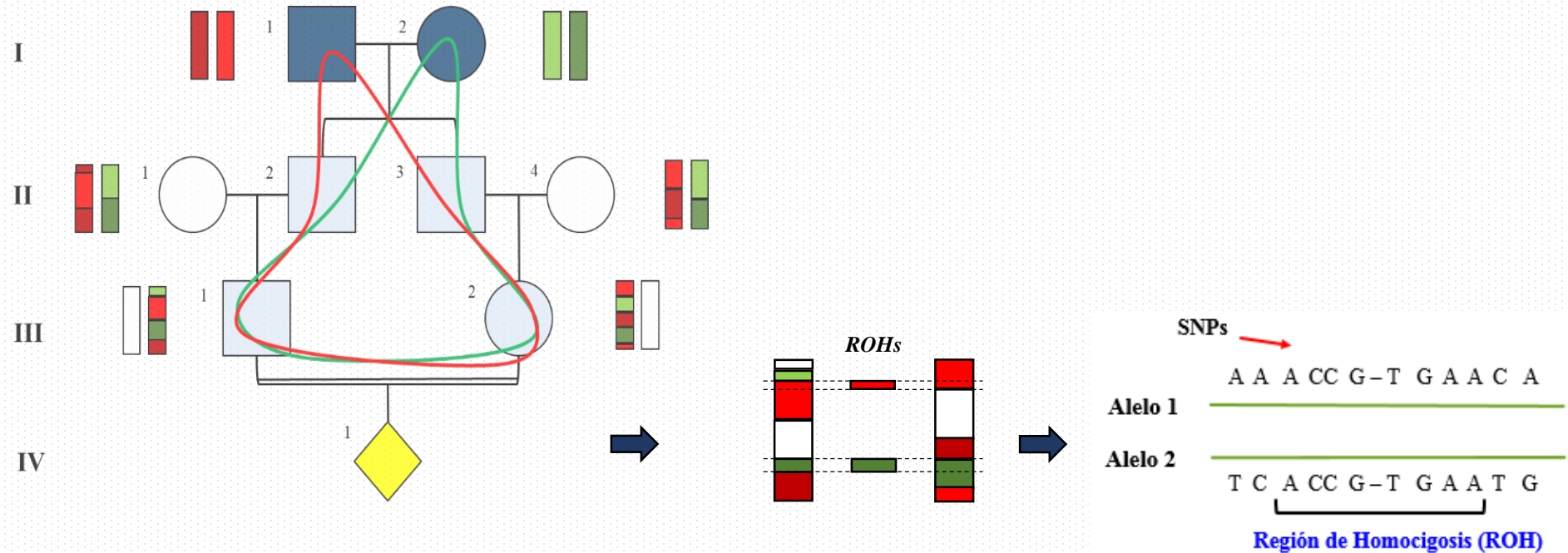




# Estimaciones del coeficiente de inbreeding genómico humano mediante la librería detectRUNs de R y otros programas bioinformáticos



**L. J. Sánchez-Martínez<sup>1</sup>, C.L. Hernández<sup>1</sup>, A. Sánchez<sup>1</sup>, R. Calderón<sup>1</sup>**

<sup>1</sup> Departamento de Biodiversidad, Ecología y Evolución. Facultad de Ciencias Biológicas  
Universidad Complutense de Madrid, UCM

# Índice

Introducción

Objetivos

Material y métodos

Resultados

Conclusiones

# El fenómeno de la consanguinidad humana

- La estructura marital entre parientes biológicos (**matrimonios consanguíneos**) ha sido uno de los comportamientos más usuales en la historia de la humanidad.

# El fenómeno de la consanguinidad humana

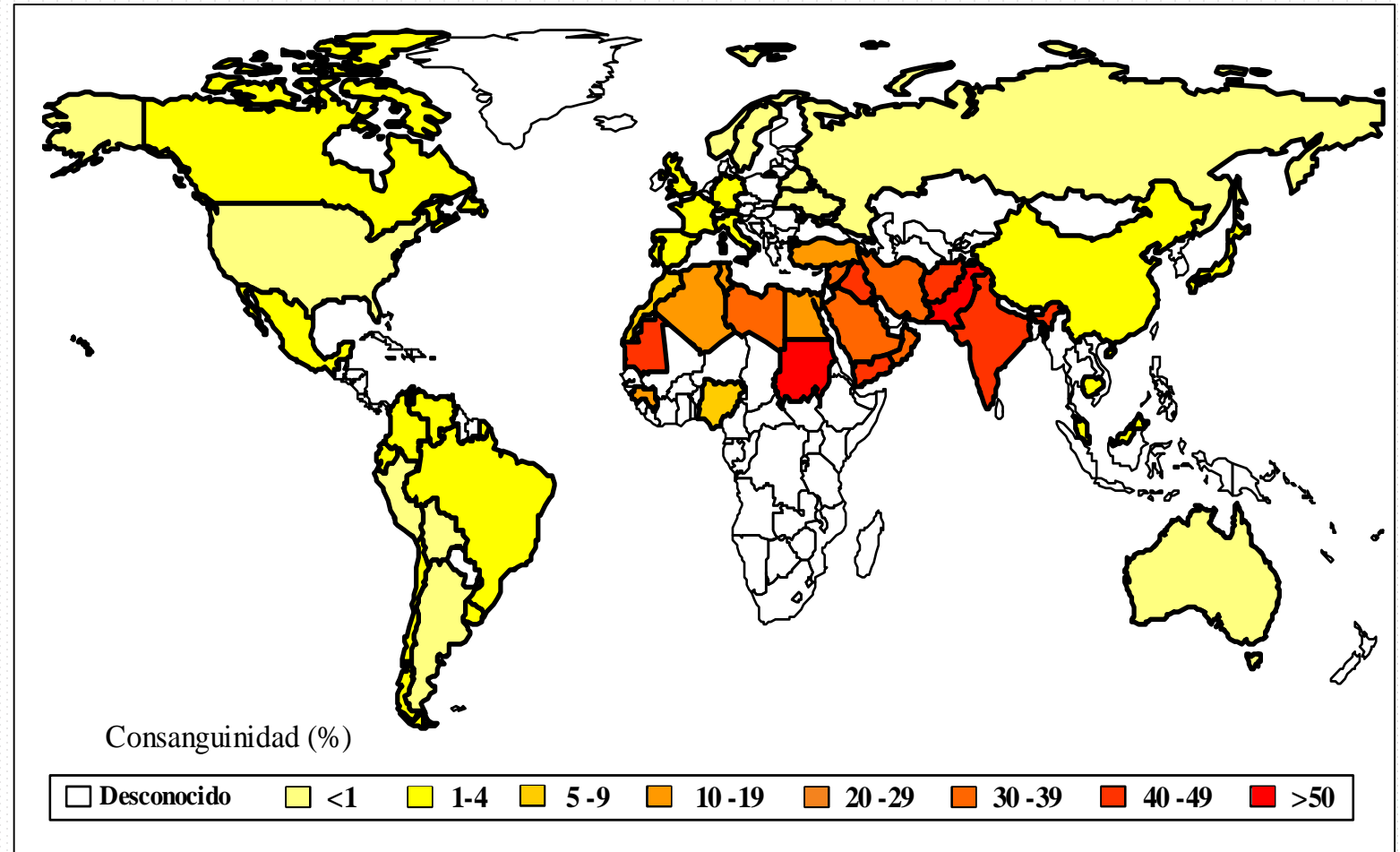
- La estructura marital entre parientes biológicos (**matrimonios consanguíneos**) ha sido uno de los comportamientos más usuales en la historia de la humanidad.
- **La consanguinidad**, constituye aun hoy un fenómeno ampliamente extendido entre diversas sociedades humanas.

Consanguinidad en la población (%)	Porcentaje de población total	Tamaño de la población (millones)
< 1	15,5	1068
1-9	43,9	3026
10-19	0,5	35
20-29	6,5	448
30-39	2,1	145
40-49	3,2	221
50+	3,3	227
Desconocido	25,0	1730

*Fuente: Global Consanguinity Website, [www.consang.net](http://www.consang.net)*

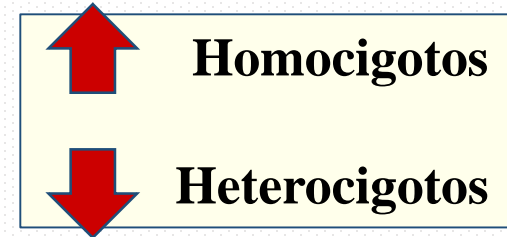
# El fenómeno de la consanguinidad humana

- La estructura marital entre parientes biológicos (**matrimonios consanguíneos**) ha sido uno de los comportamientos más usuales en la historia de la humanidad.
- **La consanguinidad**, constituye aun hoy un fenómeno ampliamente extendido entre diversas sociedades humanas.



# La consanguinidad y sus efectos biológicos

- A nivel de la población, la ocurrencia de emparejamientos (matrimonios) consanguíneos durante generaciones, puede alterar las frecuencias genotípicas respecto a las esperadas bajo equilibrio Hardy-Weinberg.



*El modelo Hardy-Weinberg describe la relación entre frecuencias alélicas y genotípicas en una población natural a lo largo del tiempo (en ausencia de procesos evolutivos)*

***Antropología Física (Genética), Genética Humana, Biomedicina, Genética de Poblaciones...***

1. Aumento de la ocurrencia de enfermedades recesivas (mendelianas) en los descendientes de padres emparentados biológicamente,
2. Aumento de la morbilidad y, una alta ocurrencia de enfermedades de etiología compleja,
3. Aumento de la mortalidad pre-reproductiva.

