy and schizophrenia [26], indicating that it is likely to be crucial for brain development. These investigations demonstrate how the identification of genetic mutations underlying a distinct severe form of disorder provide entry points into mechanisms that are relevant to the wider processes underlying the initial deficit.

In 2008, Villanueva et al described a population who are affected by an unusually high prevalence of language impairment [33]. This admixed population inhabits the Robinson Crusoe Island which forms part of the Juan Fernandez Archipelago in the South Pacific Ocean, approximately 400 miles off the coast of Chile. The Island was last colonized in 1876 by 64 individuals of European and South American descent. In the 2002 census, the Island population was 633, the majority of whom were descendants of the founder families. More than 70% of the current population has a surname from the colonizing families and 14% of marriages involve consanguineous unions [34]. In their 2008 study, Villanueva et al completed psychometric profiling of 66 island children aged between 3 and 9 years of age, of whom 40 were descendants of the founder party. They found that 35% of the founder-related children (14 of 40) were affected by specific language impairment. No evidence for a male bias was observed in this group. A further 27.5% of the founder-related child population (11 of 40) had language abilities below that expected for their age but presented with additional developmental concerns or low non-verbal IQ, precluding a diagnosis of SLI. The remaining 37.5% of founder-related children (15 of 40) had typical language development. In contrast, only one of 26 children whose parents are not related to the founder families (3.8%) had evidence of language impairment, a frequency of language impairment that coincided with that seen in mainland Chile (3%) [33]. Furthermore, when the genealogical records of the islanders were recompiled, 90% of the individuals affected by SLI were direct descendants of a single pair of founder brothers who formed part of the founder party [33,35]. Given the clear phenotypic differences between founder-related and non-founder-related children on the Island, we postulated that the founder brothers may have carried a rare causative genetic mutation or, alternatively, combinations of common genetic variations that together confer a high risk of language impairment. A previous genome-wide linkage study of 34 families from the Robinson Crusoe Island identified significant linkage to several chromosome regions, the most consistent of which included a large section (48Mb) of chromosome 7q (SLI4 – OMIM#612514) that included many genes which represent good candidates for language impairment, including *FOXP2* and *CNTNAP2* [35]. However, in depth genomic profiling has yet to be performed within this population.

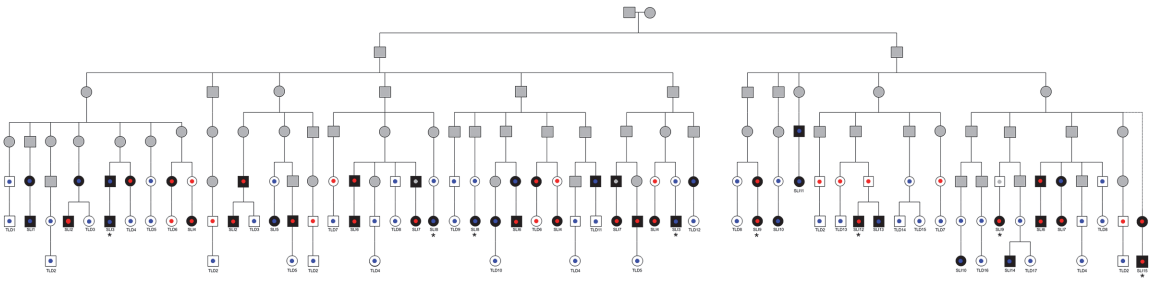


Fig 1. Pedigree showing direct lines of descent between founder brothers and children in Robinson Crusoe validation cohort. Founder brothers are individuals on the second line of the pedigree. Individuals with language impairment are colored in black. Individuals with typical language are denoted in white. Individuals with unknown phenotype are shaded grey. Genotypes at rs144169475 are represented by small circles; blue circles represent homozygote reference allele, red circles represent variant carriers, grey circles represent unknown genotype. Note that each individual may be represented through multiple lines of descent and so might appear more than once on this diagram. Children are labelled according to affection status – SLI1 to SLI15 and TLD1 to TLD17. Cases whose exomes were sequenced are indicated by asterisks. Three children (1 affected, 2 unaffected, none of whom carried the rs144169475 variant) are not represented on this figure since they were related to alternative founder families. SLI15 is known to be related to one of the founder brothers but the exact line of descent is unknown.

doi:10.1371/journal.pgen.1004925.g001

In this study, we make use of this admixed isolated population and assess the possibility of a founder mutation, by completing exome sequencing of five individuals from the Robinson Crusoe population affected by SLI. We substantiate the findings of the exome screen by performing association analyses of selected putative functional variants in the wider Robinson Crusoe population. The contribution of identified risk variants is subsequently validated by performing targeted sequencing of candidate genes in a UK-based cohort of individuals affected by SLI.

Results

We selected five related individuals with SLI from the Robinson Crusoe cohort for exome sequencing (Fig. 1). From the exome sequence data, we selected all novel variants (i.e. not reported in publically available or in-house databases) that caused nonsynonymous changes or changes to canonical splice sites and were shared by at least three of the five individuals sequenced. A flow diagram of our methodology can be found in S1 Fig.. All such variants were subsequently genotyped in 111 founder-related cases and controls from the Robinson Crusoe Island (Robinson Crusoe validation cohort) and tested for association to language impairment using a method that takes into account familial relationships. To substantiate the findings of the exome screen and association analyses, we then went on to sequence the coding regions of candidate genes implicated from these investigations in an independent cohort of 117 British children affected by SLI (SLIC cohort).

Exome sequencing

On average, 47,276 (median = 49,543, range = 43,075–50,112) genic variants were identified in each of the five exomes. This included an average of 17,405 (median = 17,326, range = 15,200–19,837) exonic variants, 8,379 (median = 8,089, range = 7,258–9,629) missense variants and 106 (median = 90, range = 72–157) nonsense (including indels) variants per individual. Across all five samples, 90.0% of targeted exome sequencing had coverage of at least 10-fold. The average coverage of targeted sequence was 56.5-fold and 21% of the reads reached this level. Sequencing metrics can be found in S1 Table. To test the hypothesis that the founder brothers carried a rare causative genetic mutation, we focused upon novel variants that caused nonsynonymous protein substitutions or altered canonical splice sites for our downstream analyses.

Comparisons between individuals found that no such variants were shared by all five individuals. However, allowing for potential genetic heterogeneity between affected individuals, we identified nine novel nonsynonymous or splice-site variants that were shared by at least 3 of the 5 children sequenced (Table 1). Eight novel nonsynonymous or splice-site variants were validated in the five exome samples by Sanger sequencing. None of these variants overlapped with the regions of suggestive linkage ($P<7.3\times10^{-4}$, chromosomes 2, 6, 7, 8, 9, 12, 13 and 17, as listed in S2 Table) previously identified in this population [35]. S3 Table provides a full list of all shared, high-quality variants that fell within the previously identified regions of linkage. All of these had previously been reported in dbSNP (138) and many were non-genic, intronic or synonymous (see notes column in S3 Table).

Association analyses of key variants in Robinson Crusoe validation cohort

All shared novel nonsynonymous or splice-site variants identified in the exome screen were subsequently genotyped in 111 members of the Robinson Crusoe population (49 individuals with language-impairment and 62 individuals with typical language ability). This validation cohort was ascertained via 35 children living on the Robinson Crusoe Island who had been diagnosed with SLI or who showed typical language development (as described in methods) and included the five children used in the exome sequencing. All children were descendants of the founder families of the Robinson Crusoe Island and, as such, the cases and controls used in these association analyses were inter-related (Fig. 1). We therefore employed an association algorithm that allowed for relatedness between cases (MQLS, [36]), and that took into account the shared ancestry of the Robinson Crusoe validation cohort (288 individuals over 5 generations). These analyses highlighted one particular coding variant (chr4:g.47,907,320A>T, hg19) that was present at a significantly higher frequency in Islanders with language impairment than in Islanders with typical language ability (Table 1). Thirty nine percent of Islanders with language impairment were found to carry this variant compared to ten percent of Islanders

Table 1. Association of novel nonsynonymous or canonical splice-site variants in 111 individuals from the Robinson Crusoe validation cohort.

Chr	Variant Position (hg19)	Ref/ variant	Average read depth across variant	Gene	Transcript ID	Gene element affected by variant	Amino Acid change	SLI/ TLD[1]	Variant Freq[2]	SLI variant freq[3]	TLD variant freq[4]	MQLS p
1	113,245,326	A/G	60	*RHOC*	NM_001042678	IVS3	SA site	49/62	0.059	0.071	0.048	0.625
1	248,308,783*	T/A	415	*OR2M5*	NM_001004690	Exon 1	C112S	49/62	0.000	0.000	0.000	-
4	47,907,320	A/T	57	*NFXL1*	NM_152995	Exon 4	N150K	49/62	0.113	0.194	0.048	0.0002
10	31,134,425	C/T	119	*ZNF438*	NM_001143766	Exon 8	R641H	49/62	0.158	0.173	0.145	0.466
11	33,054,503	T/G	36	*DEPDC7*	NM_139160	Exon 8	N444K	40/60	0.131	0.149	0.117	0.399
16	27,363,901	G/A	30	*IL4R*	NM_000418	Exon 7	R185H	49/61	0.095	0.143	0.057	0.053
21	47,359,924	C/T	52	*PCBP3*	NM_001130141	IVS-12	SA site	48/59	0.266	0.292	0.246	0.228
22	41,257,834	T/TA	37	*DNAJB7*	NM_145174	Exon 1	V55VX	49/62	0.261	0.245	0.274	0.554
X	48,682,972	A/G	30	*HDAC6*	NM_006044	Exon 29	N1200D	49/62	0.419	0.378	0.452	0.456

1 – The number of individuals with SLI genotyped / the number of individuals with typical language ability genotyped.

2 – Frequency of discovered variant in all genotyped Islanders

3 – Frequency of discovered variant in genotyped Islanders with SLI

4 – Frequency of discovered variant in genotyped Islanders with typical language ability

Note that all Islanders (both cases and controls) were related

*- this variant was not validated with Sanger sequencing and represents a false positive finding from the exome sequencing

doi:10.1371/journal.pgen.1004925.t001

with typical language skills ($p = 2.04 \times 10^{-4}$) (Table 1). Across the Robinson Crusoe validation cohort, the minor allele frequency was 11.3% (25 of 222 chromosomes sampled) (Table 1).

Predicted functional effects of chr4:g.47,907,320A>T

Chr4:g.47,907,320A>T (hg19) falls in exon 4 of the *Homo sapiens* nuclear transcription factor, X-box binding-like 1 (*NFXL1*) gene (Fig. 2). The variant causes a nonsynonymous change yielding an asparagine to lysine substitution in the encoded protein (p.N150K, uncharged amino acid to positively charged amino acid). This change is predicted to be "disease-causing" by MutationTaster with a confidence probability of 0.98 (SIFT = 0.67, PolyPhen-2 = 0.178). The position is conserved at both the amino acid and nucleotide level (PhyloP = 0.66, PhastCons = 1); the amino acid N150 is invariant across 36 of the 38 vertebrate species in which an alignment could be made and the thymine nucleotide at this position is conserved across all six ENSEMBL primate species investigated (Human, chimp, gorilla, orangutan, macque and marmoset) (Fig. 2).

Chr4:g.47,907,320A>T, hg19 in control populations

The variant at chr4:47,907,320 was not observed in 127 independent European population controls that were genotyped (Table 2). We therefore went on to genotype an additional 320 independent individuals from a Colombian population cohort and 121 independent individuals from a Chilean control population cohort. In these cohorts, the variant was present with a minor allele frequency of 4.2% (27 of 640 chromosome sampled) and 7.4% (18 of 242 chromosome sampled) respectively (Table 2). Subsequent data released by the 1000 genomes project confirmed that this variant is specific to admixed American populations (AMR) with an

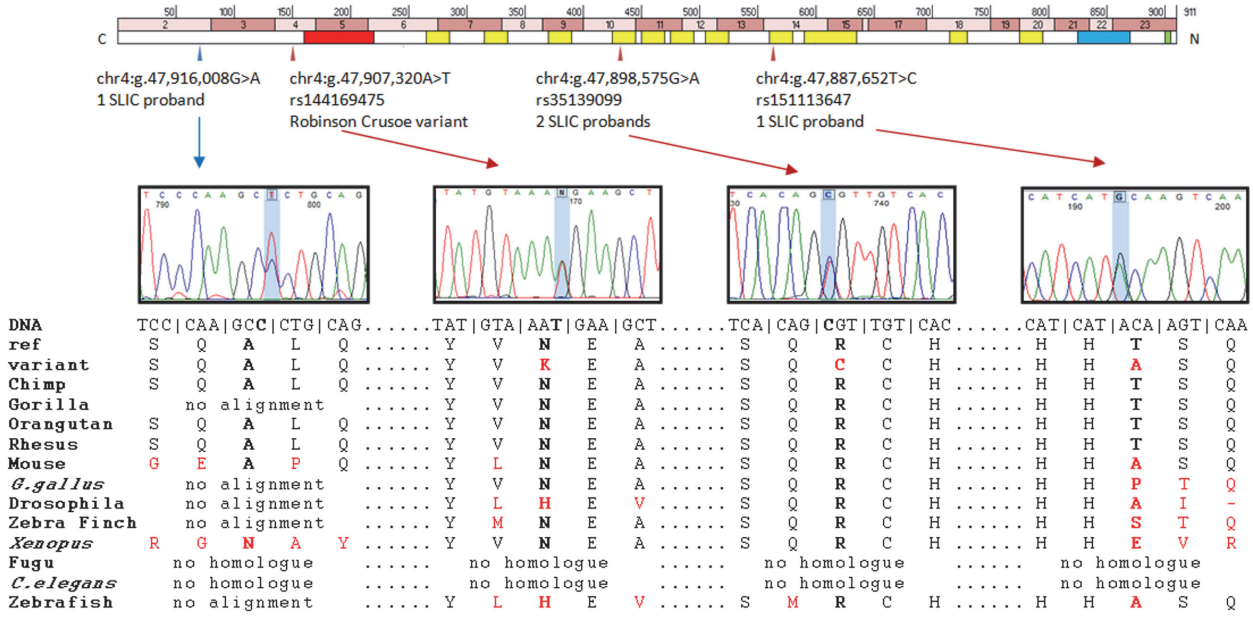


Fig 2. Putative contributory coding variants identified in *NFXL1* by this study. Position of putative *NFXL1* coding variants with respect to exons and protein coding sequence. Genomic coding exons (exons 2–23) are shown by pink bands at the top. Protein motifs are represented by colored bands in the lower boxes. The red box represents a Znf RING motif, the yellow boxes represent Znf NFX1 motifs, the blue box represents a coiled-coil domain and the green box a transmembrane domain. Putative contributory coding variants are shown by arrows. Blue arrows denote synonymous changes, red arrows nonsynonymous changes. Sanger sequencing plots are given for all variants identified. Conservation of amino acid sequences across 11 species shown for all variants identified. The ref row shows the human reference allele and the variant row shows the observed variant in our samples. All sequences that differ from the reference sequence are shown in red.

doi:10.1371/journal.pgen.1004925.g002

average minor allele frequency of 4.1%. In the sub-populations of the AMR grouping, the minor allele frequency is reported as 0.9% in Puerto Ricans (PUR – 1 in 110 chromosomes sampled), 3.3% in Colombians in Medellin (CLM – 4 in 120 chromosomes sampled) and 7.6% in individuals of Mexican ancestry in Los Angeles (MXL – 10 of 132 chromosomes sampled) (Table 2). The variant has recently been designated as rs144169475 accordingly.

Linkage analyses of chromosome 4 (46–49Mb)

Parametric and nonparametric linkage analyses were performed for 55 SNPs across the *NFXL1* region of chromosome 4 (46–49Mb, hg19) within seven extended pedigrees from the Robinson Crusoe validation cohort (S2 Fig.). In these analyses, we did not observe evidence of linkage (maximum LOD score = 0.62, S3 Fig.).

Sequencing of *NFXL1* in a language-impaired cohort (SLIC)

We sequenced the entire coding region of the *NFXL1* gene in 117 unrelated probands affected by SLI (from the UK SLI Consortium (SLIC) cohort [7,37–39]), to assess whether we could replicate a role for *NFXL1* in SLI etiology. In total, we identified 166 high-quality sequence variants across the *NFXL1* gene. 155 of the variants detected were intronic, 4 were in the 3'UTR and 7 affected the coding region. Of the coding variants, three were nonsynonymous and four were synonymous substitutions (Table 3).

Nonsynonymous variants and those with estimated allele frequencies of <5% were verified across all the pools of DNA in which they were observed using Sanger sequencing. This allowed the derivation of accurate allele frequencies within the SLIC cohort.

One of the synonymous variants (chr4:g.47,916,008G>A, hg19) was found in a heterozygous state in one SLIC proband (allele frequency of 0.43%) but had not been documented in any European individuals in the 1000 genomes project [40] or the NHLBI GO ESP Exome Variant Server (EVS), which together consist of data from 4679 control individuals and therefore have the ability to detect rare variants with a population frequency of 0.0001. A comparison of allele frequencies between SLIC probands (1 of 234 chromosomes tested) and controls (0 of 9358 chromosomes tested) yielded a significant *P*-value of 0.0244. Intriguingly, although it is synonymous, this variant was predicted to be "disease-causing" by MutationTaster with a confidence probability of 0.98 (SIFT = 1.0). This variant falls in the most 5' coding exon of *NFXL1* and is part of a CpG island, indicating that it may be important for the regulation of gene expression. Furthermore, ENCODE data shows that it is part of a H3K4Me3 mark (which is often associated with promoters) and binds multiple transcription factors, particularly POLR2A c-MYC and PHF8 (www.genome.ucsc.edu, accessed April 2014).

The remaining three synonymous variants (rs2053404, rs6818556 and rs35139099) found in SLIC probands were also found at similar frequencies in control databases. All had allele frequencies of >5% and are therefore thought to represent common polymorphisms (Table 3).

One nonsynonymous substitution (chr4:g.47,887,652T>C, hg19 – rs151113647) was found in a heterozygous state in a single SLIC proband (allele frequency of 0.43%) and again, was not observed in 4679 independent European individuals in the control public databases (Table 3), yielding a significant *P*-value of 0.024 (1 of 234 SLIC chromosomes tested vs 0 of 9358 control chromosomes tested). Further investigations found that this variant had been observed in a heterozygous state in a single African American individual from the EVS. Principal components analysis of genome-wide SNP data in the SLIC proband against the hapmap-3 populations did not detect any African ancestry. The rarity of the rs151113647 variant and its position within a zinc-finger motif (Fig. 2) indicates that it may confer negative effects upon protein function. Nonetheless, because the nucleotide is not highly conserved across species (phylo

Table 2. Allele and genotype frequencies of rs144169475 in the Robinson Crusoe validation cohort.

	Robinson Crusoe population	Founder-related Islanders[1]	Non-founder-related Islanders[2]	SLI[3]	TLD[4]	Male Islanders[5]	Female Islanders[6]	European controls[7]	Colombian Controls[8]	Chilean Controls[9]	PUR[10]	CLM[11]	MXL[12]
freq allele T (variant)	0.113	0.125	0.000	0.194	0.048	0.116	0.132	0.000	0.042	0.074	0.009	0.033	0.076
freq genotype TT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
freq genotype AT	0.225	0.250	0.000	0.388	0.097	0.233	0.263	0.000	0.084	0.149	0.018	0.067	0.152
freq genotype AA	0.775	0.750	1.000	0.612	0.903	0.767	0.737	1.000	0.916	0.851	0.982	0.933	0.848
No. individuals	111	100	11	49	62	43	57	127	320	121	55	60	66

1 – Islanders who are directly related to one of the eight founder families (NB this sample includes affected andunaffected individuals)

2 – Individuals who live on the Island but have no known genetic connection to the eight founder families (NB this sample includes 4 affected and 7 unaffected individuals)

3 – Islanders who have been diagnosed with SLI as described in methods (NB, this sample included 45 related, founder-related individuals and 4 non-founder-related parents).

4 – Islanders who have been classified as having typical language ability as described in methods (NB, this sample included 55 founder-related Islanders and 7 non-founder-related parents).

5 – Male individuals who are directly related to one of the eight founder families (NB this sample includes affected and unaffected individuals)

6 – Female individuals who are directly related to one of the eight founder families (NB this sample includes affected and unaffected individuals)

7 - 127 in-house European controls (ECACC, HRC-1 DNA Panel)

8 - 320 South American (Colombian) controls (provided by Luis Carvajal-Carmona and Maria Magdalena Echeverry)

9 – 121 Chilean controls (provided by Lillian Jara and Pia Villanueva)

10 – 1000 genomes Puerto Ricans from Puerto Rico (Integrated phase I, accessed March 2014)

11 – 1000 genomes Colombians from Medellin, Colombia (Integrated phase I, accessed March 2014)

12 – 1000 genomes Mexican Ancestry from Los Angeles USA (Integrated phase I, accessed March 2014)

doi:10.1371/journal.pgen.1004925.t002

Table 3. *NFXL1* coding variants observed in 117 UK (SLIC) probands affected by SLI.

Position (hg19)	Ref	Var	Estimated VAF in SLI probands[1]	Median read depth[2]	dbSNP ID	1000G population VAF (ALL:AFR:AMR:ASN:EUR)[3]	EVS VAF (EA:AA)[4]	European VAF[5]	Confirmed VAF in SLI probands[6]	*NFXL1* Exon	Amino Acid change[7]	Fishers exact between European controls & SLIC[8]
Chr4: g.47887536	T	C	0.7763	4531	rs2053404	0.73:0.65:0.75:0.72:0.77	0.75:0.68	0.75	NT	14	A601A	NT
Chr4: g.47887652	T	C	0.0035	5123	rs151113647	0.00:0.00:0.00:0.00:0.00	0.00:0.0002	0.00	0.0043	14	T563A	0.0244
Chr4: g.47887991	G	A	0.7835	6433.5	rs6818556	0.73:0.65:0.75:0.72:0.77	0.75:0.68	0.75	NT	13	T523T	NT
Chr4: g.47898575	G	A	0.0071	4817.5	rs35139099	0.0005:0.00:0.00:0.00:0.0013	0.005:0.0005	0.004	0.0085	10	R432C	ns
Chr4: g.47901088	C	T	0.0642	2986	rs34323060	0.02:0.002:0.03:0.00:0.04	0.047:0.0098	0.05	NT	7	K292K	NT
Chr4: g.47901476	G	A	0.3195	1212.5	rs12651301	0.42:0.33:0.35:0.63:0.35	0.31:0.33	0.31	NT	6	P246L	NT
Chr4: g.47916008	G	A	0.0071	2576.5	NA	0.00:0.00:0.00:0.00:0.00	0.00:0.00	0.00	0.0043	2	A71A	0.0244

1 – Variant allele freq (VAF) in 117 UK SLIC probands is estimated by Syzygy using the proportion of reads that have the variant

2 – Median read depth for given base across all pools

3 - Variant allele frequency (VAF) in 1000 genomes super-populations (Integrated phase I data, accessed March 2014). ALL – all 1000 genomes populations combined (No. alleles $\sim$ 2184), AFR – African populations (YRI, LWK, GWD, MSL, ESN, ASW & ACB, No. chromosomes = 492), AMR – Ad mixed Americans (MXL, PUR, CLM, PEL, No. chromosomes = 362),ASN – East Asian (CHB, JPT, CHS, CDX & KHV, No. chromosomes = 572), EUR-European (TSI, FIN, GBR, IBS, no. chromosomes = 758).

4 – Exome Sequencing Project (ESP) variant allele frequency (VAF). EA – European Americans (no. chromosomes = 8600), AA – African Americans (no. chromosomes = 4268).

5 – Combined variant allele frequency across European controls from 1000 genomes and EVS (no. chromosomes = 9358)

6 – Allele frequency in SLI probands after confirmatory Sanger sequencing (no. chromosomes = 234)

7 – Amino acid change conferred by given sequence variant in protein NP_694540.3. If the change occurs within a conserved motif, this is noted.

8 – Fisher's exact test for differences in allele frequencies between EVS European Americans and SLIC probands. ns = non-significant $P<0.05$

NT = not tested

Ns = not significant

doi:10.1371/journal.pgen.1004925.t003

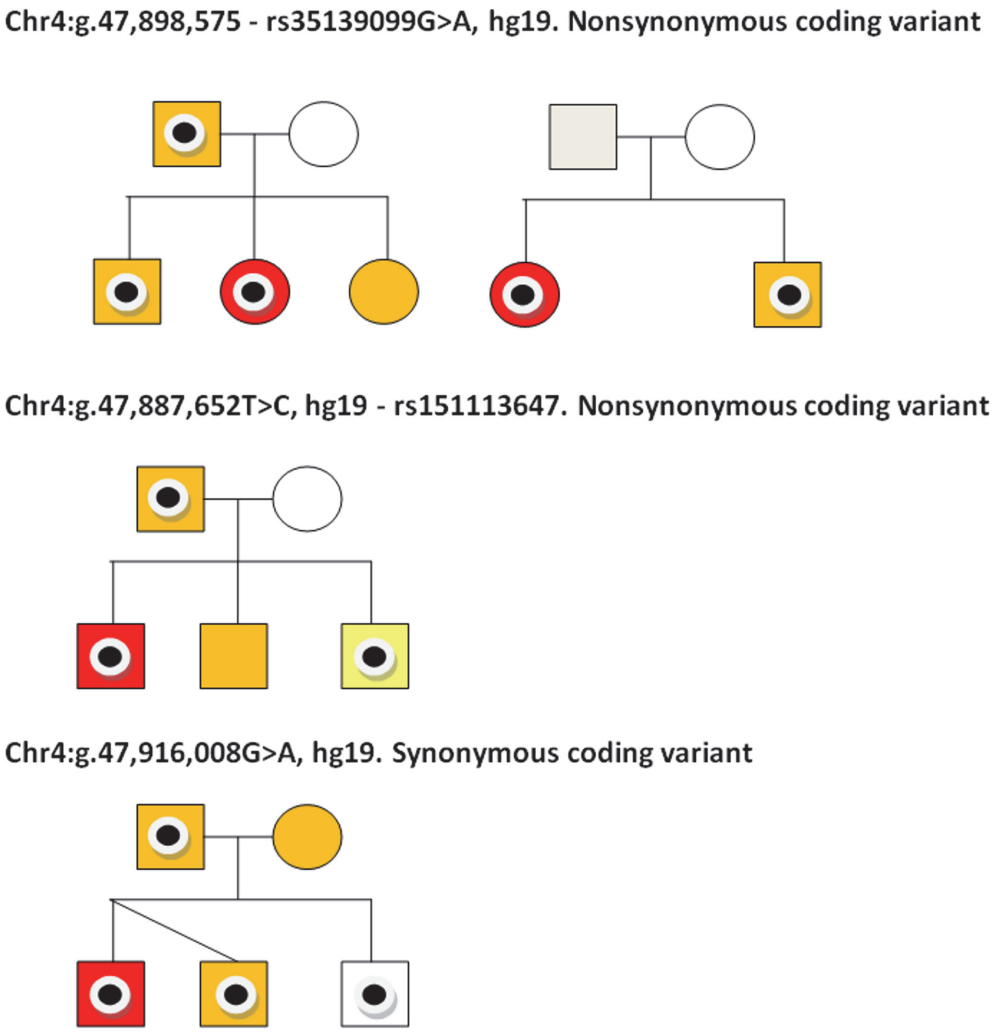


Fig 3. Coding variants observed in SLIC probands and their families. Pedigrees are shown for nuclear families of SLIC individuals carrying three coding variations in *NFXL1*. Individuals carrying the variants are identified with a black circle. Sequencing traces of each variant is shown. SLIC probands are colored in red and other family members with SLI (defined as expressive and/or receptive language skills >1.5SD below that expected for their age) are colored in orange. In pedigree 3 (rs151113647), the youngest sibling (colored in yellow) did not meet the criteria for SLI but had expressive and receptive language scores ~1SD below that expected for his age. Individuals with no shading have typical language ability. DNA was not available for individuals colored in grey.

doi:10.1371/journal.pgen.1004925.g003

P = −0.418, phastCons = 0.925), the change was predicted to be a polymorphism by MutationTaster with a confidence probability of 0.99 (SIFT = 0.68, polyphen-2 = 0.00) (Fig. 2).

A second nonsynonymous substitution (chr4:47,898,575G>A, hg19 - rs35139099) was observed in a heterozygous state in two independent SLIC probands (allele frequency of 0.85%). This variant was also found in 44 of 4679 independent European control individuals from public databases (allele frequency of 0.47%, Table 3) yielding a P value of 0.3097. Although, it was not observed to occur at a significantly increased frequency in SLIC probands, the rs35139099 variant occurs at a conserved residue (phyloP = 1.466, phastCons = 1) within a zinc-finger motif (Fig. 2) and is therefore predicted to be damaging by MutationTaster with a confidence probability of 0.99 (SIFT = 0.00, Polyphen-2 = 1.00) (Fig. 2).

The remaining nonsynonymous variant (chr4:g.47,901,476G>A, hg19 - rs12651301) was observed to occur across all the sequence pools with an estimated allele frequency of 32% (Table 3). This common variant was also observed in independent European controls from public databases with a frequency of 31% (Table 3) and falls outside of any protein motifs and is thus likely to represent a polymorphism.