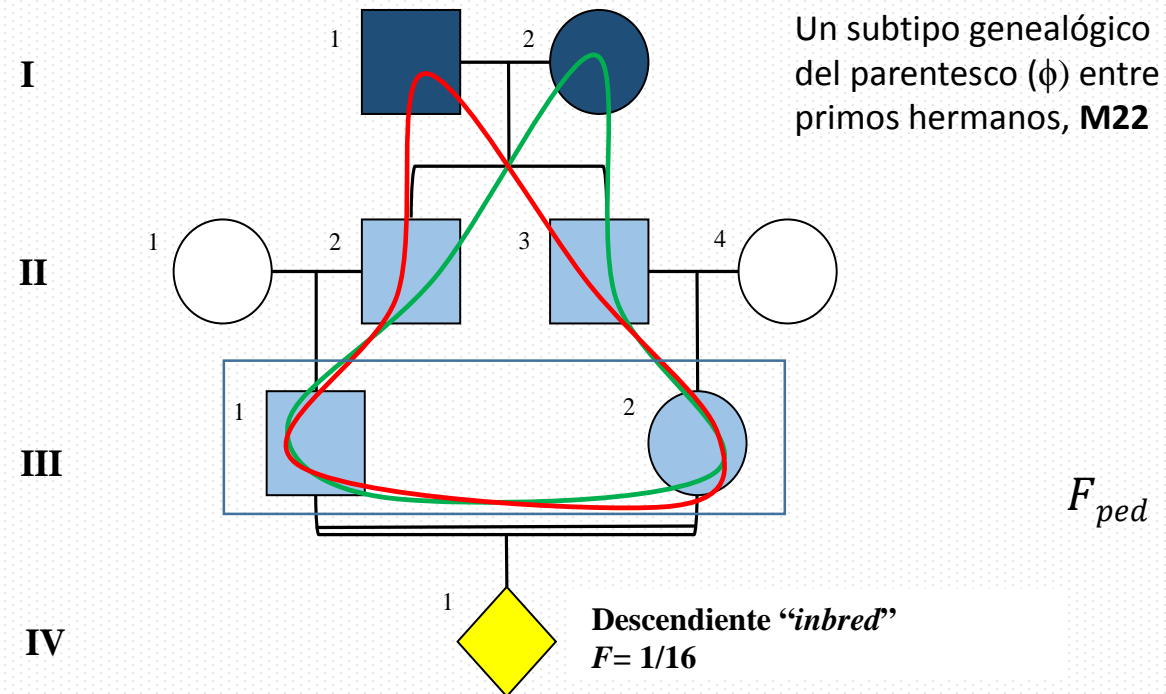
# Metodologías tradicionales de estima.

## El coeficiente de inbreeding basado en pedigrees, $F_{ped}$

El coeficiente  $F$ , hace referencia a la probabilidad de que un descendiente de padres emparentados biológicamente, reciba en un *locus* dado dos genes idénticos por descendencia (*IBD*)

### El análisis de pedigrees

- ☐ En la práctica es una información costosa de obtener,
- ☐ Habitualmente solo se recogen relaciones de parentesco tres o cuatro generaciones atrás,
- ☐  $F_{ped}$  es una expectativa del porcentaje real de genoma que es homocigoto.



$$F_{ped} = \sum_{i=1}^m \left(\frac{1}{2}\right)^k$$

# Metodologías genómicas para la estima del coeficiente de inbreeding, $F$

**Tecnologías de secuenciación de nueva generación**  
*(NGS, Next Generation Sequencing)*

**Genotipado masivo**  
*(Genome-Wide analysis, GW)*



# Metodologías genómicas para la estima del coeficiente de inbreeding, $F$

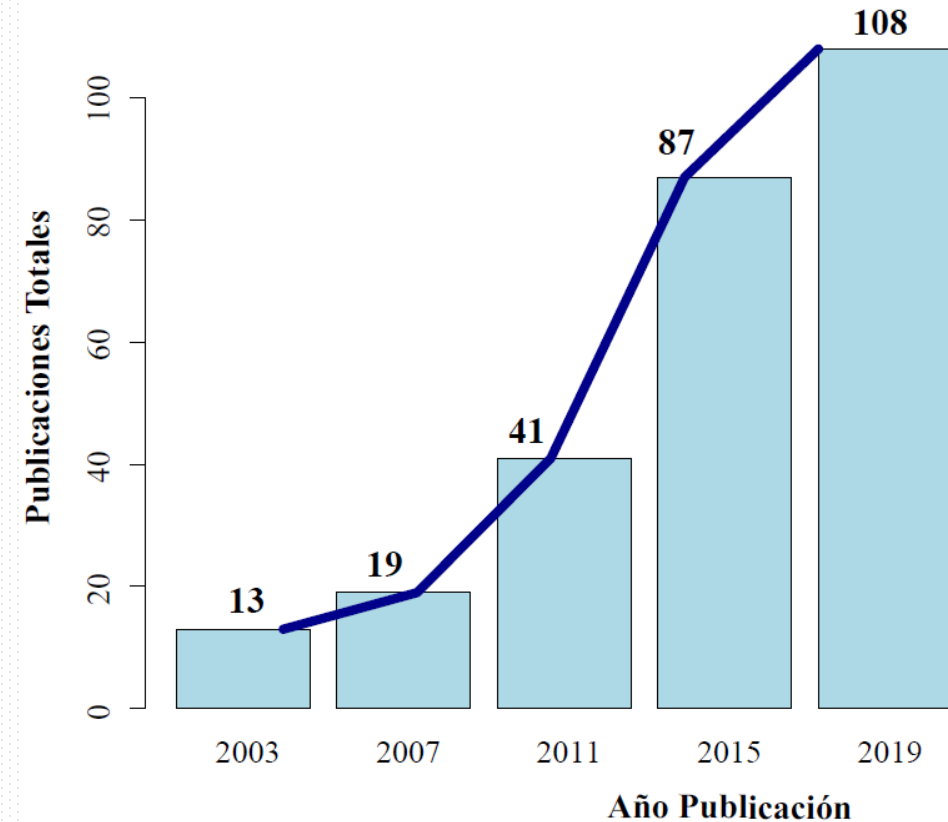
*Un cambio de paradigma a la hora de estudiar la consanguinidad humana*

- ✓ Proporción real del genoma en homocigosis,
- ✓ Capacidad de medir el efecto de relaciones de parentesco muy distantes temporalmente,
- ✓ El análisis de cualquier individuo, incluso en aquellos sin información de pedigrees,
- ✓ Conocer la posición exacta a nivel genómico y cromosómico de la homocigosis,
- ✓ Abaratamiento de los costes.

# Metodologías genómicas para la estima del coeficiente de inbreeding, $F$

*Un cambio de paradigma a la hora de estudiar la consanguinidad humana*

- ✓ Proporción real del genoma en homocigosis,
- ✓ Capacidad de medir el efecto de relaciones de parentesco muy distantes temporalmente,
- ✓ El análisis de cualquier individuo, incluso en aquellos sin información de pedigrees,
- ✓ Conocer la posición exacta a nivel genómico y cromosómico de la homocigosis,
- ✓ Abaratamiento de los costes.



*Términos de búsqueda: ROHs & Human Populations*

# Metodologías genómicas para la estima del coeficiente de inbreeding, $F$

➤ Single-point methods

➤ Multi-point methods

## PLINK v.1.9

$F_h$  estima el valor de  $F$  basándose en el exceso de homocigosis a nivel de SNPs.

$$F_{h_j} = \frac{O(H_j) - E(H)}{m - E(H)}$$

$O(H_j)$  es la homocigosis observada en todos los SNPs para la persona  $j$ ,  
 $E(H)$  es la homocigosis esperada:  
 $\sum_i 1 - 2p_i(1 - p_i)$

## F Suite

Método de máxima verosimilitud (LHM). Usa **Hidden Markov Model (HMM)** para modelizar el proceso de los fragmentos **IBD** en un individuo.

$$P(X_k = 1 | X_{k-1} = 1) = (1 - e^{-at_k})f + e^{-at_k},$$

$$P(X_k = 0 | X_{k-1} = 1) = (1 - e^{-at_k})(1 - f),$$

$$P(X_k = 1 | X_{k-1} = 0) = (1 - e^{-at_k})f, \text{ and}$$

$$P(X_k = 0 | X_{k-1} = 0) = (1 - e^{-at_k})(1 - f) + e^{-at_k},$$

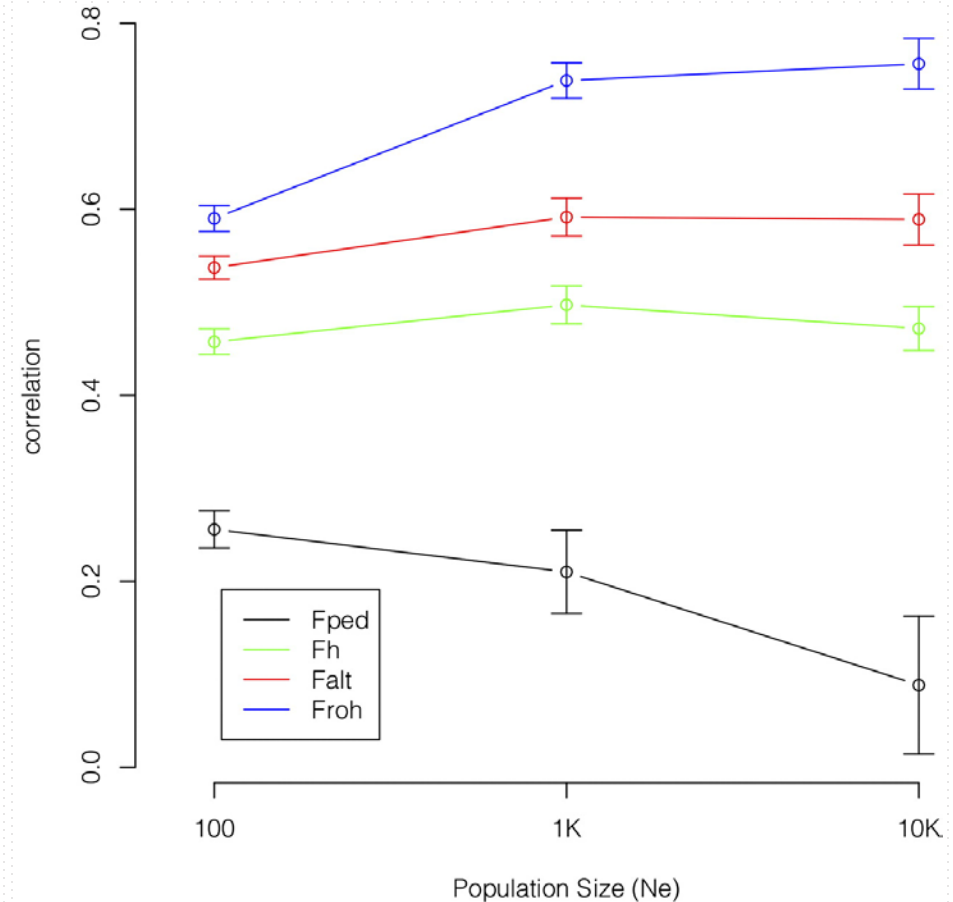
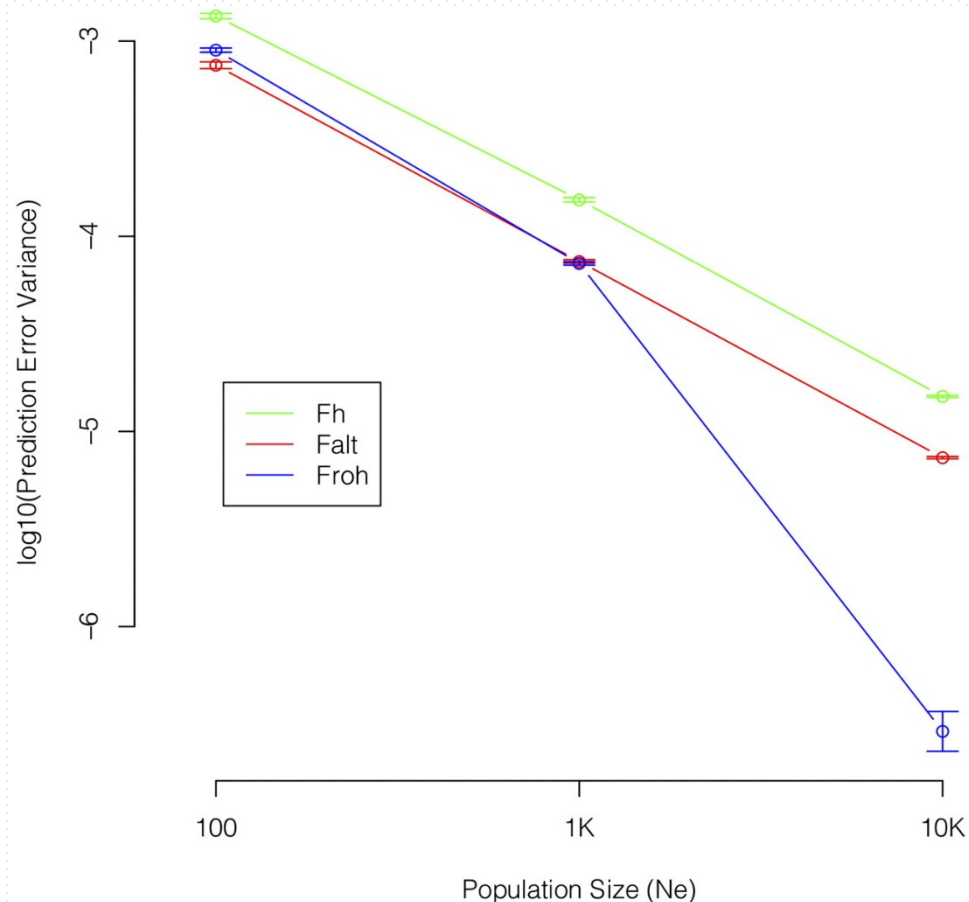
## Librería *detectRUNs* de R

$F_{ROH}$  se refiere a la proporción del genoma que está ocupado por *Runs of Homozygosity (ROHs)*.

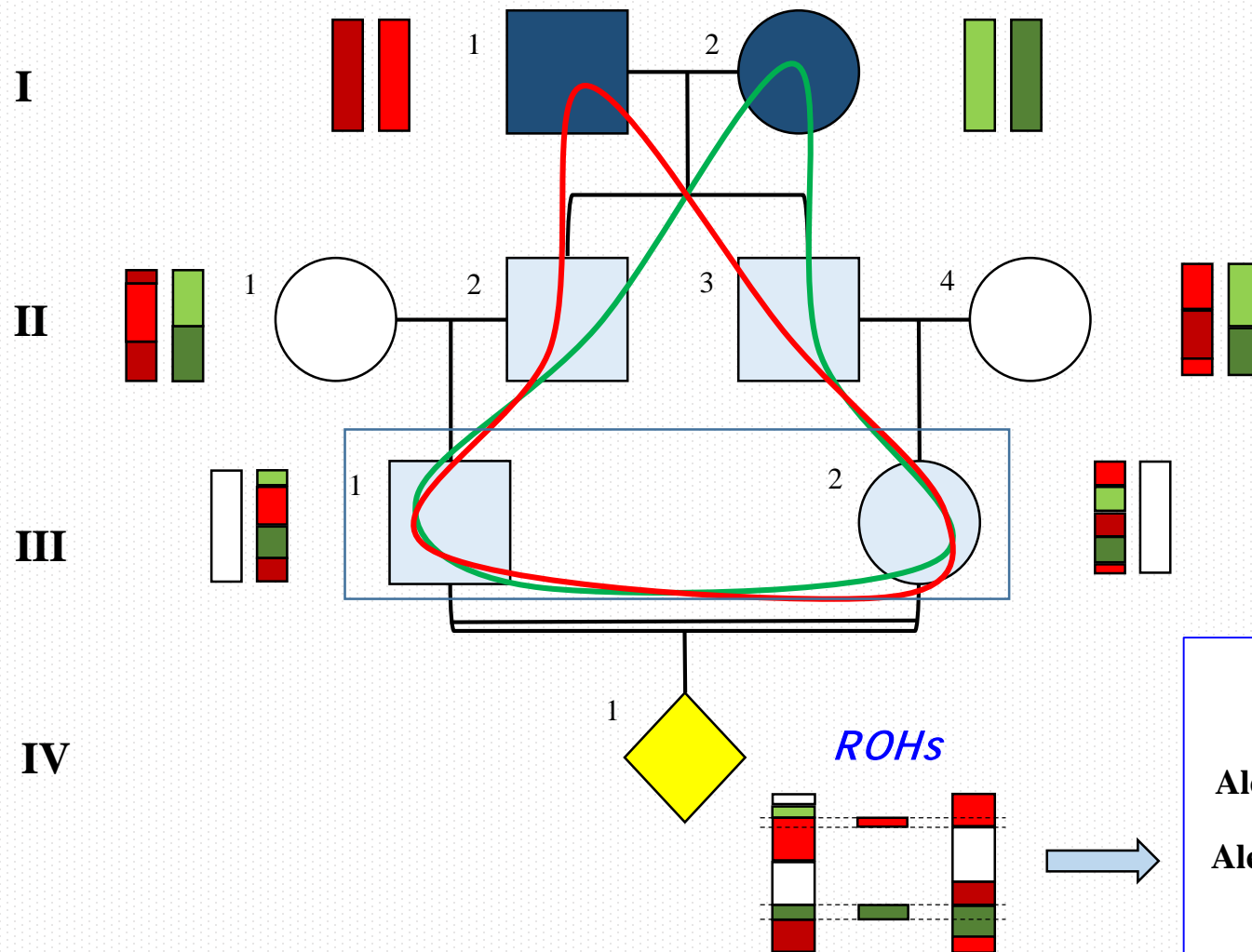
$$F_{ROH} = \frac{\sum Length(ROHs)}{L_{auto}}$$

- A pesar del desarrollo de diferentes metodologías conducentes a estimar **el valor de  $F$  a nivel genómico**, los estudios que emplean procesos de simulación genómica han llegado a la conclusión de que, el estimador más poderoso para detectar los efectos de la consanguinidad es el denominado  **$F_{ROH}$**

(Keller et al., 2011; Gazal et al., 2014)



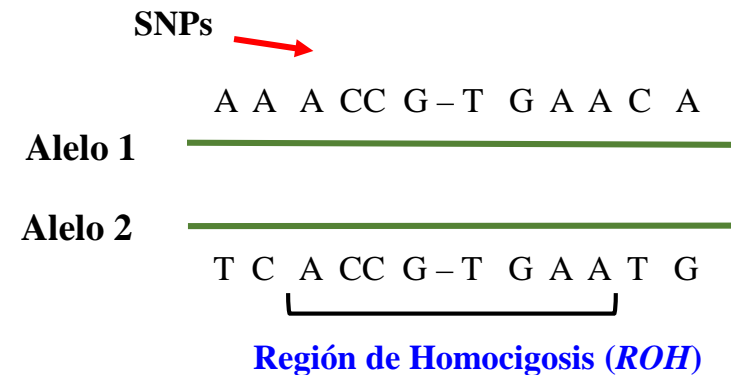
# Metodología genómica para la estima del coeficiente de inbreeding, $F_{ROH}$



**Fragmentos ROHs (Runs of Homozygosity):**

Segmentos cromosómicos idénticos por descendencia (IBD).

$$F_{roh} = \frac{\sum L_{roh}}{L_{auto}}$$



# LAS APLICACIONES BIOMEDICAS Y POBLACIONALES DEL INBREEDING GENOMICO

- La posibilidad de posicionar a nivel genómico la localización cromosómica de cada fragmento **ROH**, ello hace que uno de los enfoques más atractivos hoy es el de relacionar la distribución y estructura de los **ROHs** a lo largo del genoma humano con enfermedades de etiología compleja (*Esquizofrenia, Alzheimer, Parkinson, diversos tipos de cáncer etc...*).