The three rare variants identified (rs151113647, rs35139099 and chr4:g.47,916,008G>A, hg19) were sequenced in all available family members of the SLIC proband in whom they were observed (Fig. 3). The chr4:g.47,916,008 variant was inherited from an affected father by two affected children and one child with typical language development (Fig. 3). The rs151113647 variant was inherited from a father, who reports a history of language and literacy problems, by the proband, who attends a special language unit, and his sibling, who also has SLI. The middle child in this family, who also showed evidence of expressive and receptive language deficits, did not inherit the variant (Fig. 3). Two SLIC families carried the rs35139099 variant; in the first, the variant is present in the father, who self-reports a history of dyspraxia, and passed onto both the proband and her elder sib, each of whom has expressive and receptive language problems. The youngest daughter in this family, who was observed to have a similar pattern of language deficits, did not inherit the variant (Fig. 3). In the second family carrying the rs35139099 variant, the change was present in both the proband and his younger sib, who had expressive and receptive language scores $\sim$1SD below that expected for his age, indicating that it is inherited (Fig. 3). The variant was not present in the mother and we did not have a DNA sample, or phenotypic data, from the father. Nonetheless, haplotype analyses of genome-wide SNP data indicated that the two children shared the same paternal chromosome in this region indicating that the rs35139099 variant was likely inherited from the father.

Discussion

In this paper, we report results from the whole exome sequencing of five individuals from an isolated Chilean island population affected by a high incidence of SLI. We identify a heterozygous nonsynonymous coding variant in the *NFXL1* gene that is shared by three of the five individuals sequenced. Association analyses within a larger Robinson Crusoe validation cohort, demonstrated that this variant occurred at a significantly increased frequency in Islanders with language impairment than those with typical language development (P = 0.0002) and is predicted to be "disease-causing". Subsequent sequencing of *NFXL1* in a cohort consisting of 117 independent UK probands (SLIC) with SLI identified four individuals with putative high-risk variants in the heterozygous state; three SLIC individuals carried rare nonsynonymous changes and one SLIC individual carried a novel variant that falls within a regulatory motif. Given the above evidence, we postulate that variants within *NFXL1* may contribute to genetic risk of language impairment. We propose that such changes are likely to function as—risk variants with a complex model of inheritance.

We used the Robinson Crusoe ancestry to trace back the relationships between individuals carrying the associated rs144169475 variant. The only common ancestors to the carriers were two founder brothers who had previously been reported to head the SLI lineage on the Island (Fig. 1). These brothers were related to all carriers of the rs144169475 variant (Fig. 1). However no single brother was related to all Islanders carrying the variant allele (Fig. 1). We therefore concluded that both founder brothers are likely to have carried the variant. These data therefore support the founder model of language impairment proposed at the outset of this study. We performed allele dropping simulations within the descendants of these founder brothers and found that a variant with an allele frequency of 3–9% in the founder population would be expected to have a frequency of 8–14% in the current population (S1 Text, S4 Table). This

prediction fits well with the observed frequency of 12.5% in the founder-related Islanders and is elevated above that observed in Chilean population controls (7.4%), indicating the presence of a founder effect at this locus. Moreover, we found that the increased frequency of the rs144169475 variant is driven by Islanders with SLI (19.4% in 49 individuals with SLI vs 4.8% in 62 individuals with typical language ability) (Table 2).

Our data further suggest that the effects of rare mutations in *NFXL1* may extend to the etiology of SLI in other populations. In a screen of the *NFXL1* coding regions in 117 independent UK probands affected by SLI, we observed four individuals who carried rare coding changes generating a combined high risk allele frequency of 1.71%. By contrast, the combined allele frequency of these three variants in 4679 independent European controls (from the 1000 genomes and EVS public databases) is 0.47%, a difference that yields a marginally significant *P*-value of 0.029 (4 of 234 SLIC chromosomes vs 44 of 9358 control chromosomes). Extending our investigations to include all private coding mutations (i.e. only found in one individual) across the entire *NFXL1* transcript, as opposed to the consideration of the three specific mutations considered above, we again observed a marginally increased frequency in the SLIC cohort (2 of 234 chromosomes tested, 0.85%) above that expected given the data reported in public European databases (EVS European American and 1000 Genomes EUR super-population – 28 of 9358 chromosome sequenced, 0.3%, P = 0.0359). Broadening our investigation to include all rare coding changes (<1%) across the entire *NFXL1* transcript, revealed a similar trend (1.71% (4 of 234 chromosomes sequenced) in the SLIC cohort, compared to 1.36% (127 of 9358 chromosome sequenced) in public European databases) but this did not reach significance (P = 0.3821).

Given our consistent findings across cohorts, and in line with the data arising from other neurodevelopmental disorders, we suggest that rare variants in *NFXL1* may represent genetic risk factors with incomplete penetrance. Given our data, it is likely that these putative risk factors are modulated by other genetic variations and/or environmental factors [41–43]. We could not identify a distinct or specific phenotypic feature that distinguished rs144169475 language-impaired carriers from language-impaired non-carriers. Nor did we observe complete co-segregation between *NFXL1* variants and the presence of SLI in either the Robinson Crusoe validation population or the UK SLIC cohort. Thirty nine percent of the Robinson Crusoe validation cohort affected by language impairment carried the rs144169475 variant, as did ten percent of the Robinson Crusoe validation cohort with typical language ability. Similarly, one of the variants observed in the SLIC probands was inherited by a child with typical language development and two children affected by language impairment did not inherit the observed variant. In addition, we observed a high phenocopy rate in the Robinson Crusoe cohort; only 39% of individuals affected by language impairment carried the rs144169475 variant. Incomplete segregation is commonly described in neurodevelopmental disorders such as autism [42,43] and intellectual disability [44,45] and represents a major challenge in the interpretation of high-throughput sequencing data [46].

The *NFXL1* gene encodes a NFX-1-type nuclear zinc-finger transcriptional repressor that is expressed at the cytoplasm [47]. Little is known regarding the function of the NFXL1 protein; no disorders have been identified that arise from the mutation of this gene and no animal knock-outs have been described. The protein has zinc-finger domains which mediate DNA binding and carries a RING domain that has E3 ubiquitin ligase function (Fig. 2) [48]. This transcription factor has been shown to be highly expressed in embryonic stem cells prior to differentiation into myelinated oligodendrocytes [49] and shows a high level of expression in the early mouse embryonic development (E11.5) and in human cerebellar structures (www.brainmap.org). *NFXL1* is so-called because it is a paralogue of the NF-X1 transcription factor which binds the X-box sequence of class II *MHC* genes [50]. This feature may be relevant in

light of a recent study that found association between HLA loci and SLI [51]. Similarly, an *NFX1* isoform functions in the regulation of the NFκB pathway [52], as does *CMIP*, a gene implicated in the etiology of SLI in UK populations [7,53].

Limitations of our study

A natural limitation of all studies of founder or isolated populations is the restricted size of the cohort. Although our study represents a comprehensive profiling of the Robinson Crusoe child population, the total sample consisted of only 111 individuals, 100 of whom were founder-related and 49 of whom had language impairment. Although it should be noted that the power of this particular sample lies in the close relationships between individuals rather than the absolute number of samples, the issue of sample sizes is especially pertinent when one is considering rare variations. Thus it is of particular importance that we observed independent evidence implicating NFXL1 rare variants in another cohort. However, in the absence of a large South American cohort of language-impaired individuals, we were unable to include the rs144169475 variant in our replication investigations (since this SNP is particular to South American populations). Thus, further studies of larger sample sizes that include language-selected controls and South American individuals will be required to fully evaluate the role of rs144169475 and rare *NFXL1* coding variants in SLI susceptibility.

Of note, none of the shared variants identified through exome sequencing co-incided with regions of suggestive linkage reported in a previous genomewide linkage study of the Robinson Crusoe population (S2 and S3 Tables) [54]. Nor did we find evidence for linkage to the *NFXL1* region of chromosome 4 (S4 Fig.). We must therefore acknowledge that the increased frequency of rs144169475 in language-impaired individuals of the Robinson Crusoe validation cohort does not directly indicate pathogenicity. The result may represent a chance finding or, alternatively, rs144169475 may be a proxy for the causal variant. Since the exome sequencing performed did not capture 100% of the exome, it is possible that the causal variant was not detected here. Full genome sequencing would be required to fully investigate this possibility. However, it is also important to note that a lack of linkage does not preclude the presence of a causal variant and may instead reflect the complexities of analyzing a pedigree of this size and complexity [55]. The pedigree, which explained the known relationships between the founder brothers and the Robinson Crusoe validation cohort, included 288 individuals (321 bits, where a bit is defined as twice the number of non-founders—the number of founders) and so had to be broken into smaller sets for linkage analyses. This segmentation process discards information and can reduce the power to detect linkage if individuals sharing the linked chromosome segment are split between sub-pedigrees [56]. Lastly, since we hypothesize that SLI in this population has a complex genetic basis and involves incomplete and a high phenocopy frequency, it is possible that the power to detect linkage is insufficient. We observed reduced penetrance at the *NFXL1* locus (of 25 variant carriers, 19 were diagnosed with SLI, penetrance of 76%) in combination with evidence of a high phenocopy rate in our cohort (of 49 individuals with language impairment, 19 carried the variant, phenocopy rate of 61%). In combination, these factors break down the correspondence between genotype and phenotype, compromising the ability to detect linkage [57].

In summary, the Robinson Crusoe admixed founder population represents a rare resource which may assist in the identification of genetic variants that contribute to SLI susceptibility. Exome sequencing of five individuals from this population identified eight shared coding variants. One of these variants (rs144169475) was found to be significantly associated ($P = 0.0002$) with language impairment in the wider Robinson Crusoe population. rs144169475 confers a nonsynonymous change (N150K) in the *NFXL1* gene at a highly conserved residue. Subsequent

sequencing of the *NFXL1* coding regions in 117 independent UK SLI cases identified four individuals with rare heterozygous variants predicted to be of functional consequence. We conclude that coding variants within *NFXL1* confer an increased risk of SLI within a complex genetic model.

Materials and Methods

Ethics

The work on the Robinson Crusoe Island was approved by the ethics department of the University of Chile. Ethical permission for each SLIC collection was granted by local ethics committees. Guys Hospital Research Ethics Committee approved the collection of families from the Newcomen Centre to identify families from the South East of England with specific language disorder. Ref No. 96/7/11. Cambridge Local Research Ethics Committee approved the CLASP project "Genome Search for susceptibility loci to language disorders" Ref No. LREC96/212. Ethical approval for the Manchester Language Study was given by the University of Manchester Committee on the Ethics of Research on Human Beings. Ref No. 03061 The Lothian Research Ethics Committee approved the project "Genetics of specific language impairment in children in Scotland" for the use of the Edinburgh samples. Ref. No. LREC/1999/6/20. The ethics department of the University of Chile approved the project "Genetic analysis of language-impaired individuals from the Robinson Crusoe Island". Project Number 001-2010. Informed consent was given by all participants and/or, where applicable, their parents.

Ascertainment of the Robinson Crusoe population

The Robinson Crusoe cohort was ascertained on the basis of phenotypic data from 61 children, between the ages of 3 years and 8 years, 11 months (i.e. the child cohort, described below) all of whom were descendants of the founder families and represents an extended cohort (including children who have turned 3 years of age since 2008) of that described in [33]. First-degree relatives of founder-related children found to meet criteria for SLI or typical language development were then also assessed for language performance (i.e. the family cohort, described below). Age constraints of available standardized tests meant that different language batteries were employed within the child and family cohorts.

Phenotyping and selection of the Robinson Crusoe child cohort

The language ability of 61 children, all of whom were related to a founder individual, was assessed by tests of expressive and receptive language (Toronto Spanish Grammar Exploratory test, TEGE [58]) and phonology (Phonological simplification test (Test para Evaluar Procesos de Simplificación Fonológica—TEPROSIF [59]). Nonverbal IQ was tested using the Colombia Mental Maturity Scale [60]. In addition, all children were subjected to an auditory screen and oral motor exam [61]. All tests were validated and normalized in Chilean populations. On the basis of these tests, all children were classified into one of the three following categories:

1. "Specific Language Impairment)" (N = 16, 7 male, 9 female, 26.2%) defined as (i) performance >2SD below expected on TEPROSIF (for children aged 6 years or less) or performance >2 years below expected for chronological age on TEPROSIF (for children aged over 6 years) and/or performance below the 10^{th} percentile on either the receptive or expressive scales of the TEGE, (ii) nonverbal IQ not below the 10^{th} percentile, (iii) normal hearing, oral motor skills and neurological development.

2. "Typical language development" (N = 23, 8 male, 15 female, 37.7%) defined as (i) performance not >2SD below expected on TEPROSIF or performance >2 years below expected for chronological age on TEPROSIF (for children aged over 6 years) and performance above the 10th percentile on both the receptive and expressive scales of the TEGE.
3. "Nonspecific language impairment" (N = 22, 13 male, 9 female, 36.1%) defined as (i) performance >2SD below expected on TEPROSIF or performance >2 years below expected for chronological age on TEPROSIF (for children aged over 6 years) and/or performance below the 10th percentile on either the receptive or expressive scales of the TEGE, and (ii) nonverbal IQ >1SD below age-expected, and/or (iii) evidence of hearing loss or oral motor disability (e.g cleft lip) or abnormal neurological development.

The observed language deficits in the individuals diagnosed with SLI were typical of those described in other SLI cohorts and involved varied deficits across grammatical, morphosyntactical and receptive aspects of language, but not dialectic variations in intonation, vocabulary or phonology.

Phenotyping and selection of the Robinson Crusoe family cohort

Since we were particularly interested in genetic contributions to SLI, our family cohort consisted of the first-degree relatives of the 39 founder-related children presenting with SLI or typical language development. All available first-degree family members (92 parents and siblings, 47 male, 45 female) were assessed for language difficulties using tests of verbal fluency (Barcelona test [62]) and verbal comprehension (Token test [63]). These family members included 11 parents who were not related to a founder member of the Island (referred to as non-founder-related parents). In addition to these formal language assessments, all individuals (or their parents or spouses) completed a family history interview (provided by P Tallal) [64], which specifically asks questions regarding language difficulties. On the basis of these data individuals were classified as either:

1. "Language-impaired" (N = 34, 15 male, 19 female, 37.0%, including 4 non-founder-related parents) if they scored below the 10th percentile on either the Barcelona test or the token test or they self-reported a need for writing or reading support at school or a history of language support in the family history questionnaire.
2. "Typical language ability" (N = 58, 32 male, 26 female, 63.0%, including 7 non-founder-related parents) if they scored above the 10th percentile on both the Barcelona test and the token test and they indicated no requirement for writing, reading or language support in the family history questionnaire.

Exome sequencing of selected Robinson Crusoe children

Five Islanders (3 male, 2 female) from the child cohort who had been diagnosed with SLI were selected for exome sequencing. The selection of individuals for sequencing was based upon the amount and quality of DNA available, the severity of observed language impairment and their known relationships with other affected individuals. The five children were selected to cover the different branches of the founder pedigree and were descendants of the founder families (Fig. 1).

Exome capture was performed using 10μg of genomic DNA with a first generation (v1) Agilent SureSelect human exome kit (Agilent, Santa Clara, CA, USA), which provide an average target coverage of 80% of the exome at 56-fold across all samples. Sequencing of the generated fragments was performed on the SOLiD 4 sequencer (Life Technologies, Carlsbad, CA, USA).

Color space reads were mapped to the human reference genome (hg18) in the SOLiD bioscope software (v1.2), which applies an iterative mapping approach. Variants were called using a diBayes algorithm [65] using high stringency settings, requiring calls on each strand. Small insertions and deletions were detected using the SOLiD Small Indel Tool. We assumed a binomial distribution with a probability of 0.5 of sequencing the variant allele at a heterozygous position. Given such a distribution, a minimum of ten reads would be required to provide a 99% probability that two or more reads contain an allele variant call. We filtered variant calls to have at least four unique (i.e. different start sites) variant reads with the variant being present in at least 15% of all reads.

To test the hypothesis that the founder brothers carried a rare causative genetic mutation, for our downstream analyses, we focused upon novel variants that were potentially deleterious. Each exome file was individually filtered to exclude nongenic, intronic (other than canonical splice sites) and synonymous variants. The remaining nonsynonymous and splice-site mutations were further filtered to exclude known sites of variation (as described in dbSNP, (build 130), publically available genome sequences and an in-house sequencing database). The remaining variants were then compared across exome samples to allow the selection of variants that occurred in 3 or more of the 5 children sequenced. A flow diagram of the methodology can be found in S1 Fig.. Shared novel, potentially deleterious variants discovered in the exome data were verified by Sanger sequencing. Primers for Sanger sequencing were designed in primer3 [66]. Primer sequences are available on request.

Association analyses of selected variants in the Robinson Crusoe population

All novel nonsynonymous or canonical splice-site variants found to occur in 3 or more of the 5 exome samples were also genotyped in the wider child and family cohorts from the Robinson Crusoe population. We were able to obtain DNA samples for 35 founder-related children (from the SLI and typical language development child groups described above) and their family members (from the family cohort described above). Forty nine of these individuals (16 children, 22 parents (4 of whom were non-founder-related), 7 siblings and 4 half-siblings) were language impaired and 62 (19 children, 32 parents (7 of whom were non-founder-related), 9 siblings and 2 half-siblings) had language ability in the normal range. These families included the five children used in the exome sequencing. DNA was extracted from EDTA whole blood samples using a standard chloroform extraction protocol. All novel nonsynonymous or canonical splice-site variants identified from the exome screen were sequenced using a standard Sanger protocol in these 111 individuals.

The resultant genotype data were used to perform a family-based test of association within the MQLS-XM package [36,67]. This algorithm calculates a quasi-likelihood score which corrects the Chi-square statistic for relationships between individuals, providing accurate type I error rates [68]. The MQLS-XM extension allows for the accurate application of this statistic to X-linked markers [67]. The MQLS algorithm distinguishes between unaffected controls and controls of unknown phenotype, can incorporate phenotypic data from individuals who have not been genotyped [36] and is robust to the mis-specification of prevalence. It allows for the presence of both linkage and association effects in the test statistic and is computationally straightforward making it particularly suitable for large complex pedigrees in which cases and controls may be inter-related, as is the case in this study [36].

A full pedigree structure was generated that accounted for all known relationships between 111 individuals from the child and family cohorts and the two identified, shared, founder brothers. This pedigree included 288 individuals (141 males, 144 females and 85

founders (i.e. individuals with no parental information available—both original founders and incoming), 203 non-founders) over 5 generations. As described above, 111 individuals (including 11 non-founder-related parents) had full genotype and phenotype data, 11 individuals were also included who had phenotype data but no genotype data and the remaining 166 individuals had no phenotype or genotype data but defined relationships between the 111 genotyped individuals and the founder brothers. In the MQLS-XM analyses, the expected prevalence of SLI was set at 0.25 for males and 0.27 for females. These figures were derived from the child cohort described above.

Any variant that was significantly associated with language impairment in the population cohort was genotyped in 127 independent European population controls (ECACC, HRC-1 DNA Panel),441 independent South American controls; 320 individuals of Colombian descent and 121 individuals of Chilean origin. The Colombian controls were collected as part of a genetic demography study in the Colombian population, where all participants had to have four grandparents of local origin (provided by Luis Carvajal-Carmona and Maria Magdalena Echeverry). The Chilean controls were ascertained from the Santiago area and consisted of DNA from 30 male Chilean students (provided by P Villanueva) and from 91 female adult controls from a breast cancer study (provided by L Jara, University of Chile). Genome-wide SNP data indicated that these samples were of Amerindian and European ancestry. Note that both the European and South American control populations were unselected and, as such, were not screened for language ability.

Linkage analysis of chromosome 4

Genome-wide linkage data for the Robinson Crusoe validation cohort have previously been reported [35]. These previous analyses included 6,090 SNPs and reported suggestive linkage ($P<7.3\times10^{-4}$) between SLI and chromosomes 2, 6, 7, 8, 9, 12, 13 and 17. In the current study, we had access to a new set of denser genotypes from the Robinson Crusoe population, generated with the Affymetrix Axiom GW-LAT 1 array (Affymetrix Inc, Santa Clara, CA, www.affymetrix.com), supplemented with a custom array designed to cover South American-specific variants which together included 1,141,741 SNPs.

929 SNPs across chromosome region chr4:46,000,000–49,000,000 (hg19) were selected to cover the chromosome region surrounding the *NFXL1* gene (reported transcript—chr4:47,849,258–47,916,680, hg19). SNP data were filtered within PLINK [69] to remove markers in close linkage disequilibrium ($r^2>0.5$) resulting in a linkage dataset of 54 independent SNPs that were appended with rs144169475 genotype data and analysed for linkage in MERLIN [70]. Linkage disequilibrium between these SNPs and rs144169475 are provided in S3 Fig.

Since linkage packages were unable to analyse genome-wide data for the 321-bit Robinson Crusoe validation pedigree as a whole, it was broken into sub-pedigrees manually selected on the basis of closest shared ancestor. We employed linkage sub-pedigrees and linkage methods analogous to those described in the previous linage study [35]; Seven extended families of 20–24 bits (where a bit is defined as twice the number of non-founders—the number of founders) were analysed for linkage under parametric and nonparametric models with MERLIN (S2 Fig.) Parametric linkage analyses were performed under a model which reflected the observed nature of rs144169475 (assuming a disease frequency of 26.2% (as observed in the Robinson Crusoe children) and penetrance of 0.76 (as observed in the Robinson Crusoe validation cohort). Nonparametric linkage results are reported as P-values derived from the Kong and Cox exponential model, which can be more powerful in large pedigrees [71]. Expected allele frequencies were derived from the 1000 Genomes AMR super-population (integrated phase 1, accessed March 2014) which includes 181 independent South American individuals (60 Colombians

from Medellin, Colombia (CLM), 66 individuals with Mexican ancestry in Los Angeles (MXL) and 55 Puerto Ricans from Puerto Rico (PUR)) [40].

Functional effects of identified variants

Putative functional effects of associated variants were evaluated using MutationTaster [72]. MutationTaster uses a Bayes classifier which integrates information from various biomedical databases and analysis tools to evaluate the possible pathogenicity of coding variants. MutationTaster considers evolutionary conservation at both the nucleotide and amino acid level, splice-site changes, loss of protein motifs or features and changes that might affect the level of mRNA expression and stability within a single tool to classify variants as a "disease mutation" or a "polymorphism". A p-value is given to indicate "the security" of the prediction [72]. The MutationTaster algorithm was trained using more than 390,000 known disease mutations from HGMD and more than 6,800,000 SNPs and Indel polymorphisms from the 1000 Genomes Project.

For each of the variants highlighted, we also present the SIFT and polyphen-2 scores. In contrast to MutationTaster, the SIFT and PolyPhen algorithms primarily consider protein sequences, motifs and structures to assign pathogenicity and therefore can only be applied to coding changes. SIFT performs a multiple alignment of closely related protein sequences to identify conserved motifs and assign a probability that a given amino acid substitution is pathogenic [73]. PolyPhen-2 uses a Bayes classifier to consider the property of the reference and variant amino acids, the amino acid conservation, protein motifs and 3D protein structure to derive a probability that a mutation is damaging [74]. SIFT scores vary between 0 and 1. Amino acid substitutions are classified as "deleterious" for scores $\leq$0.05 and "tolerated" for scores $>$0.05. In Polyphen-2, two training models are available—HumDiv, which is more appropriate for the identification of fully penetrant Mendelian mutations and HumVar, which is more appropriate for the classification of rare alleles at loci potentially involved in complex phenotypes. PolyPhen scores from both of these models vary from 0 to 1, where 0 represents a variant with no functional effect. Functional effects are classified as "benign", "possibly damaging", or "probably damaging", depending on whether the posterior probability falls above or below the appropriate false positive thresholds.

Sequencing of candidate genes in SLIC cohort

In order to further investigate the role of NFXL1 variants in SLI, the coding regions of the *NFXL1* gene were subsequently sequenced in 117 unrelated British children affected by SLI. These children formed part of the SLI Consortium (SLIC) collection, which has previously been described in detail [7,37,39]. In short, the probands were collected from four sites across the UK (The Newcomen Centre at Guy's Hospital, London, the Cambridge Language and Speech Project (CLASP) [75], the Child Life and Health Department at the University of Edinburgh [76] and the Manchester Language Study [77]). All probands were selected to have receptive and/or expressive language skills (as assessed by the Clinical Evaluation of Language Fundamentals (CELF-IV-R) [78]) more than 1.5SD below the normative mean for his or her age and non-verbal IQ (as measured by the Wechsler Intelligence Scale for Children [79]) in the "normal" range ($>$80).

The concentration of genomic DNA samples from 117 independent SLIC probands was quantified by picogreen and each sample normalized to 10ng/μl. Individual DNAs were pooled prior to PCR amplification. Following PCR, the amplicons were fragmented, end-repaired and adapter-ligated. The prepared and tagged libraries were then multiplexed before paired-end sequencing in a single lane of flow-cell on an Illumina HiSeq 2000 (Illumina Inc, SanDiego, CA,

www.illumina.com). Sequences were aligned against human reference sequence (37d5) using STAMPY [80] and variants called by the Syzygy (1.2.6) algorithm to create a VCF file. Syzygy implements a Bayes likelihood calculation to allow a base calling strategy that is particularly suited to the calling of variants in pooled samples, in which the frequency of reads containing a rare variant will be lower than expected [81]. Identified sequence variants were annotated within the SNPeff package allowing the identification of coding variants [82]. Individual DNAs from all pools that contained a nonsynonymous coding variant with an expected frequency of <5% were resequenced by Sanger sequencing using primers designed with the primer3 software [66]. This allowed the verification of the variants, the derivation of true variant frequencies across pools and the identification of the individuals who carried the variant.

The allele frequencies of coding variants discovered in SLIC probands were compared to those observed in 4679 individuals of European ancestry across publically available control databases; the 1000 genomes project (the European (EUR) super-population from integrated phase 1, accessed March 2014) [40] which includes 379 independent European individuals (89 British in England and Scotland, 93 Finnish in Finland, 14 Iberian populations in Spain, 98 Toscani in Italy and 85 Utah residents with Northern and Western European ancestry) and the European American (EA) cohort from the exome variant server (ESP6500 SI-V2, accessed March 2014) (http://evs.gs.washington.edu/EVS/) which includes data from 4300 independent individuals of European American ancestry. The 1000 genomes samples are unselected controls while the EVS samples are selected to include controls, extremes of specific traits (LDL and blood pressure) and specific diseases (early onset myocardial infarction and early onset stroke). Allele frequencies were compared between SLIC probands and controls using a two-tailed Fisher's exact test with 1 degree of freedom. Calculations were performed in the graphpad online calculator (http://www.graphpad.com/). Where given variants were observed in alternative populations, these data are reported but were not included in the statistical analyses since population admixture and stratification can lead to false positives, especially when investigating rare variants [83].

Supporting Information

S1 Fig. A flow diagram showing the filtering of the exome data. Blue boxes show each filter step and red boxes describe exclusion criteria involved in each step
(PDF)

S2 Fig. Structure of pedigrees used for linkage analyses (redrawn using data from [35])). Seven pedigrees of no more than 24-bits were used for linkage analyses. Individuals with language impairment are colored in black. Individuals with typical language are denoted in white. Individuals with unknown phenotype are shaded grey.
(PDF)

S3 Fig. Linkage disequilibrium between markers across *NFXL1* region. a—Linkage disequilibrium between all genotyped markers (n = 929) across chr4:46–49Mb (hg19). b—Linkage disequilibrium between all analyzed markers (n = 55) across chr4:46–49Mb (hg19), after pruning for r^2>0.5. Position of *NFXL1* gene is shown by red box. Plots were generated in haploview (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview) using linkage pedigrees (as shown in S2 Fig.). Color scheme is standard haploview colour scheme (blue—D' = 1, LOD<2; white—D'<1, LOD<2; pink/red—LOD≥2).
(PDF)

S4 Fig. Linkage across the *NFXL1* region. No linkage was observed to the NFXL1 region of chromosome 4 under parametric and non-parametric models using a dense SNP array. The approximate position of the *NFXL1* gene is indicated by the red box on the X axis.
(PDF)

S1 Table. Exome sequencing metrics.
(PDF)

S2 Table. Regions of suggestive linkage in the Robinson Crusoe population (as presented in [35]).
(PDF)

S3 Table. All variants found under the peaks of previous linkage (as reported in [35]) that were shared across all 5 exome samples.
(PDF)

S4 Table. Genotype reconstruction simulations. The *NFXL1* variant has an expected population frequency of between 0.033 (1000 genomes CLM) and 0.09 (1000 genomes PUR) and is predicted to be present in both founder brothers (frequency in founder brothers of 0.5). Given the population structure, it would therefore be expected to be present in the current population at a frequency of between 0.08 (MAF = 0.03) and 0.14 (MAF = 0.10). Although the frequency of the *NFXL1* variant in the founder-related individuals of the Robinson Crusoe validation cohort was at this expected level (0.125), the variant allele was found to cosegregate with language impairment; the frequency of the *NFXL1* variant in the founder-related individuals with SLI was above expected (0.194) while that of founder-related individuals with typical language was below expected (0.048), supporting a pathogenic role for this allele.
(PDF)

S1 Text. Genotype reconstruction simulations.
(PDF)

Acknowledgments

We would like to thank all the families, professionals and individuals who participated in this research. In particular we are extremely grateful to the inhabitants of Robinson Crusoe Island who have agreed to participate in this study. We would also like to thank Mr. Felipe Paredes, the mayor of the Ilustre Municipalidad de Juan Fernández for his infinite assistance and patience in the development of this research. Also to the authorities of schools of medicine and dentistry for giving us the necessary permits to travel to the island of Juan Fernandez.