OPEN ACCESS Freely available online

PLOS GENETICS

## Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor

Matthew C. Keller<sup>1,2\*</sup>, Matthew A. Simonson<sup>1,2</sup>, Stephan Ripke<sup>3,4,5</sup>, Ben M. Neale<sup>3,4,5</sup>, Pablo V. Gejman<sup>6,7</sup>, Daniel P. Howrigan<sup>1,2</sup>, Sang Hong Lee<sup>8</sup>, Todd Lencz<sup>9,10,11</sup>, Douglas F. Levinson<sup>12</sup>, Patrick F. Sullivan<sup>13,14,15,16</sup>, The Schizophrenia Psychiatric Genome-Wide Association Study Consortium<sup>1</sup>



Available online at www.jbr-pub.org

Open Access at PubMed Central

The Journal of Biomedical Research, 2013, 27(3):208–214

Research Paper

JBR

## Genome-wide analysis of runs of homozygosity identifies new susceptibility regions of lung cancer in Han Chinese

Cheng Wang<sup>a,Δ</sup>, Zhengfeng Xu<sup>b,Δ</sup>, Guangfu Jin<sup>a,c,d</sup>, Zhibin Hu<sup>a,c,d</sup>, Juncheng Dai<sup>a</sup>, Hongxia Ma<sup>a</sup>, Yue Jiang<sup>a</sup>, Lingmin Hu<sup>a</sup>, Minjie Chu<sup>a</sup>, Songyu Cao<sup>a</sup>, Hongbing Shen<sup>a,c,d,✉</sup>

<sup>a</sup>Department of Epidemiology and Biostatistics and Ministry of Education (MOE) Key Lab for Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 210029, China;

OPEN ACCESS Freely available online

PLOS one

## Cooperative Genome-Wide Analysis Shows Increased Homozygosity in Early Onset Parkinson's Disease

Javier Simón-Sánchez<sup>1,3</sup>, Laura L. Kilarski<sup>2,6,9</sup>, Michael A. Nalls<sup>3</sup>, Maria Martinez<sup>4,5</sup>, Claudia Schulte<sup>7</sup>, Peter Holmans<sup>2,6</sup>, International Parkinson's Disease Genomics Consortium<sup>1,8</sup>, Wellcome Trust Case Control Consortium<sup>2,10</sup>, Thomas Gasser<sup>7</sup>, John Hardy<sup>8</sup>, Andrew B. Singleton<sup>3</sup>, Nicholas W. Wood<sup>8</sup>, Alexis Brice<sup>9,10,11</sup>, Peter Heutink<sup>1</sup>, Nigel Williams<sup>2,6\*</sup>, Huw R. Morris<sup>2,6\*</sup>

1 Section of Medical Genetics, Department of Clinical Genetics, VU University Medical Centre, Amsterdam, The Netherlands, 2 MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, United Kingdom, 3 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, 4 Inserm, UMR 1043, Toulouse, France, 5 Paul Sabatier University, Toulouse, France, 6 Department of Psychological Medicine & Neurology, Cardiff University School of Medicine, Cardiff, United Kingdom, 7 Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany, 8 Department of Molecular Neuroscience, UCL Institute of Neurology, London, United Kingdom, 9 Université Pierre et Marie Curie-Paris6, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière, UMR-S975, Paris, France, 10 Inserm, U975, Paris, France, 11 Cnrs, UMR 7225, Paris, France

SCIENTIFIC REPORTS

## OPEN Inbreeding and homozygosity in breast cancer survival

Hauke Thomsen<sup>1</sup>, Miguel Inacio da Silva Filho<sup>1</sup>, Andrea Woltmann<sup>1</sup>, Robert Johansson<sup>2</sup>, Jorunn E. Eyfjörð<sup>3</sup>, Ute Hamann<sup>4</sup>, Jonas Manjer<sup>5,6</sup>, Kerstin Enquist-Olsson<sup>7</sup>, Roger Henriksson<sup>2,8</sup>, Stefan Herms<sup>9,10</sup>, Per Hoffmann<sup>9,10</sup>, Bowang Chen<sup>1</sup>, Stefanie Huhn<sup>1</sup>, Kari Hemminki<sup>1,11</sup>, Per Lenner<sup>2</sup> & Asta Försti<sup>1,11</sup>

Received: 15 June 2015

Accepted: 14 October 2015

Published: 12 November 2015

# LAS APLICACIONES BIOMEDICAS Y POBLACIONALES DEL INBREEDING GENOMICO

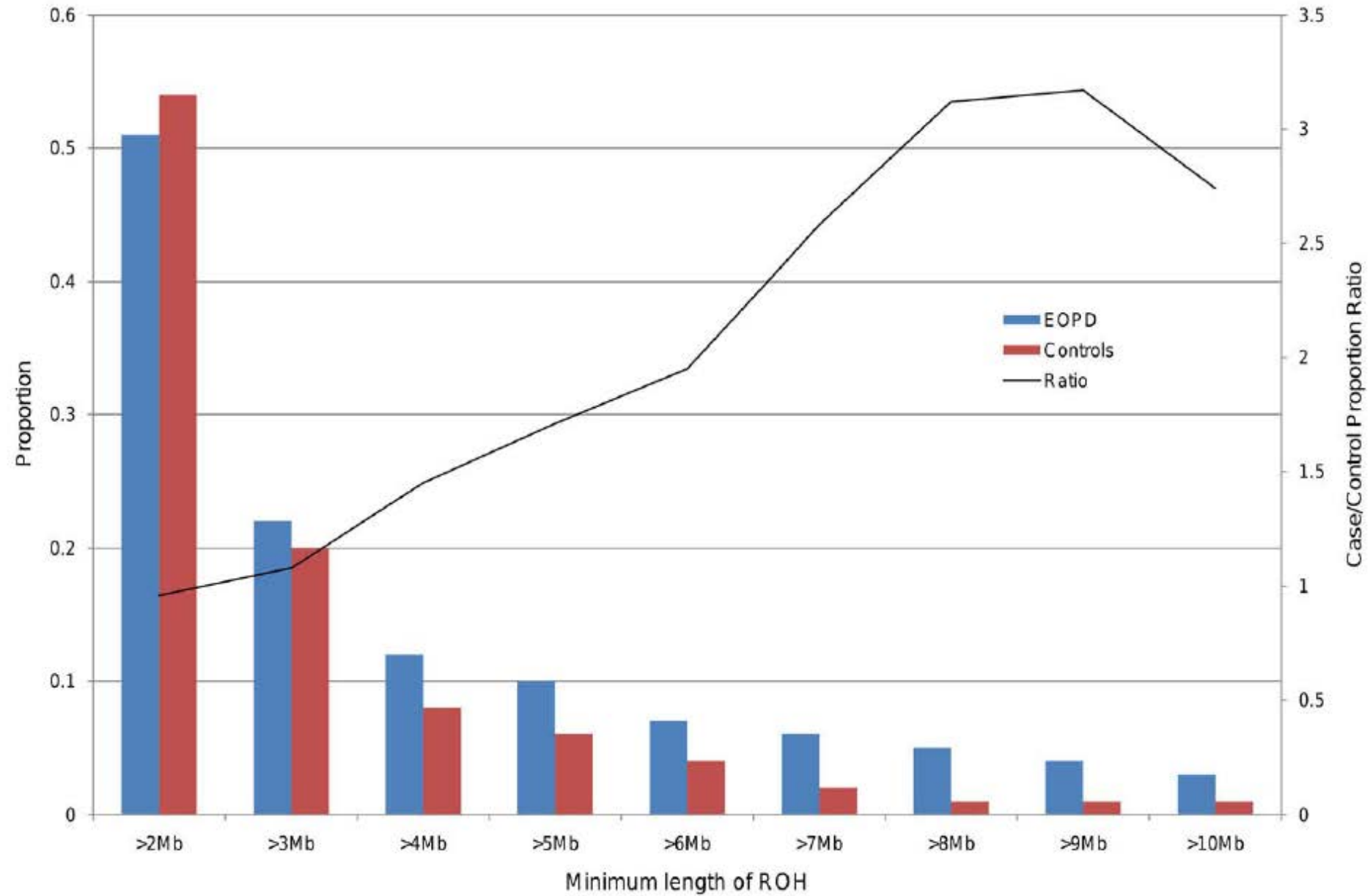
- Vinculación entre el aumento de los niveles de  $F_{ROH}$  con el incremento de la incidencia de algunas enfermedades humanas.