We are very grateful to all members of the SLI Consortium for their contributions to this work: V. Slonims (Newcomen Centre, Evelina Children's Hospital, London, UK), A. Clark, J Watson (Speech and Hearing Sciences, Queen Margaret University, Edinburgh, UK), E. Simonoff, A Pickles (King's College London, Institute of Psychiatry); A. Everitt (University Child Health and DMDE, University of Aberdeen); J. Seckl (Molecular Medicine Centre, University of Edinburgh); H. Cowie (Department of Speech and Language Therapy, Royal Hospital for Sick Children, Edinburgh); W. Cohen (Psychological Sciences and Health, University of Strathclyde); J. Nasir (Division of Biomedical Sciences, St George's University of London); D. V.M. Bishop (Department of Experimental Psychology, University of Oxford); Z. Simkin (School of Psychological Sciences, University of Manchester).

Author Contributions

Conceived and designed the experiments: PV LJ GB GCR AO PFB ERH SLIC CF HP LCC JAV JBC ZDB SEF DFN. Performed the experiments: PV RN AH MAF NHS CG RHR MME LCC ZDB DFN. Analyzed the data: PV RN AH MAF CG LCC JBC DFN. Contributed reagents/materials/analysis tools: PV AH MME CF GB GCR AO PFB ERH DFN. Wrote the paper: PV DFN.

References

1. Harel S, Greenstein Y, Kramer U, Yifat R, Samuel E, et al. (1996) Clinical characteristics of children referred to a child development center for evaluation of speech, language, and communication disorders. Pediatr Neurol 15: 305–311. PMID: 8972529
2. American-Psychiatric-Association (2000) Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IVTR). Washington, DC: American Psychiatric Publishing.
3. Law J, Boyle J, Harris F, Harkness A, Nye C (2000) Prevalence and natural history of primary speech and language delay: findings from a review of the literature. Int J Lang Commun Disord 35: 165–188. PMID: 10912250
4. Whitehouse AJ (2010) Is there a sex ratio difference in the familial aggregation of specific language impairment? A meta-analysis. J Speech Lang Hear Res 53: 1015–1025. doi: 10.1044/1092-4388(2009/09-0078) PMID: 20605945
5. Bishop DV, Laws G, Adams C, Norbury CF (2006) High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. Behav Genet 36: 173–184. PMID: 16485179
6. Bishop DV (2001) Genetic and environmental risks for specific language impairment in children. Philos Trans R Soc Lond B Biol Sci 356: 369–380. PMID: 11316485
7. Newbury DF, Winchester L, Addis L, Paracchini S, Buckingham LL, et al. (2009) CMIP and ATP2C2 modulate phonological short-term memory in language impairment. Am J Hum Genet 85: 264–272. doi: 10.1016/j.ajhg.2009.07.004 PMID: 19646677
8. Luciano M, Evans DM, Hansell NK, Medland SE, Montgomery GW, et al. (2013) A genome-wide association study for reading and language abilities in two population cohorts. Genes Brain Behav 12: 645–652. doi: 10.1111/gbb.12053 PMID: 23738518
9. Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, et al. (2014) Genome-wide screening for DNA variants associated with reading and language traits. Genes Brain Behav 13: 686–701. doi: 10.1111/gbb.12158 PMID: 25065397
10. St Pourcain B, Cents RA, Whitehouse AJ, Haworth CM, Davis OS, et al. (2014) Common variation near ROBO2 is associated with expressive vocabulary in infancy. Nat Commun 5: 4831. doi: 10.1038/ncomms5831 PMID: 25226531
11. Eicher JD, Powers NR, Miller LL, Akshoomoff N, Amaral DG, et al. (2013) Genome-wide association study of shared components of reading disability and language impairment. Genes Brain Behav 12: 792–801. doi: 10.1111/gbb.12085 PMID: 24024963
12. Nudel R, Simpson NH, Baird G, O'Hare A, Conti-Ramsden G, et al. (2014) Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment (SLI). Genes, Brain, Behavior 13: 418–429. doi: 10.1111/gbb.12127 PMID: 24571439
13. Ceroni F, Simpson NH, Francks C, Baird G, Conti-Ramsden G, et al. (2014) Homozygous microdeletion of exon 5 in ZNF277 in a girl with specific language impairment. Eur J Hum Genet 22: 1165–1171. doi: 10.1038/ejhg.2014.4 PMID: 24518835
14. Simpson NH, Addis L, Brandler WM, Slonims V, Clark A, et al. (2013) Increased prevalence of sex chromosome aneuploidies in specific language impairment and dyslexia. Dev Med Child Neurol 56: 346–353. doi: 10.1111/dmcn.12294 PMID: 24117048
15. Newbury DF, Paracchini S, Scerri TS, Winchester L, Addis L, et al. (2011) Investigation of dyslexia and SLI risk variants in reading—and language-impaired subjects. Behav Genet 41: 90–104. doi: 10.1007/s10519-010-9424-3 PMID: 21165691
16. Scerri TS, Morris AP, Buckingham LL, Newbury DF, Miller LL, et al. (2011) DCDC2, KIAA0319 and CMIP are associated with reading-related traits. Biological Psychiatry 70: 237–245. doi: 10.1016/j.biopsych.2011.02.005 PMID: 21457949
17. Shu W, Yang H, Zhang L, Lu MM, Morrisey EE (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 276: 27488–27497. PMID: 11358962

18. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413: 519–523. PMID: 11586359
19. Feuk L, Kalervo A, Lipsanen-Nyman M, Skaug J, Nakabayashi K, et al. (2006) Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. Am J Hum Genet 79: 965–972. PMID: 17033973
20. MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, et al. (2005) Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet 76: 1074–1080. PMID: 15877281
21. Shriberg LD, Ballard KJ, Tomblin JB, Duffy JR, Odell KH, et al. (2006) Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. J Speech Lang Hear Res 49: 500–525. PMID: 16787893
22. Zeesman S, Nowaczyk MJ, Teshima I, Roberts W, Cardy JO, et al. (2006) Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. Am J Med Genet A 140: 509–514. PMID: 16470794
23. Fisher SE (2006) Tangled webs: tracing the connections between genes and cognition. Cognition 101: 270–297. PMID: 16764847
24. Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, et al. (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 81: 1144–1157. PMID: 17999357
25. Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, et al. (2007) High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech & language disorders. Am J Hum Genet 81: 1232–1250. PMID: 17999362
26. Rodenas-Cuadrado P, Ho J, Vernes SC (2014) Shining a light on CNTNAP2: complex functions to complex disorders. Eur J Hum Genet 22: 171–178. doi: 10.1038/ejhg.2013.100 PMID: 23714751
27. Stein MB, Yang BZ, Chavira DA, Hitchcock CA, Sung SC, et al. (2011) A common genetic variant in the neurexin superfamily member CNTNAP2 is associated with increased risk for selective mutism and social anxiety-related traits. Biol Psychiatry 69: 825–831. doi: 10.1016/j.biopsych.2010.11.008 PMID: 21193173
28. Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, et al. (2008) A functional genetic link between distinct developmental language disorders. N Engl J Med 359: 2337–2345. doi: 10.1056/NEJMoa0802828 PMID: 18987363
29. Peter B, Raskind WH, Matsushita M, Lisowski M, Vu T, et al. (2011) Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. J Neurodev Disord 3: 39–49. doi: 10.1007/s11689-010-9065-0 PMID: 21484596
30. Whitehouse AJ, Bishop DV, Ang QW, Pennell CE, Fisher SE (2011) CNTNAP2 variants affect early language development in the general population. Genes Brain Behav 10: 451–456. doi: 10.1111/j.1601-183X.2011.00684.x PMID: 21310003
31. Whalley HC, O'Connell G, Sussmann JE, Peel A, Stanfield AC, et al. (2011) Genetic variation in CNTNAP2 alters brain function during linguistic processing in healthy individuals. Am J Med Genet B Neuropsychiatr Genet 156B: 941–948. doi: 10.1002/ajmg.b.31241 PMID: 21987501
32. Kos M, van den Brink D, Snijders TM, Rijpkema M, Franke B, et al. (2012) CNTNAP2 and language processing in healthy individuals as measured with ERPs. PLoS One 7: e46995. doi: 10.1371/journal.pone.0046995 PMID: 23115634
33. Villanueva P, de Barbieri Z, Palomino HM, Palomino H (2008) [High prevalence of specific language impairment in Robinson Crusoe Island. A possible founder effect]. Rev Med Chil 136: 186–192. PMID: 18483672
34. Villanueva P, Fernandez MA, Z DEB, Palomino H (2013) Consanguinity on Robinson Crusoe Island, an Isolated Chilean Population. J Biosoc Sci: 1–10. PMID: 23931260
35. Villanueva P, Newbury DF, Jara L, De Barbieri Z, Mirza G, et al. (2011) Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. Eur J Hum Genet 19: 687–695. doi: 10.1038/ejhg.2010.251 PMID: 21248734
36. Thornton T, McPeek MS (2007) Case-control association testing with related individuals: a more powerful quasi-likelihood score test. Am J Hum Genet 81: 321–337. PMID: 17668381
37. SLIC (2002) A genomewide scan identifies two novel loci involved in Specific Language Impairment. Am J Hum Genet 70: 384–398. PMID: 11791209
38. SLIC (2004) Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. Am J Hum Genet 74: 1225–1238. PMID: 15133743

39. Falcaro M, Pickles A, Newbury DF, Addis L, Banfield E, et al. (2008) Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in specific language impairment. Genes Brain Behav 7: 393–402. PMID: 18005161

40. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.

41. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, et al. (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43: 585–589. doi: 10.1038/ng.835 PMID: 21572417

42. Leblond CS, Jutta H, Delorme R, Proepper C, Betancur C, et al. (2012) Genetic and Functional Analyses of SHANK2 Mutations Suggest a Multiple Hit Model of Autism Spectrum Disorders. Plos Genetics 8: e1002521. doi: 10.1371/journal.pgen.1002521 PMID: 22346768

43. Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, et al. (2008) Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet 82: 165–173. doi: 10.1016/j.ajhg.2007.09.017 PMID: 18179895

44. Horn D, Kapeller J, Rivera-Brugues N, Moog U, Lorenz-Depiereux B, et al. (2010) Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. Hum Mutat 31: E1851–1860. doi: 10.1002/humu.21362 PMID: 20848658

45. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, et al. (2011) Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes. PLoS Genet 7: e1002334. doi: 10.1371/journal.pgen.1002334 PMID: 22102821

46. Gilissen C, Hoischen A, Brunner HG, Veltman JA (2012) Disease gene identification strategies for exome sequencing. European Journal of Human Genetics 20: 490–497. PMID: 22258526

47. Tang W, Yuan J, Chen X, Shan Y, Luo K, et al. (2005) Cloning and characterization of the CDZFP gene which encodes a putative zinc finger protein. DNA Seq 16: 391–396. PMID: 16323267

48. Mussig C, Schroder F, Usadel B, Lisso J (2010) Structure and putative function of NFX1-like proteins in plants. Plant Biol (Stuttg) 12: 381–394. doi: 10.1111/j.1438-8677.2009.00303.x PMID: 20522174

49. Chaerkady R, Letzen B, Renuse S, Sahasrabuddhe NA, Kumar P, et al. (2011) Quantitative temporal proteomic analysis of human embryonic stem cell differentiation into oligodendrocyte progenitor cells. Proteomics 11: 4007–4020. doi: 10.1002/pmic.201100107 PMID: 21770034

50. Song Z, Krishna S, Thanos D, Strominger JL, Ono SJ (1994) A novel cysteine-rich sequence-specific DNA-binding protein interacts with the conserved X-box motif of the human major histocompatibility complex class II genes via a repeated Cys-His domain and functions as a transcriptional repressor. J Exp Med 180: 1763–1774. PMID: 7964459

51. Nudel R, Simpson NH, Baird G, O'Hare A, Conti-Ramsden G, et al. (2014) Associations of HLA alleles with specific language impairment. J Neurodev Disord 6: 1. doi: 10.1186/1866-1955-6-1 PMID: 24433325

52. Xu M, Katzenellenbogen RA, Grandori C, Galloway DA (2010) NFX1 plays a role in human papillomavirus type 16 E6 activation of NFkappaB activity. J Virol 84: 11461–11469. doi: 10.1128/JVI.00538-10 PMID: 20739528

53. Kamal M, Valanciute A, Dahan K, Ory V, Pawlak A, et al. (2009) C-mip interacts physically with RelA and inhibits nuclear factor kappa B activity. Mol Immunol 46: 991–998. doi: 10.1016/j.molimm.2008.09.034 PMID: 19019440

54. Zweier C (2012) Severe Intellectual Disability Associated with Recessive Defects in CNTNAP2 and NRXN1. Mol Syndromol 2: 181–185. PMID: 22670139

55. Cummings AC, Lee SL, McCauley JL, Jiang L, Crunk A, et al. (2011) A genome-wide linkage screen in the Amish with Parkinson disease points to chromosome 6. Ann Hum Genet 75: 351–358. doi: 10.1111/j.1469-1809.2011.00643.x PMID: 21488853

56. Cummings AC, Torstenson E, Davis MF, D'Aoust LN, Scott WK, et al. (2013) Evaluating power and type 1 error in large pedigree analyses of binary traits. PLoS One 8: e62615. doi: 10.1371/journal.pone.0062615 PMID: 23658753

57. Nyholt D (2008) Principles of linkage analysis. In: Neale B, Ferreira M, Medland SE, Posthuma D, editors. Statistical genetics: gene mapping through linkage and association. New York; Abingdon: Taylor & Francis. pp. 113–134.

58. Pavez M (2003) Test exploratorio de Gramática española de A. Toronto. Aplicación en Chile. Santiago: Ediciones Universidad católica de Chile.

59. Pavez MM, Maggiolo M (2000) Test para evaluar los procesos fonologicos de simplificacion TEPROSIF; De Barbieri Z, editor. Santiago, Chile: Ediciones Escuela de Fonoaudiologia. v. p.