Modelo  
regresión logística

Por cada incremento de 0.01 en el valor de  $F_{roh}$  *las probabilidades de presentar esquizofrenia aumentaban un 17%.*

*Keller et al., (2012)*

- La población de enfermos frente a la población control, destaca por albergar en su genoma fragmentos *ROHs* de tamaño (longitud) más grande (*Simón-Sánchez et al., 2012; Ghani et al., 2015*)

*continua...*

**EOPD: Early Onset  
Parkinson Disease**



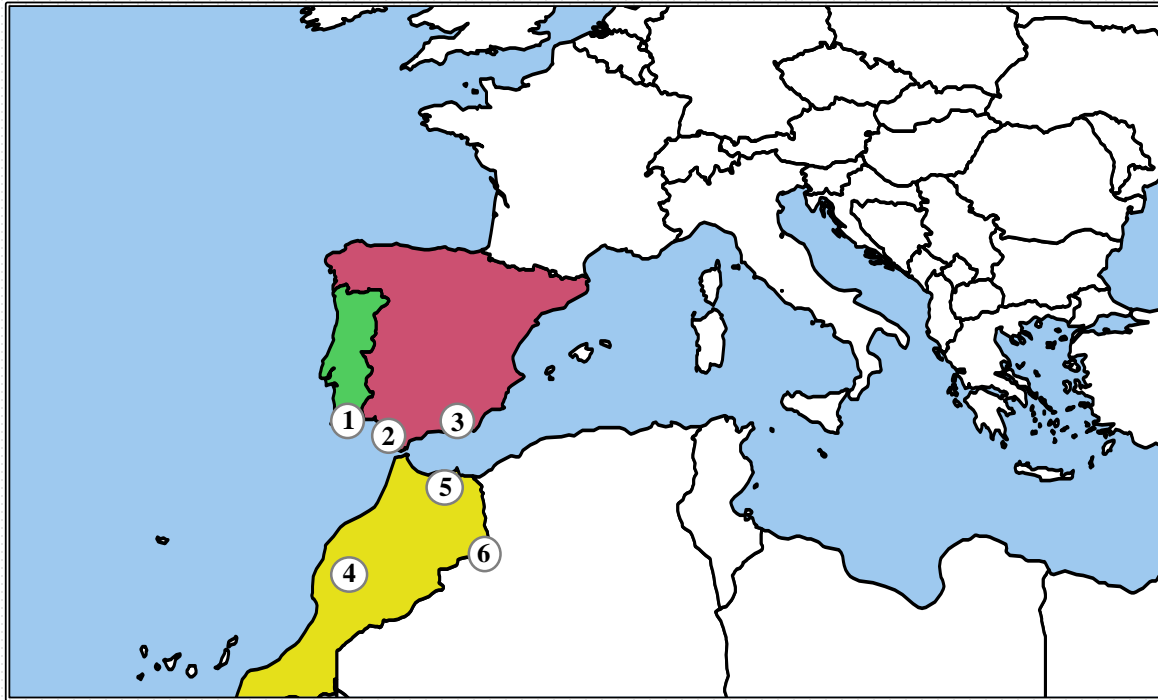
# Los objetivos del presente estudio.....

- Inferir la valoración de los Coeficientes de *Inbreeding* genómicos mediante distintas metodologías basadas en aproximaciones Single-point y Multi-point en poblaciones asentadas en el extremo occidental del Mediterráneo.

# Los objetivos del presente estudio.....

- Inferir la valoración de los Coeficientes de *Inbreeding* genómicos mediante distintas metodologías basadas en aproximaciones Single-point y Multi-point en poblaciones asentadas en el extremo occidental del Mediterráneo.
- Valorar el grado de acuerdo entre los valores hallados para las distintas medidas de *Inbreeding* y contrastar la rentabilidad de las diferentes metodologías bioinformáticas disponibles hoy día en un caso real de análisis poblacional.

# LAS POBLACIONES ANALIZADAS



1: Sur de Portugal, 2: Oeste de Andalucía (Huelva), 3: Este de Andalucía (Granada), 4: Asni (Bereberes), 5: Bouhria (Bereberes), 6: Figuig (Bereberes).

Sur de Iberia

## Andalucía

Oeste de Andalucía

n = 35

Este de Andalucía

n = 35

## Sur de Portugal

Sur de Portugal

n = 35

## Bereberes de Marruecos

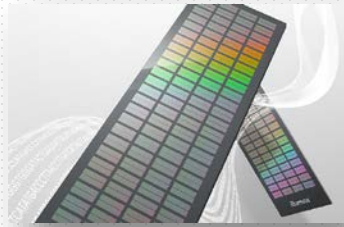
Asni n = 15

Bouhria n = 12

Figuig n = 9

Marruecos

# Proceso de genotipado y “*workflow*” para el tratamiento y filtrado de SNPs



Illumina Omni 2.5 array (~2.5 millones de SNPs)

142 muestras  
Human Omni 2.5  
2.372.784 SNPs



142 muestras  
2.312.580 SNPs



142 muestras  
2.270.963 SNPs



139 muestras  
2.270.887 SNPs

## Datos genotipados:

- 74 varones y 68 mujeres
- Tasa total de genotipado: 0.9953

Retirar chr X, Y, XY,  
mtDNA, MIND 0.05

0 individuos retirados  
60.204 SNPs retirados

GENO 0.05

41.617 SNPs retirados

- Estimación de parentesco → 3 individuos retirados

HWE 0.0000001

76 SNPs retirados

**PLINK v.1.9**

# **Las herramientas bioinformáticas. Estimación de los fragmentos *ROHs* y de los coeficientes de *Inbreeding* genómico, *F*.**

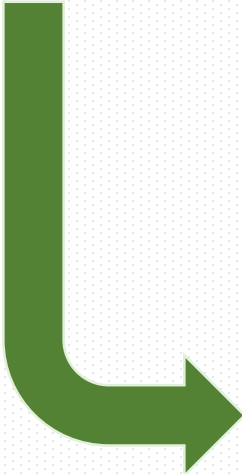
***PLINK v.1.9***

*Homoz 0.5 40 100*

Las herramientas bioinformáticas. Estimación de los fragmentos ROHs y de los coeficientes de Inbreeding genómico, *F*.

PLINK v.1.9

Homoz 0.5 40 100



Fragmentos ROHs

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	FID	Population	ID	PHE	CHR	SNP1	SNP2	POS1	POS2	KB	NSNP	DENSITY	PHOM	PHET
2	35	Huelva	299532	1	1	rs12085929	rs79459747	31372734	31898547	525814	48	10.954	0.979	0.021
3	35	Huelva	299532	1	1	rs79121844	rs74084501	53917385	54441853	524469	90	5.827	0.978	0.022
4	35	Huelva	299532	1	1	rs11207950	rs17316198	62733939	63247772	513834	80	6.423	0.975	0.025
5	35	Huelva	299532	1	1	rs72682829	rs76999445	68071462	68584231	512770	100	5.128	0.980	0.020
6	35	Huelva	299532	1	1	rs78731300	rs77662524	86308365	86884365	576001	80	7.200	0.988	0.013
7	35	Huelva	299532	1	1	rs80354837	rs116682102	90017105	90558859	541755	104	5.209	0.981	0.019
8	35	Huelva	299532	1	1	rs114554902	rs28708456	98085141	98706205	621065	82	7.574	0.988	0.012
9	35	Huelva	299532	1	1	rs742682	rs9429548	114897076	115668368	771293	76	10.149	0.974	0.026
10	35	Huelva	299532	1	1	rs115190046	rs7526319	151007824	151524558	516735	64	8.074	0.969	0.031
11	35	Huelva	299532	1	1	rs75720912	rs17656483	158390676	158907239	516564	113	4.571	0.991	0.000
12	35	Huelva	299532	1	1	rs112620748	rs12029220	186804285	187338342	534058	92	5.805	0.978	0.022
13	35	Huelva	299532	1	1	rs60996022	rs521719	188755377	189530964	775588	68	11.406	0.971	0.029
14	35	Huelva	299532	1	1	rs10920983	rs72732973	191622901	192366043	743143	63	11.796	0.984	0.016
15	35	Huelva	299532	1	1	rs73062282	rs76764548	192768787	193440379	671593	52	12.915	0.981	0.019
16	35	Huelva	299532	1	1	rs10494698	rs2839897	194096582	194841920	745339	79	9.435	0.975	0.025
17	35	Huelva	299532	1	1	rs10449246	rs116303679	224936295	225584974	648680	65	9.980	0.969	0.031
18	35	Huelva	299532	1	2	rs79270739	rs76665514	39211234	39875554	664321	74	8.977	0.973	0.027
19	35	Huelva	299532	1	2	rs76081756	rs74635159	63166198	64399313	1233116	76	16.225	0.974	0.026
20	35	Huelva	299532	1	2	rs74262579	rs35073965	65934026	66550979	616954	113	5.460	0.973	0.027
21	35	Huelva	299532	1	2	rs12713914	rs115834754	77705469	78247944	542476	96	5.651	0.979	0.021
22	35	Huelva	299532	1	2	rs114761069	rs76504362	83898828	84405076	506249	48	10.547	0.979	0.021
23	35	Huelva	299532	1	2	rs62158021	rs7589893	99320597	99932255	611659	48	12.743	0.979	0.021
24	35	Huelva	299532	1	2	rs34617783	rs13416708	102262110	102899304	637195	77	8.275	0.974	0.026
25	35	Huelva	299532	1	2	rs78467089	rs11123454	117949973	118616023	666051	88	7.569	0.989	0.011

# Las herramientas bioinformáticas. Estimación de los fragmentos *ROHs* y de los coeficientes de *Inbreeding* genómico, *F*.

*PLINK v.1.9*

*Homoz 0.5 40 100*



**Fragmentos *ROHs* +  $F_{PLINK}$**

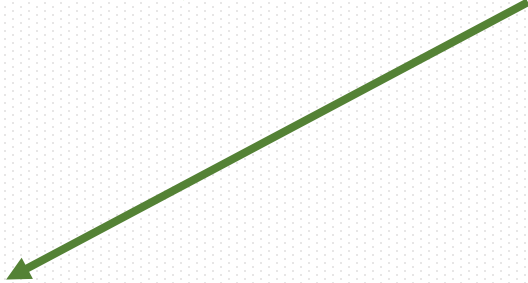
# Las herramientas bioinformáticas. Estimación de los fragmentos *ROHs* y de los coeficientes de *Inbreeding* genómico, *F*.

*PLINK v.1.9*

*Homoz 0.5 40 100*



**Fragmentos *ROHs* +  $F_{PLINK}$**



*Librería detectRUNs de R*



**Estimación del valor de  $F_{ROH}$  basado en los datos de *ROHs***



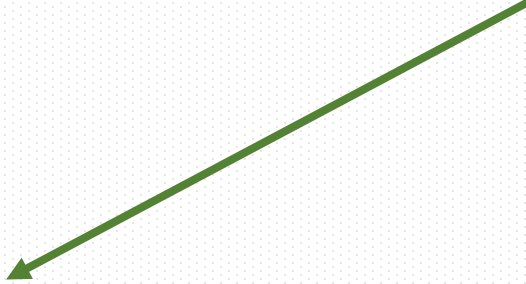
# Las herramientas bioinformáticas. Estimación de los fragmentos *ROHs* y de los coeficientes de *Inbreeding* genómico, *F*.

*PLINK v.1.9*

*Homoz 0.5 40 100*



**Fragmentos *ROHs* +  $F_{PLINK}$**



*Librería detectRUNs de R*



**Estimación del valor de  $F_{ROH}$  basado en los datos de *ROHs***

- Nivel genómico

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{auto}}$$

# Las herramientas bioinformáticas. Estimación de los fragmentos ROHs y de los coeficientes de Inbreeding genómico, $F$ .

**PLINK v.1.9**

*Homoz 0.5 40 100*



**Fragmentos ROHs +  $F_{PLINK}$**

**Librería detectRUNs de R**



**Estimación del valor de  $F_{ROH}$  basado en los datos de ROHs**

■ Nivel genómico 
$$F_{ROH} = \frac{\sum L_{ROH}}{L_{auto}}$$

**F Suite**



**$F_{mean}$  de F Suite**

# Las herramientas bioestadísticas

- **Análisis Descriptivo  $F$**

- Media y su desviación estándar
- Density plots (*violin plots*)
- Scatter plots

- **Bland-Altman plots y Coeficiente de correlación intraclass**

*Librería: psych*

$$ICC(3,1) = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$$

- **Modelización del coeficiente de *inbreeding* genómico,  $F$**

*Librerías: fitdistrplus, actuar*

- **Modelo de regresión logística empleando el valor  $F$ .  
Aplicación en un caso real.**



**LOS RESULTADOS DEL PRESENTE ESTUDIO.....**

# El Análisis Descriptivo *F*

Coeficiente de inbreeding, <i>F</i>	<i>F<sub>ROH</sub></i>			<i>F<sub>Suite</sub></i>			<i>F<sub>PLINK</sub></i>		
Poblaciones	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>
O. Andal. (Huelva)	0.053	0.008	(0.037; 0.068)	0.017	0.016	(-0.015; 0.048)	0.133	0.007	(0.118; 0.147)
E. Andal. (Granada)	0.049	0.005	(0.039; 0.059)	0.006	0.007	(-0.008; 0.020)	0.130	0.005	(0.121; 0.140)
Sur Portugal	0.047	0.005	(0.037; 0.057)	0.008	0.010	(-0.012; 0.028)	0.123	0.006	(0.111; 0.134)
Marruecos (Bereberes)	0.064	0.027	(0.011; 0.118)	0.034	0.032	(-0.029; 0.097)	0.105	0.031	(0.043; 0.167)

<sup>1</sup> Desviación estándar; <sup>2</sup> Intervalo de confianza al 95%

# El Análisis Descriptivo *F*

Coeficiente de inbreeding, <i>F</i>	<i>F<sub>ROH</sub></i>			<i>F<sub>Suite</sub></i>			<i>F<sub>PLINK</sub></i>		
Poblaciones	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>
O. Andal. (Huelva)	0.053	0.008	(0.037; 0.068)	0.017	0.016	(-0.015; 0.048)	0.133	0.007	(0.118; 0.147)
E. Andal. (Granada)	0.049	0.005	(0.039; 0.059)	0.006	0.007	(-0.008; 0.020)	0.130	0.005	(0.121; 0.140)
Sur Portugal	0.047	0.005	(0.037; 0.057)	0.008	0.010	(-0.012; 0.028)	0.123	0.006	(0.111; 0.134)
Marruecos (Bereberes)	0.064	0.027	(0.011; 0.118)	0.034	0.032	(-0.029; 0.097)	0.105	0.031	(0.043; 0.167)

<sup>1</sup> Desviación estándar; <sup>2</sup> Intervalo de confianza al 95%

El Análisis Descriptivo *F*

Coeficiente de inbreeding, <i>F</i>	<i>F<sub>ROH</sub></i>			<i>F<sub>Suite</sub></i>			<i>F<sub>PLINK</sub></i>		
Poblaciones	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>	media	s.d. <sup>1</sup>	IC <sub>95%</sub> <sup>2</sup>
O. Andal. (Huelva)	0.053	0.008	(0.037; 0.068)	0.017	0.016	(-0.015; 0.048)	0.133	0.007	(0.118; 0.147)
E. Andal. (Granada)	0.049	0.005	(0.039; 0.059)	0.006	0.007	(-0.008; 0.020)	0.130	0.005	(0.121; 0.140)
Sur Portugal	0.047	0.005	(0.037; 0.057)	0.008	0.010	(-0.012; 0.028)	0.123	0.006	(0.111; 0.134)
Marruecos (Bereberes)	0.064	0.027	(0.011; 0.118)	0.034	0.032	(-0.029; 0.097)	0.105	0.031	(0.043; 0.167)

<sup>1</sup> Desviación estándar; <sup>2</sup> Intervalo de confianza al 95%

# El Análisis Descriptivo de $F$

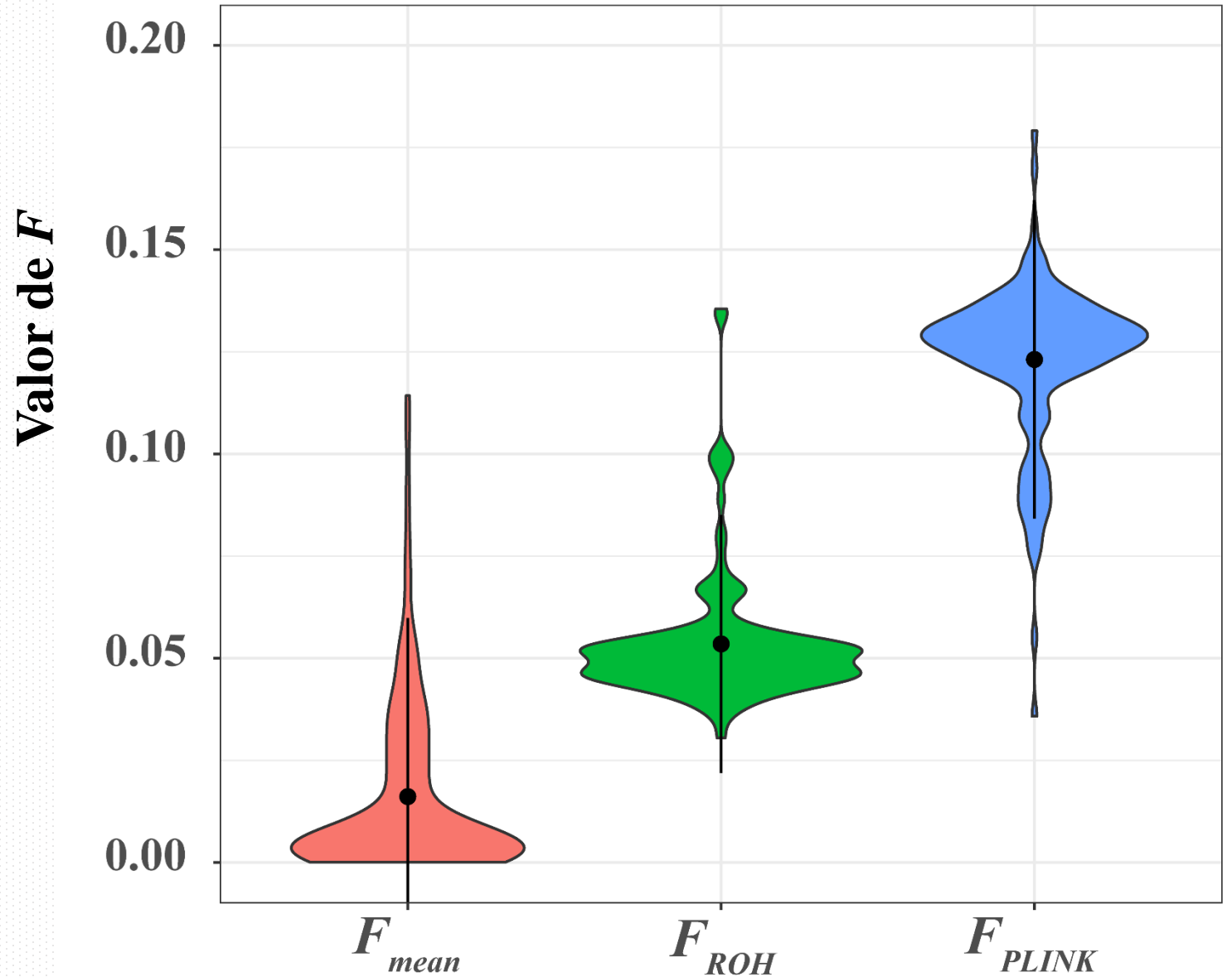
## (Violin Plots)

```
library(ggplot2)
```

```
datos_new=data.frame(F=c(rep("FROH",139),  
rep("Fmean",139),rep("FPLINK",139)),valor=c  
(datos$Froh,datos$Fmean,datos$FPLINK))
```

```
p <- ggplot(datos_new, aes(x=F,  
y=valor,fill=F)) + geom_violin()
```

```
p + theme_bw() + stat_summary  
(fun.data=mean_sdl, mult=1,  
geom="pointrange", color="black") +  
coord_cartesian(ylim = c(0,0.20))
```