60. Burgemeister B, Blue L, Lorge I (1998) Escala de madurez mental. Columbia: Ed. TEA, Madrid.

61. Villanueva P (2000) Pauta de Examen en Habla y motricidad Orofacial. Santiago: Escuela de Fonoaudiología, Facultad de Medicina, Universidad de Chile.
62. Peña-Casanova J (1991) Programa integrado de exploración Neuropsicológica. Test de Barcelona. Masson, Barcelona.
63. De Renzi E, Vignolo L (1962) The Token Test: a sensitive tests to detect receptive disturbances in aphasics. Brain 85: 665–678. PMID: 14026018
64. Tallal P, Hirsch LS, Realpe-Bonilla T, Miller S, Brzustowicz LM, et al. (2001) Familial aggregation in specific language impairment. J Speech Lang Hear Res 44: 1172–1182. PMID: 11708534
65. Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, et al. (1999) A general approach to single-nucleotide polymorphism discovery. Nat Genet 23: 452–456. PMID: 10581034
66. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365–386. PMID: 10547847
67. Thornton T, Zhang Q, Cai X, Ober C, McPeek MS (2012) XM: association testing on the X-chromosome in case-control samples with related individuals. Genet Epidemiol 36: 438–450. doi: 10.1002/gepi.21638 PMID: 22552845
68. Bourgain C, Hoffjan S, Nicolae R, Newman D, Steiner L, et al. (2003) Novel case-control test in a founder population identifies P-selectin as an atopy-susceptibility locus. Am J Hum Genet 73: 612–626. PMID: 12929084
69. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575. PMID: 17701901
70. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30: 97–101. PMID: 11731797
71. Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61: 1179–1188. PMID: 9345087
72. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7: 575–576. doi: 10.1038/nmeth0810-575 PMID: 20676075
73. Dollaghan CA (2011) Taxometric analyses of specific language impairment in 6-year-old children. J Speech Lang Hear Res 54: 1361–1371. doi: 10.1044/1092-4388(2011/10-0187) PMID: 21646422
74. Bedore LM, Leonard LB (2001) Grammatical morphology deficits in Spanish-speaking children with specific language impairment. J Speech Lang Hear Res 44: 905–924. PMID: 11521782
75. Burden V, Stott CM, Forge J, Goodyer I (1996) The Cambridge Language and Speech Project (CLASP). I. Detection of language difficulties at 36 to 39 months. Dev Med Child Neurol 38: 613–631. PMID: 8674912
76. Clark A, O'Hare A, Watson J, Cohen W, Cowie H, et al. (2007) Severe receptive language disorder in childhood—familial aspects and long-term outcomes: results from a Scottish study. Arch Dis Child 92: 614–619. PMID: 17405857
77. Conti-Ramsden G, Botting N (1999) Characteristics of children attending language units in England: a national study of 7-year-olds. Int J Lang Commun Disord 34: 359–366. PMID: 10884906
78. Semel EM, Wiig EH, Secord W (1992) Clinical Evaluation of Language Fundamentals—Revised. San Antonio: Phychological Corporation.
79. Wechsler D (1992) Wechsler Intelligence Scale for Children - Third UK Edition. London: Psychological Corporation.
80. Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Research 21: 936–939. doi: 10.1101/gr.111120.110 PMID: 20980556
81. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, et al. (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43: 1066–1073. PMID: 21983784
82. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6: 80–92. doi: 10.4161/fly.19695 PMID: 22728672
83. Mathieson I, McVean G (2012) Differential confounding of rare and common variants in spatially structured populations. Nat Genet 44: 243–246. doi: 10.1038/ng.1074 PMID: 22306651

Int. J. Morphol.,
33(2):538-543, 2015.

Dental Morphological Markers as a Proxy for Ethnicity in Robinson Crusoe Islanders

Marcadores Morfológicos Dentarios en la Estimación de la Etnicidad Poblacional de la Isla Robinson Crusoe

Pía Villanueva[*,,***,****]; Mónica Quevedo[*****]; Zulema de Barbieri[***]; Soledad Piñeiro[******]; Carolina Herrero[******]; María Angélica Fernández[***] & Hernán Palomino[**]**

VILLANUEVA, P.; QUEVEDO, M.; DE BARBIERI, Z.; PIÑEIRO, S.; HERRERO, C.; FERNÁNDEZ, M. A. & PALOMINO, H. Dental morphological markers as a proxy for ethnicity in Robinson Crusoe islanders. ***Int. J. Morphol., 33(2)***:538-543, 2015.

SUMMARY: Chilean Robinson Crusoe Island is a semi-isolated location with unusually high rates of both consanguinity and language disorder. The current population of 633 inhabitants is descended almost exclusively from the colonization at the end of the 19th century, as there were few preceding immigrations to the island. This study investigates the genetic composition and degree of miscegenation within the island population, using dental morphological markers. The universe of island children was studied (n= 128, 3 to 15 years of age) using clinical exams, dental cast, and identification of each individual within a previously-constructed extensive genealogy for the island. The frequencies for Carabelli's cusp (61.7%), shovel-shaped incisor (9.4%), and sixth cusp (2.3%), along with the absence of seventh cusp, are consistent with a primarily Caucasian population. The estimated degree of miscegenation suggests an Amerindian component of 4.3%, which is consistent with the extensive known genealogies of the founders. Characterizing the genetic profile of Robinson Crusoe Island, a location with a remarkably high prevalence of language disorder, facilitates the comparison of the genetic variants underlying this pathology with those identified in European populations.

KEY WORDS: Dental morphological markers; Carabelli's cusp; Shovel-shaped incisor; Robinson Crusoe Island.

INTRODUCTION

Existing methods for studying human variation have progressed enormously from the anthropological definitions that gave origin to the concept of race. Various levels of observation have been carried out, including isonomy, qualitative and quantitative anthropometric measures, and immunologic methods, including DNA analysis. Results from one type of analysis, however, are not always consistent with results from another (Cavalli-Sforza & Bodmer, 1971; Rothhammer & Llop, 2004).

Human populations live in varied conditions, some isolated for thousands of years and others with variable rates of genetic mixture among groups. Studying the effects of population mixture on genetic variation allows for a detailed evaluation of the role of the environment in diseases that have both genetic and environmental causes (Weiss, 1995).

Current Chilean populations originate from different groups, and, therefore, identifying their genetic components (gene pool) is important in epidemiologic genetic research when studying the hereditary factors underlying common diseases. Analysis of the size and shape of maxillofacial and dental structures has contributed to the study of anthropology and human evolution. Teeth have a considerable genetic component, and, under normal conditions, dental structures are the organism's most stable, changing neither in shape nor size with aging (Rothhammer & Llop).

Dental morphological variation can be used to characterize and compare populations, as well as establish measures of biological distance similar to those established using serological or molecular markers (Palomino *et al*., 1977; Palomino, 1978; Rodríguez-Flórez, 2004; Bollini *et*

[*] Department of Pediatric Dentistry and Dentomaxillary Orthopedics, Faculty of Dentistry, Universidad de Chile, Santiago, Chile.
[**] Institute of Biomedical Sciences, Human Genetics Division, Faculty of Medicine, Universidad de Chile, Santiago, Chile.
[***] Department of Speech Language and Hearing Sciences, Faculty of Medicine, Universidad de Chile, Santiago, Chile.
[****] Doctoral Program of Psychology, Graduate School, Universidad de Granada, Granada, Spain.
[*****] Robinson Crusoe Island Health Public Centre, Juan Fernandez Archipelago, Chile.
[******] Private Practice of Dentistry, Santiago, Chile.

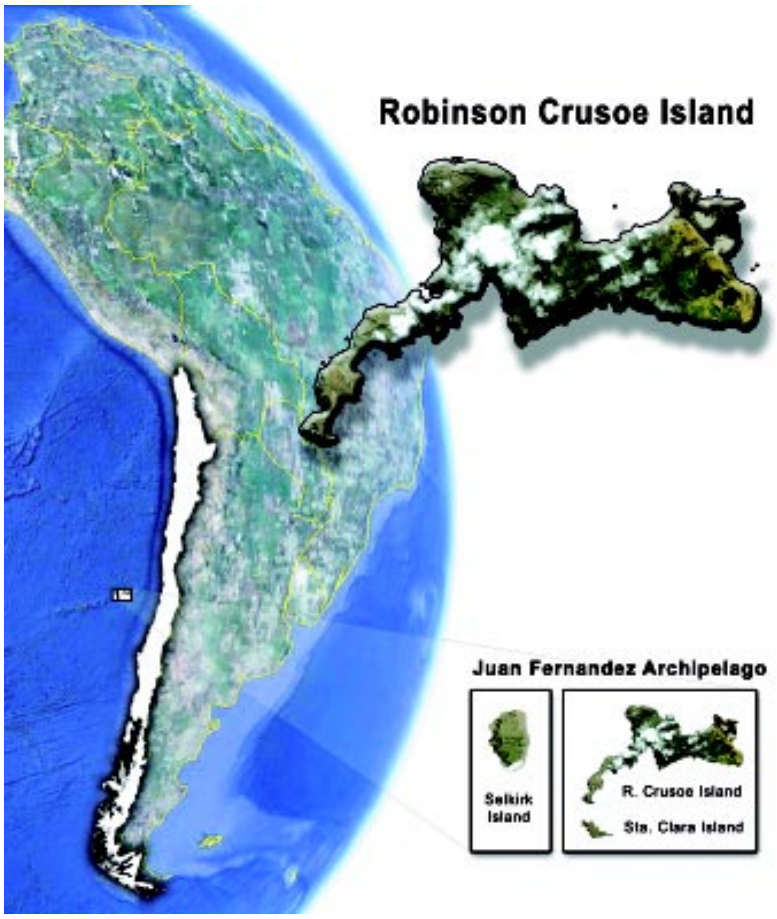
al., 2006; Aragón *et al*., 2008; Díaz *et al*., 2014).

Among dental phenotype is Carabelli's cusp, located on the mesiopalatine surface of the first superior molar. This is a high-frequency trait among populations of Caucasian origin (Bailit, 1975; Palomino; Villavicencio *et al*. 1996; Rodriguez & Gavilanes, 2002; Rothhammer & Llop).

Shovel-shaped incisor, a morphological variation on the palatine face of the superior incisor with variable expression, is a trait of Asian origin. It appears with high frequency among Amerindian and Eskimo, as well as Japanese and Chinese, populations (Krauss *et al*., 1969; Bailit; Palomino *et al*.).

The sixth cusp is sometimes present in the inferior molars of Asian and American Indigenous populations, and the seventh or intermedium cusp is characteristic of the inferior molars of African populations (Kraus *et al*.; Rothhammer & Llop; Delgado-Burbano, 2007)

The present study was carried out in Chilean Robinson Crusoe Island, which is part of Juan Fernandez Archipelago, located 677 kilometers from central continental Chile (Fig. 1). Robinson Crusoe is a unique island, with 633 inhabitants according to the national population census (INE, 2002).



The population's genetic structure is semi-isolated, small, and of recent origin. The archipelago was discovered in 1574 and colonized in 1750; however, it was depopulated in 1850. The most recent colonization occurred at the end of the 19th century, with 8 families led by Swiss Baron Alfred von Rodt (Vicuña Mackenna, 1883; Ruh, 1979). Of the current population of children, 77% have at least one of the colonizer's last names, consistent with the high degree of consanguinity of the island´s population (Villanueva *et al*., 2014).

The present research group has described a high prevalence of Specific Language Impairment (SLI; SLI, MIM 602081) in this population (Villanueva *et al*., 2008), at a rate seven times higher than those of the continental Chilean population (De Barbieri 1999) and five times higher than those of the global population (Tomblin *et al*., 1997). As this language pathology has a genetic basis, it may be associated with a founder effect, that is, the genetic homogeneity that develops in isolated locations with low population variance (Cavalli-Sforza & Bodmer; Villanueva *et al*., 2010, 2011).

The objective of this study is to estimate the genetic component and degree of miscegenation of the population of Robinson Crusoe Island, using dental morphological markers. The results will allow for a more detailed analysis of the genetic and environmental contribution to the risk for the disease in this population.

Fig. 1. Robinson Crusoe Island, geographic area studied (33° .37´ 56´´ S.; 78° .49´ 45´´ W). This Island belongs governmentally to the 5th region of Chile

MATERIAL AND METHOD

The socioeconomic conditions of the island inhabitants are relatively homogenous in terms of education level and quality of life. According to data from the national population census, inhabitants are largely fishermen and belong to a middle- and lower-middle-class socioeconomic level (INE).

After signing informed consent, there were created dental casts of the superior and inferior teeth of the universe of island children aged 3 to 15 years (n= 128). Each child was identified within the extensive genealogy previously created.

It was recorded the following dental morphological characteristics of the teeth: shovel-shaped superior incisor, Carabelli's cusp on the superior molars, sixth and seventh cusps on the inferior molars. The evaluation criteria are described in Kraus *et al*.

The estimation of genetic mixture was determined based on frequency of shovel-shaped incisor and Carabelli's cusp, using the Berstein method and assuming a population of mixed Amerindian and Caucasian origin (Cavalli-Sforza & Bodmer).

The data were compared using the Chi square method and Student t-test.

The study was approved by the Ethics Committee of the School of Medicine and undertaken as part of the UCHILE DID TNAC 01-02/01 and UCHILE DI MULT 05/05-02 projects funded by the University of Chile. The Municipality of Juan Fernandez Archipelago signed a legal contract with the University, permitting the realization of this project.

RESULTS

The universe of preschool- and school-aged children on Robinson Crusoe Island had an average age of 9.5 years ±3.2 for males and 9.5±3.1 for females, with no significant differences of age by sex (t= 0.07; p= 0.94). Type of dentition was mainly mixed (58.6%), with no differences by sex (x_= 3.04; df =3; p= 0.52) (Table I).

Of the 128 children studied, 61.7% had Carabelli's cusp, and 9.4% had shovel-shaped incisor, with no differences by sex (x_= 1.98; df= 3; p= 0.79) (Table II).

Sixth cusp was found in 2.3% of subjects with erupted inferior first molars, independent of the sex. The seventh cusp was absent in all Island children.

Table III shows the frequencies of shovel-shaped incisor and Carabelli's cusp for the Island population in comparison with estimated frequencies for larger racial groups and various continental Chilean populations. Shovel-shaped incisors were found at significantly lower frequency in comparison to the other populations (x_= 69.17; p= 0.0001), and Carabelli's cusp is common among Islanders. Carabelli's cusp is found at high frequency in mainly Caucasian populations and low frequency in Asian and Indigenous populations. The frequency found for Island children is significantly higher (x_= 201.5; p= 0.0001) than for children attending public school in Santiago and southern Chile. Students atprivate schools, especially in southern Chile, show values closer to those of the children of the island (Palomino *et al.*).

The ethnic composition of the populated was calculated using the Bernstein method for markers Carabelli's cusp and shovel-shaped incisor. The degree of miscegenation estimated, which represents the average of the two markers, indicates that the Amerindian component of the population of Robinson Crusoe Island was 4.34%. This value is compared with those obtained for the continental Chilean population and larger Caucasian and Amerindian racial groups in Table III.

Table I. Distribution of the school- and preschool-aged population of Chilean Robinson Crusoe Island by sex and type of dentition

	Males		Females		Total	
	n	%	n	%	n	%
Stage	11	17.2	10	15.6	21	16.4
1st Stage Mixed	18	28.1	23	35.9	41	32.0
2nd Stage Mixed	21	32.8	13	20.3	34	26.6
Permanent	14	21.9	18	28.1	32	25.0
Total	64	100	64	100	128	100

Table II. Dental morphological markers in the preschool- and school-aged population of Chilean Robinson Crusoe Island, by sex.

	Males		**Females**		**Total**	
	n	**%**	**n**	**%**	**n**	**%**
Carabelli's cusp	38	59.4	41	64.1	79	61.7
Shovel-shaped incisor	8	12.5	4	6.3	12	9.4
Sixth cusp	2	3.1	1	1.6	3	2.3
Seventh cusp	0	0.0	0	0.0	0	0.0

Table III. Frequency of dental morphological characteristics in the Chilean Robinson Crusoe Island, in continental Chile, and in larger racial groups.

	Carabelli's cusp	**Shovel-shaped incisor**	**GENETIC MIX**
Populations	%	%	%
Robinson Crusoe Island [1]	**61.7**	**9.4**	**4.3**
Caucasian [2]	59.5	12.6	---
Amerindian [2]	7.2	88.6	---
Santiago public school [2]	15.5	41.7	70.7
Santiago private school [2]	24.7	19.5	37.8
Southern Chile public school [3]	23.7	62.4	66.6
Southern Chile Private school [3]	65.8	21.8	7.9

1= Data obtained in the present study.
2= Data published by Palomino *et al*. Morfología Dentaria en la Evaluación de la Etnicidad Poblacional. Odontología Chilena 43, 91-94, 1995.
3= Data published by Rothhammer & Llop, Poblaciones chilenas. Cuatro décadas de investigaciones bioantropológicas. Editorial Universitaria, Santiago, pp. 115-128, 2004

DISCUSSION

It was described a high prevalence of Specific Language Impairment in children on Robinson Crusoe Island, at rates well above those reported for populations in Santiago de Chile or Europe (Villanueva *et al*., 2008). This language disorder has been described as highly heritable, complex, and multifactorial (Newbury *et al*., 2010; Villanueva *et al*., 2011). Therefore, identifying the genetic components of this interesting population (gene pool) allows for a more detailed evaluation of the role of genetic and environmental contributions in this developmental disorder.

Current Chilean populations are descended from different origins and, therefore, their population structure and degree of genetic mixture cannot be assumed. It is known that the country's population is descended primarily from the mix of Amerindian inhabitants and European settlers that colonized the country after its discovery and conquest (Rothhammer & Llop). It is also known that the Amerindian populations were genetically homogeneous but that the proportion of European or African genetic contribution varies within the population (Salzano & Callegari-Jacques, 1988). In Chile, the impact of African genes is quite low (Rothhammer & Llop).

Chilean Robinson Crusoe Island is highly endogamic, with a recent and known origin, with little prior immigration. The names of 77% of current children remain the same as those of the founding families (Villanueva *et al*., 2008). The families that colonized the island originate from a rural zone in the center of Chile, descended primarily from Spanish settlers with a relatively low indigenous component. Few of the settlers had other European backgrounds including German, French, and Swiss.

There are various methods for establishing the ethnic composition of a population, each of which has been demonstrated to be useful and to provide equivalent results.

Dental markers have been widely used because of their stable morphology over the lifespan. Studies of intra- and interpopulation dental morphological variability in groups with varying levels of indigenous American genetic contribution have concluded that philogenetic comparisons based on dental morphology are a useful indicator of biological distance (Palomino *et al*.).

In the Robinson Crusoe Islanders, it was analyzed four markers that are clearly different among racial groups. Out results show that the frequency of indigenous genes in the population is low (4%) and that there is no presence of African genes, which is consistent with historical and demographic data regarding the origin of this population (Vicuña Mackenna; Arana, 2010).

Furthermore, studies of schoolchildren in Santiago and Southern Chile show that dental morphology is a good phenotypic marker for estimating the degree of mixture within the population, with results equivalent to those obtained using serological markers in the same individuals. Populations of private and public school students in Santiago evaluated from a socioeconomic and ethnic perspective showed lower levels of Amerindian component than their public school counterparts (Valenzuela *et al*., 1987). In the rural zone of Southern Chile, especially in the Lake District, there has been a recent immigration of individuals from Germany resulting in relatively closed groups with respect to the surrounding resident population, and especially

exclusive of the numerous indigenous settlements. Currently, the public schools in the zone are attended by both indigenous and non-indigenous students from the mixed Chilean population, while the private schools of German origin are attended mainly by students from the colonizing German families. Studies have shown a similar pattern for school populations in Santiago (Rothhammer & Llop).

The high frequency of Carabelli's cusp and low frequency of shovel-shaped incisor described in the present work allows us to conclude that the population of Robinson Crusoe Island has remained semi-isolated, with scarce genetic variability. The population has maintained the characteristics of the founding families originating mainly from Europe. The absence of the seventh cusp rules out the existence of African genes.

Significantly, the frequency of shovel-shaped incisor on the Island is similar to that described for Caucasians and lower than that described for indigenous South American populations, while Carabelli's cusp is more common than in urban continental Chilean populations. Populations influenced by racial mixing and less isolation generally show higher frequencies of traits of Caucasoid origin and decreased frequencies of Amerindian traits (Palomino *et al*.; Aragón *et al*.).

The frequency of Specific Language Impairment in the island (35%) (Villanueva *et al*., 2008) is higher than in English spoken children (Tomblin *et al*.) or continental Chile (De Barbieri *et al*., 1999), where rates are below 10%. This island population with a low frequency of indigenous genes can be considered to be mainly European in origin. Therefore, it would be useful to compare the genetic profiles of SLI cases reported in Europe with those in Robinson Crusoe Island.

In fact, four genes associated with spoken language disorder have been identified in European populations (for more detail see Newbury *et al*.).

In the population of Robinson Crusoe, a marker has been identified, along with five zones highly associated with the pathology on other chromosomes (Villanueva *et al*., 2010, 2011). Among these, a zone in chromosome 7 coincides with the location of genes identified as an area of susceptibility for SLI, autism, and vocabulary and grammar disorders. Furthermore, it coincides with the location of the gene FOXP2 recognized as an area of susceptibility for speech and language disorders over a decade ago, and described as a biological modulator of the evolution and development of language in the human species (Konopka *et al*., 2009).

In conclusion, this study confirmed a scarce indigenous genetic contribution to the semi-isolated Chilean population of Robinson Crusoe Island, according to dental morphological markers. This is the first step in describing the genetic profile of this population, in which a high prevalence of language disorder is reported. This characterization paves the way for comparisons of the genetic variants underlying the language pathologies described in European populations with findings for this Chilean island population.

ACKNOWLEDGEMENTS

The authors are extremely grateful to the inhabitants of Robinson Crusoe Island who have agreed to participate in this study. We would also like to thank to the Mayor of the Ilustre Municipalidad de Juan Fernandez, for his infinite assistance in the development of this research.

FUNDING. The characterization of the Robinson Crusoe population was funded by two grants of Vicerrectoría de Investigación, Universidad de Chile [grant number UCHILE DID TNAC 01-02/01], [grant number UCHILE DI MULT 05-05/02].

VILLANUEVA, P.; QUEVEDO, M.; DE BARBIERI, Z.; PIÑEIRO, S.; HERRERO, C.; FERNÁNDEZ, M. A. & PALOMINO, H. Marcadores morfológicos dentarios en la estimación de la etnicidad poblacional de la Isla Robinson Crusoe. ***Int. J. Morphol., 33(2)***:538-543, 2015.

RESUMEN: La isla chilena Robinson Crusoe es un semiaislado geográfico de alta consanguinidad. Su población actual de 633 habitantes proviene de la última colonización ocurrida a finales del siglo XIX y pocas migraciones posteriores, en quienes recientemente se ha descrito una alta incidencia de trastorno de lenguaje. Este estudio estimó el componente genético y grado de miscegenación de la población isleña usando marcadores morfológicos dentarios. Se estudió al universo de niños isleños (n= 128, 3 a 15 años de edad) con exámenes clínicos, modelos dentales y ubicación de cada individuo en genealogías extensas confeccionadas previamente. La frecuencia de Tubérculo de Carabelli fue 61,7%, Diente en Pala 9,4%, tubérculo sexto 2,3% y ausencia del rasgo tubérculo séptimo, lo que concuerda con una población eminentemente caucásica. El grado de miscegenación estima que el componente amerindio de esta población es de 4,3%, que también se evidencia al analizar las genealogías extensas originadas por los colonizadores. La descripción del perfil genético de esta población, donde se han reportado altas prevalencias de trastorno de lenguaje, permitirá comparar con las variantes genéticas subyacentes a esta patología descritas para poblaciones europeas.

PALABRAS CLAVES: Marcadores morfológicos dentarios; Tubérculo de Carabelli; Incisivo en forma de pala; Isla Robinson Crusoe.

REFERENCES

Aragón, N.; Bastidas, C.; Bedón, L.K.; Duque, P.; Sánchez, M.; Rivera, S.; Triana, F.; Bedoya, N. & Moreno, F. Rasgos morfológicos dentales coronales en dentición temporal y permanente: Distancia biológica entre tres grupos indígenas del Amazonas Colombiano. *Rev. Odontol. Mex., 12(1)*:13-28, 2008.

Arana, P. M. *La isla de Robinson Crusoe*. Valparaiso, Ediciones Universitarias de Valparaiso, Pontificia Universidad Católica de Valparaíso, 2010.

Bailit, H. L. Dental variation among populations. An anthropologic view. *Dent. Clin. North Am., 19(1)*:125-39, 1975.

Bollini, G. A.; Rodríguez-Flórez, C. D.; Colantonio, S. E.; Méndez, M. G. Morfología dental de una serie prehistórica de Araucanos provenientes de la Patagonia Argentina y su relación biológica con otras poblaciones prehistóricas argentinas y del mundo. *Int. J. Morphol., 24(4)*:705-12, 2006.

Cavalli-Sforza, L. L. & Bodmer, W. F. *The genetics of Human Populations*. San Francisco, W. H. Freeman & Co. Dover Publications, 1971.

De Barbieri, Z.; Maggiolo, M. & Alfaro, S. Trastornos de la comunicación oral en niños que asisten a control de salud en un consultorio de atención primaria. *Rev. Chil. Pediatr., 70(1)*:36-40, 1999.

Delgado-Burbano, M. E. Population affinities of African Colombians to Sub-Saharan Africans based on dental morphology. *Homo, 58(4)*:329-56, 2007.

Díaz, E.; García, L.; Hernández, M.; Palacio, L.; Ruiz, D.; Velandia, N.; Villavicencio, J. & Moreno, F. Frequency and variability of dental morphology in deciduous and permanent dentition of a Nasa indigenous group in the municipality of Morales, Cauca, Colombia. *Colomb. Med. (Cali), 45(1)*:15-24, 2014.

Instituto Nacional de Estadísticas (INE). Chile: Censo Nacional de Población y Vivienda 2002. Santiago, Gobierno de Chile, 2002. Disponible en: http://espino.ine.cl/cgibin/RpWebEngine.exe/PortalAction?&MODE=MAIN&BASE=CPCHL2KCOM&MAIN=WebServerMain.inl

Konopka, G.; Bomar, J. M.; Winden, K.; Coppola, G.; Jonsson, Z. O.; Gao, F.; Peng, S.; Preuss, T. M.; Wohlschlegel, J. A. & Geschwind, D. H. Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature, 462(7270)*:213-7, 2009.

Kraus, B. S.; Jordan, R. E. & Abrams, L. *Dental anatomy and occlusion*. Baltimore, Williams & Wilkins Company, 1969.

Newbury, D. F.; Fisher, S. E. & Monaco, A. P. Recent advances in the genetics of language impairment. *Genome Med., 2(1)*:6, 2010.

Palomino, H.; Chakraborty, R. & Rothhammer, F. Dental morphology and population diversity. *Hum. Biol., 49(1)*:61-70, 1977.

Palomino, H. The Aymara of western Bolivia: III. Occlusion, pathology, and characteristics of the dentition. *J. Dent. Res., 57(3)*:459-67, 1978.

Rodríguez, C. D. & Gavilanes, D. M. Morfología dental de la población enterrada en el cementerio prehispánico de Obando al norte de Valle del Cauca entre los siglos VIII y XIII d.C. *Rev. Fed. Odontol. Colomb., 63*:100-14, 2002.

Rodríguez-Flórez, C. D. La antropología dental y su importancia en el estudio de los grupos humanos prehispánicos. *Rev. Antropol. Exp., 4*:1-7, 2004.

Rothhammer, R. & Llop, E. *Poblaciones chilenas: Cuatro décadas de investigaciones bioantropológicas*. Santiago, Editorial Universitaria, 2004.

Ruh, M. Alfredo Von Rodt, Subdelegado de Juan Fernández. Santiago, Depto. de Ciencias Antropológicas y Arqueológicas Universidad de Chile, 1979.

Salzano, F. M. & Callegari-Jacques, S. M. *South American Indians: A case study in evolution*. Oxford, Clarendon Press, 1988.

Tomblin, J. B.; Records, N. L.; Buckwalter, P.; Zhang, X.; Smith, E. & O'Brien, M. Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res., 40(6)*:1245-60, 1997.

Valenzuela, C. Y.; Acuña, M. P. & Harb, Z. Gradientes sociogenéticos en la población chilena. *Rev. Med. Chil., 115(4)*:295-9, 1987.

Vicuña Mackenna, B. *Juan Fernández. Historia verdadera de la isla de Robinson Crusoe.* Tomos I y II. Valparaíso, Ediciones Universitarias de Valparaíso, 1883.

Villanueva, P.; de Barbieri, Z.; Palomino, H. M. & Palomino, H. Alta prevalencia de trastorno específico de lenguaje en isla Robinson Crusoe. Un probable efecto fundador. *Rev. Med. Chil., 136(2)*:186-92, 2008.

Villanueva, P.; Jara, L. & Palomino, H. Association of D16S515 microsatellite with specific language impairment on Robinson Crusoe Island, an isolated Chilean population: a possible key to understanding language development. *Hum. Biol., 82(4)*:395-408, 2010.

Villanueva, P.; Newbury, D. F.; Jara, L.; De Barbieri, Z.; Mira, G.; Palomino, H. M.; Fernández, M.A.; Cazier, J. B.; Mónaco, A. & Palomino, H. Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. *Eur. J. Hum. Genet., 19(6)*:687-95, 2011.

Villanueva, P.; Fernández , M. A.; De Barbieri, Z. & Palomino, H. Consanguinity on Robinson Crusoe Island, an isolated Chilean population. *J. Biosoc. Sci., 46(4)*:546-55, 2014.

Villavicencio, J. A.; Fernández, M. A. & Magaña, L. *Evolución de las Piezas Dentarias.* En: Villavicencio, J. A.; Fernández, M. A. & Magaña, L. (Eds.). Ortopedia Dentofacial: Una Visión Multidisciplinaría. Caracas, Actualidades Médico Odontológicas Latinoamérica, 1996. pp.77-87.

Weiss, K. M. *Genetic variation and human disease: Principles and evolutionary approaches*. Cambridge, Cambridge University Press, 1995.

Correspondence to:
Pia Villanueva
Sergio Livingstone Pohlhammer 943
Independencia
ZC: 8380492
Santiago
CHILE

Email: piavilla@u.uchile.cl

Received: 30-11-2014
Accepted: 19-03-2015

DISCUSIÓN

GENÉTICA MOLECULAR

La Isla Robinson Crusoe es un semiaislado geográfico, colonizado a fines del siglo XIX en base a 8 familias fundadoras. La estructura genética de la población actual es considerada muy joven y con muy poca variabilidad poblacional.