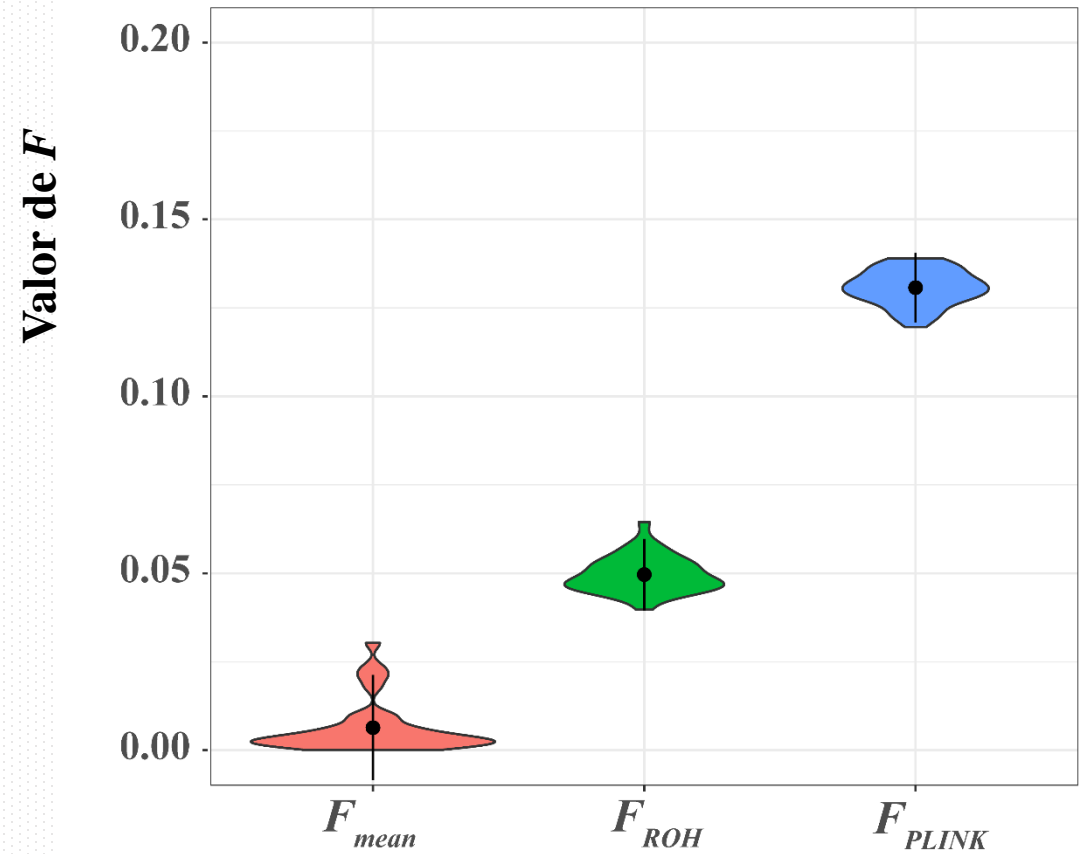




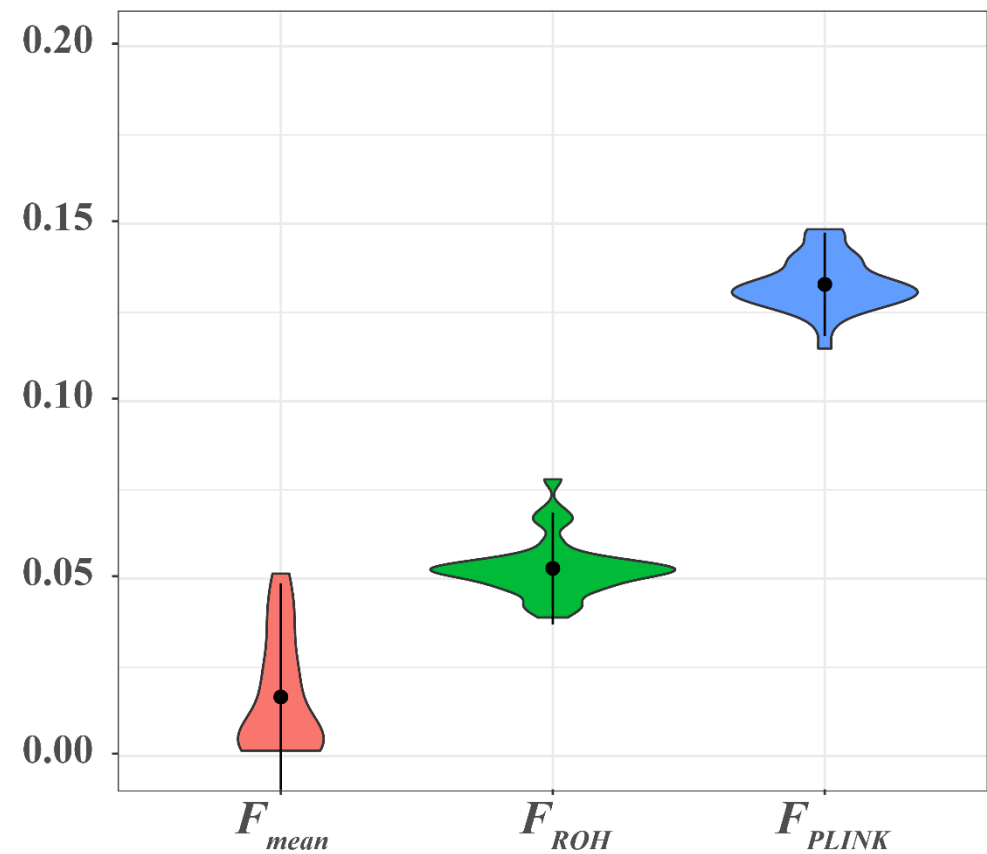
# El Análisis Descriptivo de $F$

(Violin Plots)

E.Andal (Granada)



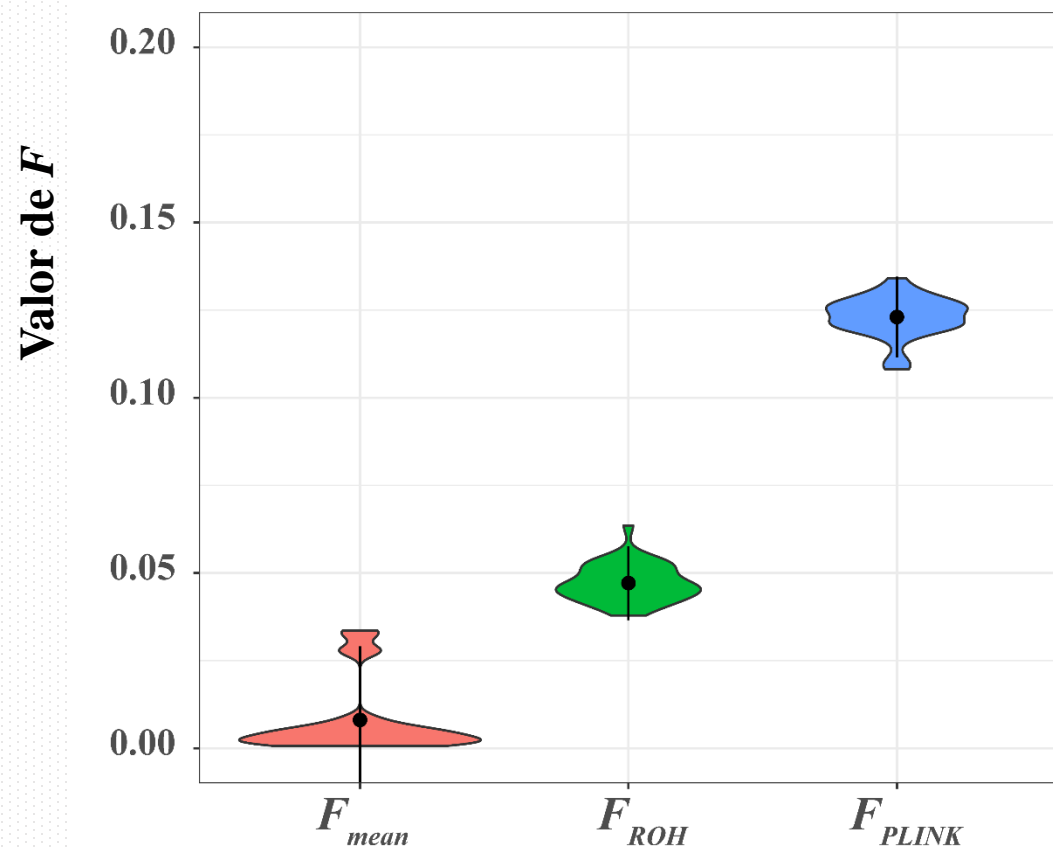
O.Andal (Huelva)