En los niños de esta población se ha reportado una alta prevalencia de TEL, probablemente consecuencia de un efecto fundador. La frecuencia de TEL en los niños originarios isleños es de 35%, cifra significativamente mayor a lo observada en otras poblaciones tanto chilenas (4%) como extranjeras (entre 2% y 8%) y a la frecuencia en los niños no isleños que habitan la isla temporalmente (3,8%). La alta prevalencia de TEL en el Archipiélago, sugiere la existencia de un efecto fundador para la susceptibilidad del TEL que podría ser consecuencia de la deriva génica propia de poblaciones pequeñas y aisladas (Villanueva et al. 2008).

Las características previamente descritas para este aislado geográfico, hacen que esta población sea única en el mundo y muy adecuada para investigar la etiología molecular del TEL y para la identificación de genes que participen en el desarrollo del lenguaje.

Los resultados de este estudio se basan en la secuenciación exómica completa de cinco individuos, habitantes de la isla Robinson Crusoe, con diagnóstico de TEL.

Se identificó una variante de código no-sinónima / heterocigótica en el gen NFXL1, que es compartida por tres de los cinco individuos secuenciados. El análisis de asociación realizado en la cohorte ampliada de Robinson Crusoe, demuestra que esta variante se presenta con una frecuencia significativamente aumentada en los sujetos isleños con trastorno de lenguaje en comparación a los que presentan desarrollo típico de lenguaje (P = 0.0002).

Al realizar la secuenciación del gen NFXL1 en la cohorte de 117 probandos de TEL ingleses (Reino Unido), se identificaron cuatro individuos con variantes heterocigóticas, tres individuos portadores de cambios no_sinónimos y un individuo con una variante novel.

Basados en los resultados anteriores, se concluye que la variante en el gen NFXL1 podría contribuir al riesgo genético de los trastornos de lenguaje. Se propone que dichos cambios presentarían un modelo de herencia complejo.

CONSANGUINIDAD

Pese a que otras poblaciones isleñas chilenas, como la población de la Isla de Pascua, cuentan con acabados estudios de consanguinidad (Cruz-Coke, 1989), la isla de Robinson Crusoe sólo cuenta con una referencia a su estructura poblacional recopilada a principios de los años 60 (Orellana et al, 1975).

Es por lo anterior que se realizó un análisis de la estructura poblacional, en base al estudio del nivel de consanguinidad y tipos de "matrimonios/uniones con descendencia" más frecuente.

Todos los matrimonios registrados en la isla de Robinson Crusoe, desde la última colonización hasta el año 2000 (417 matrimonios en total) fueron incluidos en genealogías extendidas de las familias fundadoras y analizados.

Los resultados demuestran un alto grado de consanguinidad, sustentado por los matrimonios entre primos primeros, así como por los matrimonios entre primos segundos, siendo estos los tipos de matrimonios de mayor frecuencia en esta población. La tasa consanguinidad fue determinada en 14.9% y el promedio de coeficiente de endogamia en α = 54.05 x 10^{-4}. Comparado con otra poblaciones chilenas (Pinto-Cisternas, 2004), este valor es superior al descrito para población Atacameña [α = 34,4 x 10^{-4} (Chapin y cols,1976)], para población Mapuche [α = 1,0 x10^{-4} (Rothhammer y cols,1971)] y para población Aymara [α =

0.011x 10 $^{-4}$ (Cruz Coke,1966)]. Además, es menor al coeficiente de endogamia reportado en las poblaciones rurales de Caleu (α = 60,4 x10 $^{-4}$) por Blanco y Covarrubias (1971) y de Olmué (α = 65,4 x10 $^{-4}$), por Quezada y Barrantes (1973).

Las características únicas de la isla - población actual pequeña, originada de pocas familias, aislada geográficamente y con poca deriva génica - podrían explicar la consanguinidad aumentada.

El desarrollo de este trabajo permitió, por primera vez, que los habitantes de la isla cuenten con información detallada de la estructura poblacional e información cualitativa del grado de parentesco que presentan. Por lo anterior, las autoridades de la isla han mostrado gran interés en el desarrollo de esta investigación, brindando el apoyo logístico necesario para el cumplimiento de los objetivos.

Concluir un alto nivel de endogamia en una población pequeña y aislada como es la de la isla Robinson Crusoe es muy interesante, debido a las importantes repercusiones médico-genéticas para los descendientes de matrimonios consanguíneos, considerando que en la población de Robinson Crusoe hemos descrito alta prevalencia de trastorno de lenguaje.

MISCEGENACIÓN

La población chilena actual presenta diferentes orígenes, por lo que su estructura y grado de mezcla no es estable. Se sabe que la población actual desciende principalmente de una mezcla entre habitantes originarios de America y colonizadores europeos, que poblaron el territorio continental después del descubrimiento y durante la conquista (Rothhammer & Llop 2004). Se sabe también que las proporciones de contribución genética de las poblaciones europeas y africanas varian entre los distintos grupos sudamericanos (Salzano & Callegari-Jacques 1988). En Chile, el impacto de genes africanos es muy bajo (Rothhammer & Llop 2004).

Para determinar el grado de miscegenación de la población se utilizaron marcadores morfológicos dentarios. En base a análisis de modelos dentarios se determinó la frecuencia de marcadores morfológicos dentarios en el universo de la población isleña (Kraus et al. 1969). La frecuencia del marcador dentario Diente en Pala fue de 9,4%, mientras que la frecuencia de Tubérculo de Carabelli fue de 61,7%. Las frecuencias de ambos marcadores dentarios concuerdan con una población eminentemente caucásica. El grado de miscegenación estimado por el método de Bernstein, indica que el componente amerindio de la población es menor de un 5%, lo que también se evidencia al analizar las genealogías extensas que a través de 5 generaciones se llega al grupo colonizador que en la mayoría de los casos corresponde a sujetos de procedencia europea o de la zona rural de Chile. Este estudio permitirá aplicar normas y/o realizar comparaciones con

grupos caucásicos, considerando que se trata de una población donde se ha descrito una alta prevalencia de trastorno de lenguaje, facilitando así la comparación de variantes genéticas subyacentes a esta patología y que han sido identificadas en poblaciones europeas.

REFERENCIAS BIBLIOGRÁFICAS

AGUADO G. Caracterización del Trastorno Específico del Lenguaje. Conceptos Básicos. En: Aguado G. Trastorno Específico del Lenguaje. Málaga: Aljibe, 2000;15-30.

ARANA P. La isla de Robinson Crusoe. Ed. Universitarias de Valparaíso. Valparaíso. 394pp. 2010

BARTLETT C, FLAX J, LOGUE M, VIELAND V, BASSETT A, TALLAL P, BRZUSTOWICZ L. A major susceptibility locus for Specific Language Impairment is located on 13q21. Am J *Hum Genet* 2002;71:45-55

BISHOP DV, NORTH T, DONLAN C. Genetic basis of Specific Language Impairment: evidence from a twin study. *Dev Med Child Neurol* 1995;37:56-71.

BLANCO R, COVARRUBIAS E. Estudio genético en Caleu. *Rev Med Chile* 1971;99:139-144

BRESCIA DE VAL M. Mares de Leyenda. Santiago: Marenostrum 1982

BRINCK G. Memorias insulares. Archipiélago Juan Fernández. Ed. Azapa. 2005

CAVALLI-SFORZA LL, BODMER WF. The genetics of human populations. San Francisco. WH Freeman & Co. Dover Publications 1971

CHAPIN J, ROTHHAMMER F, BLANCO R, PALOMINO H. Consanguinity in Chilean Populations. *Rev. Med Chile* 1976:104: 202- 206

CONTI – RAMSDEN G. Genes, language and specific language impairment (SLI). *First Lang.* 1997;17:321-332

CONAF Corporación Nacional Forestal. 1976. Plan de manejo, Parque Nacional Juan Fernández, Documento Técnico de Trabajo; 22:61

CONAF : Corporación Nacional Forestal, 2009. Priorización y definición de objetos de conservación y líneas de acción para la elaboración del plan de conservación de la Biodiversidad Terrestre del Archipiélago de Juan Fernández. CONAF, Santiago, 24 pp

CRUZ-COKE R, CRISTOFFANINI AP, ASPILLAGA M, BIANCANI F. Evolutionary forces in an envionmental gradient in Arica, Chile. *Hum Biol* 1996;38:422-438

DE BARBIERI Z, MAGGIOLO M, ALFARO S. Trastornos de la comunicación oral en niños que asisten a control de salud en un consultorio de atención primaria. *Rev Chil Pediart*, 1999;70:36-40

FALCARO FALCARO M, PICKLES A, NEWBURY DF, ADDIS L, BANFIELD E, ET AL. Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in specific language impairment. Genes Brain Behav 2008:7: 393–402.

FIGUEROA S. Por amor al vuelo. Mago editores, Santiago. 412 pp 2006.

FISHER SE, VARGHA-KHADEM, WATKINS KE, MONACO AP, PEMBREY ME. Localisation of a gene implicated in a severe speech and language disorder. *Nat genet* 1998;18:168-170.

INE, Instituto Nacional de Estadísticas. Censo 2002 REDATAM. www.INE.cl (Consultado el 22 de julio de 2008)

KRAUS B.S, JORDAN R.E , ABRAMS L. Dental anatomy and occlusion. 1969 Baltimore:The Williams and OMIM. Online Mendelian Inheritance in Man. www.ncbi.nlm.nih.gov/OMIM/

LAI C, FISHER S, HURST J, LEVY E, HODGSON S, FOX M, et al. The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocation associated with speech and language disorders. *Am J Hum Genet* 2000;67:357-368

LAI C, FISHER S, HURST J, VARGA-KHADEM F, MONACO A. A forkhead-domain gene is mutated in severe speech and language disorder. *Nature* 2001;416:519-523

LAW J, BOYLE J, HARRIS F, HARKNESS A, NYE C. Screening for speech and language delay: a systematic review of the literatura. *Health Technol Assess* 1998;2:1-184

MARTÍNEZ L, HERRERA C, VALLE J, VASQUEZ M. Razonamiento analógico verbal y no verbal en niños con trastorno específico de lenguaje. *Rev Chil Fonoaud* 2002;3:5-24

MENDOZA E. Criterio de Identificación del TEL. En: Mendoza E, Trastorno Específico de Lenguaje. Madrid: Ed. Pirámide, 2001;45-66

MONACO, A. P. Multivariate Linkage Analysis of Specific Language Impairment (SLI). Ann Hum Genet.2007: 71:660-673

NARBONA J & CHEVRIE-MULLER C. El Lenguaje del Niño. Desarrollo Normal, Evaluación y Trastornos. Barcelona: Ed.Masson,1997;183-200

NEWBURY DF, FISHER SE, MONACO AP.Recent advances in the genetics of language impairment Genome Med. 2010: 26;2:6

OMIM. Online Mendelian Inheritance in Man. www.ncbi.nlm.nih.gov/OMIM/ (Consultado el 22 de julio de 2015).

ORELLANA, E. El desarrollo del lenguaje oral y sus implicancias en el aprendizaje lector. *Rev Pensam Educat* 1996;19:207-224

PINTO – CISTERNAS J. Consanguinidad de la población chilena. En Rothhammer F & Llop E. Poblaciones chilenas: cuatro décadas de investigación bioantropológica.Santiago: Ed Universitaria 2004

QUEZADA M, BARRANTES R. Análisis de la endogamia en una población chilena. *Rev Med Chil* 1973;101:227-230

ROTHHAMMER F, BENADO M, PEREIRA G. Variability of two dental traits in chilean indian and mixed populations. *Hum Biol* 1971;43:317-319

ROTHHAMMER R & LLOP, E. Poblaciones chilenas cuatro décadas de investigaciones bio-antropológicas. Editorial Universitaria, Santiago. 2004

RUH M. ALFREDO VON RODT, Subdelegado de Juan Fernández. En Orellana M, Medina A, Morel T, Ruh M, Hernández R y Monleón J. Las islas de Juan Fernández. Santiago: Departamento de Ciencias Antropológicas y Arqueológicas Universidad de Chile, 1975

SCHNYDER O. El diario de Alfredo de Rodt. Subdelegado e Inspector de Colonia de las islas Juan Fernández. Editorial Camberland, isla Robisnon Crusoe, 440 pp 2005

SHRIBERG LD, TOMBLIN JB, MCSWEENY JL. Prevalence of speech delay in 6 year-old children and co morbidity with language impairment. *J Speech Lang Hear Res* 1999;42:1461-1481

STROMSWOLD K. The genetics of spoken language disorders. *Hum Biol* 1998;70:297-324

STROMSWOLD K. The heredability of language: a review and metaanalysis of twin, adoption, and linkaje studies. *Language* 2001;77:647-723

TAKAHASHI D. Insearch of Robinson Crusoe. Edit. Cooper Square Press, New York, 229 pp. 2002

THE SLI CONSORTIUM. A genowide Scan Identifies Two Novel Loci Involved in Specific Language Impairment. *Am J Hum Genet* 2002;70:384-398.

THE SLI CONSORTIUM. Highly significance linkage to SLI1 locus in an expanded sample of individual affected by specific language impairment. *Am J Hum Genet* 2004;74:1225-1238.

TOMBLIN JB & BUCKWALTER PR. Heritability of poor language achievement among twins. *J Speech Lang Hear Res* 1998;41:188–199.

TOMBLIN JB, RECORDS NL, BUCKWALTER PR, ZHANG X, SMITH E, O'BRIEN M. Prevalence of specific language impairment in kindergarten children. *J Speech Lang. Hear Res* 1997; 40:1245-1260.

TOMBLIN JB, SHRIBERG L, NISHIMURA C, ZHANG X, MURRAY J. Association of specific language impairment with loci 7q3. Simposium of Communication Disorders; 2000, Wisconsin, USA.

VICUÑA MACKENNA B. Juan Fernández. Historia verdadera de la isla de Robinson Crusoe. Santiago, Rafael Jover, 1883

VILLANUEVA P, DE BARBIERI Z, PALOMINO HM, PALOMINO H. Alta prevalencia de trastorno específico de lenguaje en isla Robinson Crusoe y probable efecto fundador. *Rev Med Chil* 2008;136:186-192

VILLANUEVA, P., JARA, L. & PALOMINO, H. Association of D16S515 microsatellite with specific language impairment on Robinson Crusoe Island, an isolated Chilean Population: a possible key to understanding language development. *Hum Biol* 2010: 82, 395–408.

VILLANUEVA, P., NEWBURY, D. F., JARA, L., DE BARBIERI, Z., MIRA, G., PALOMINO, H. M., FERNANDEZ, M. A. ET AL. Genome wide analysis of genetic susceptibility to language impairment in a isolated Chilean population. European Journal of Human Genetics 2011: 19, 687–695.