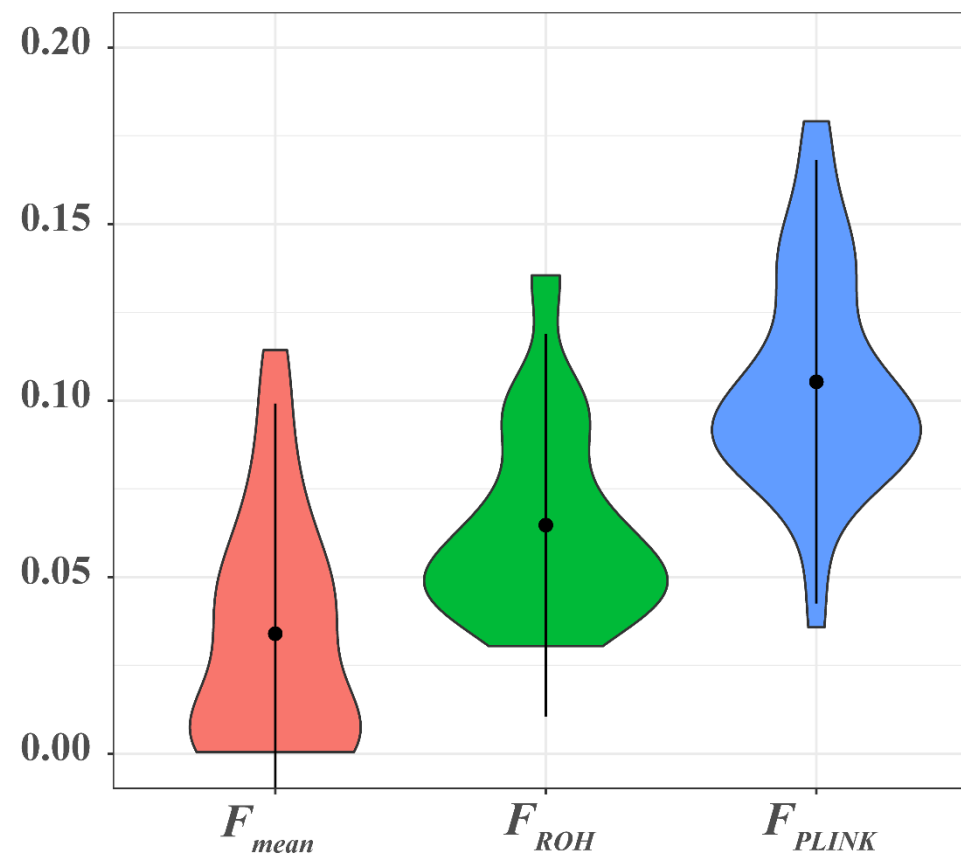
# El Análisis Descriptivo de $F$

(*Violin Plots*)

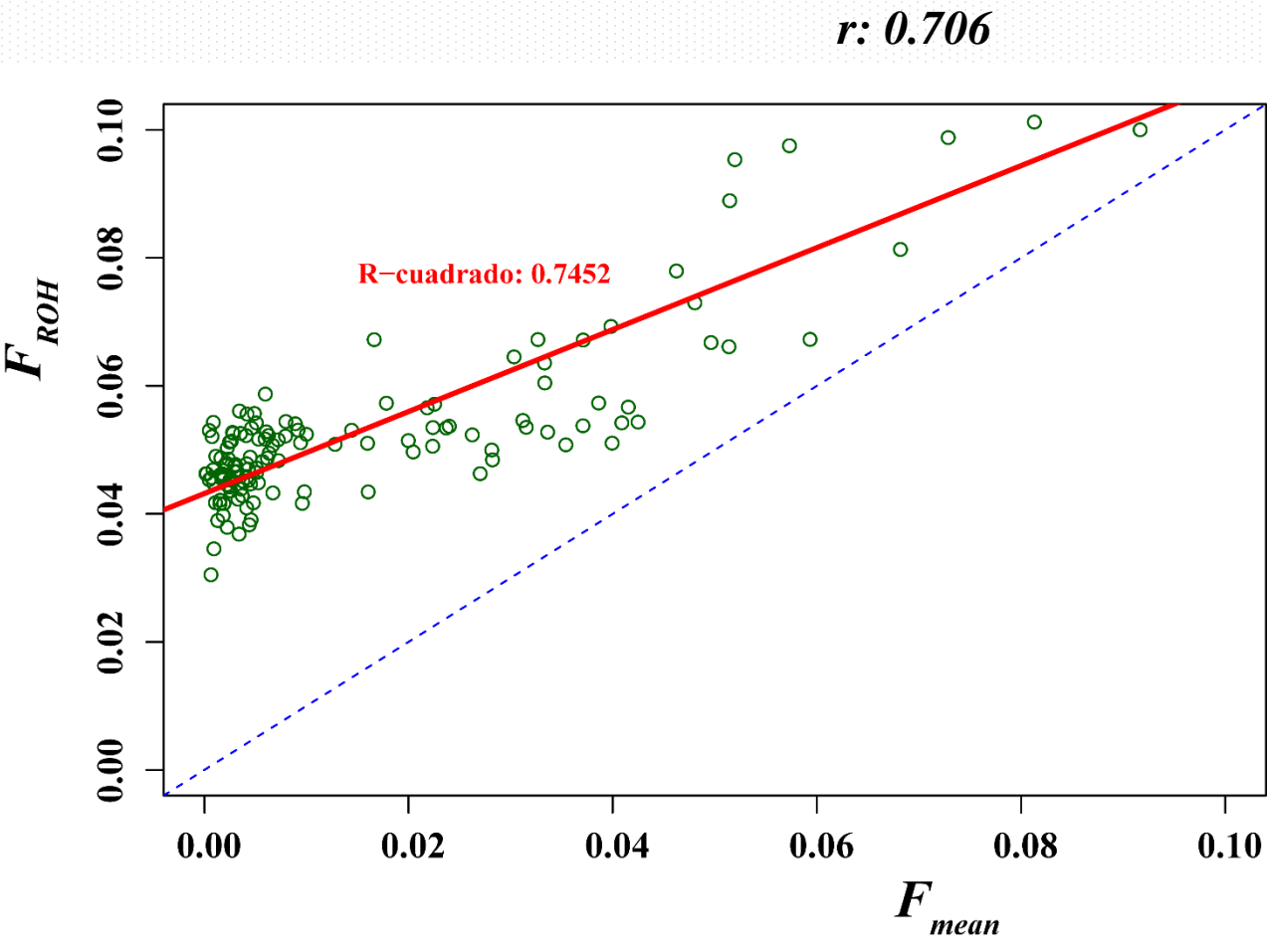
Sur de Portugal



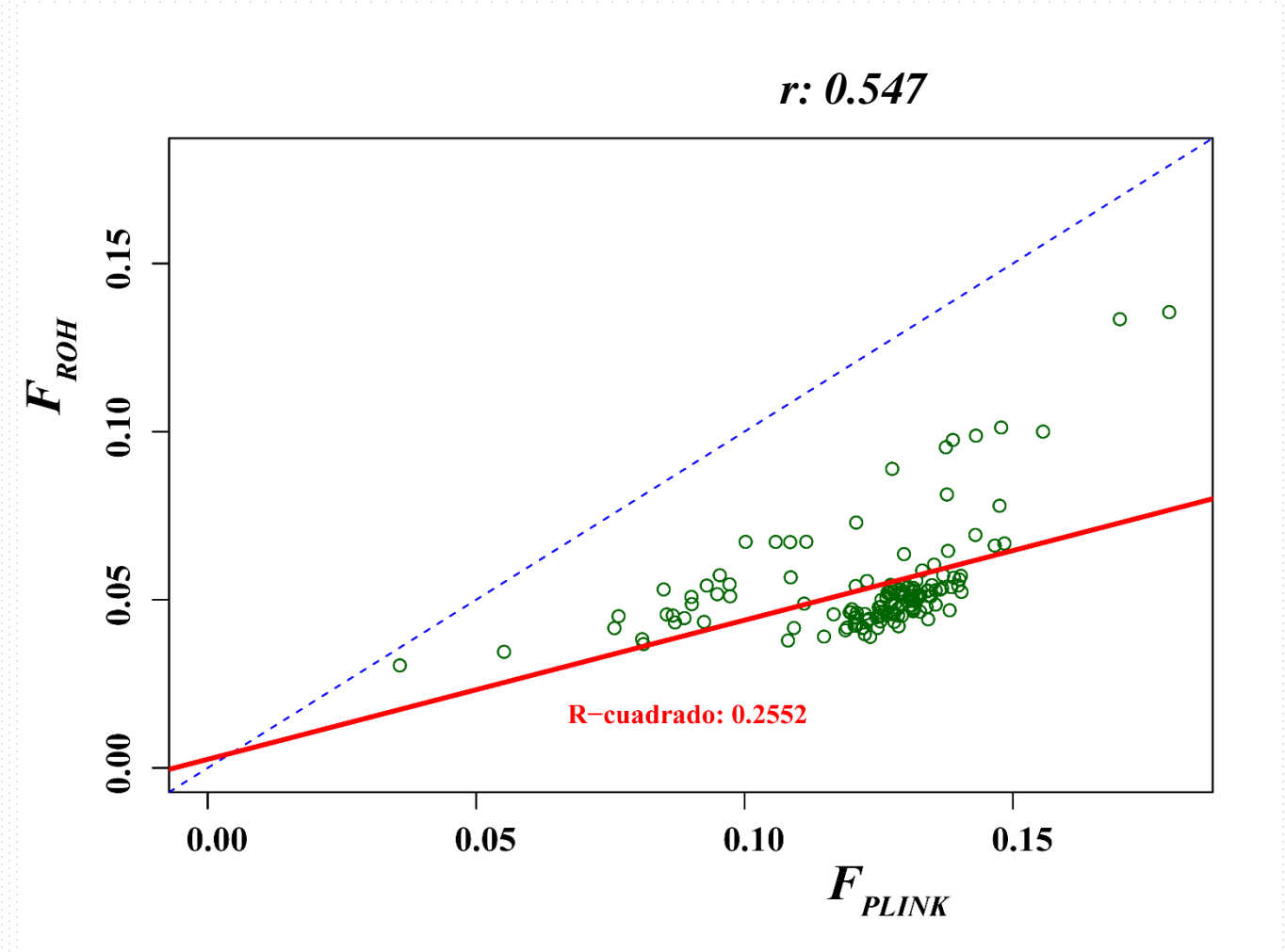
Marruecos (Bereberes)



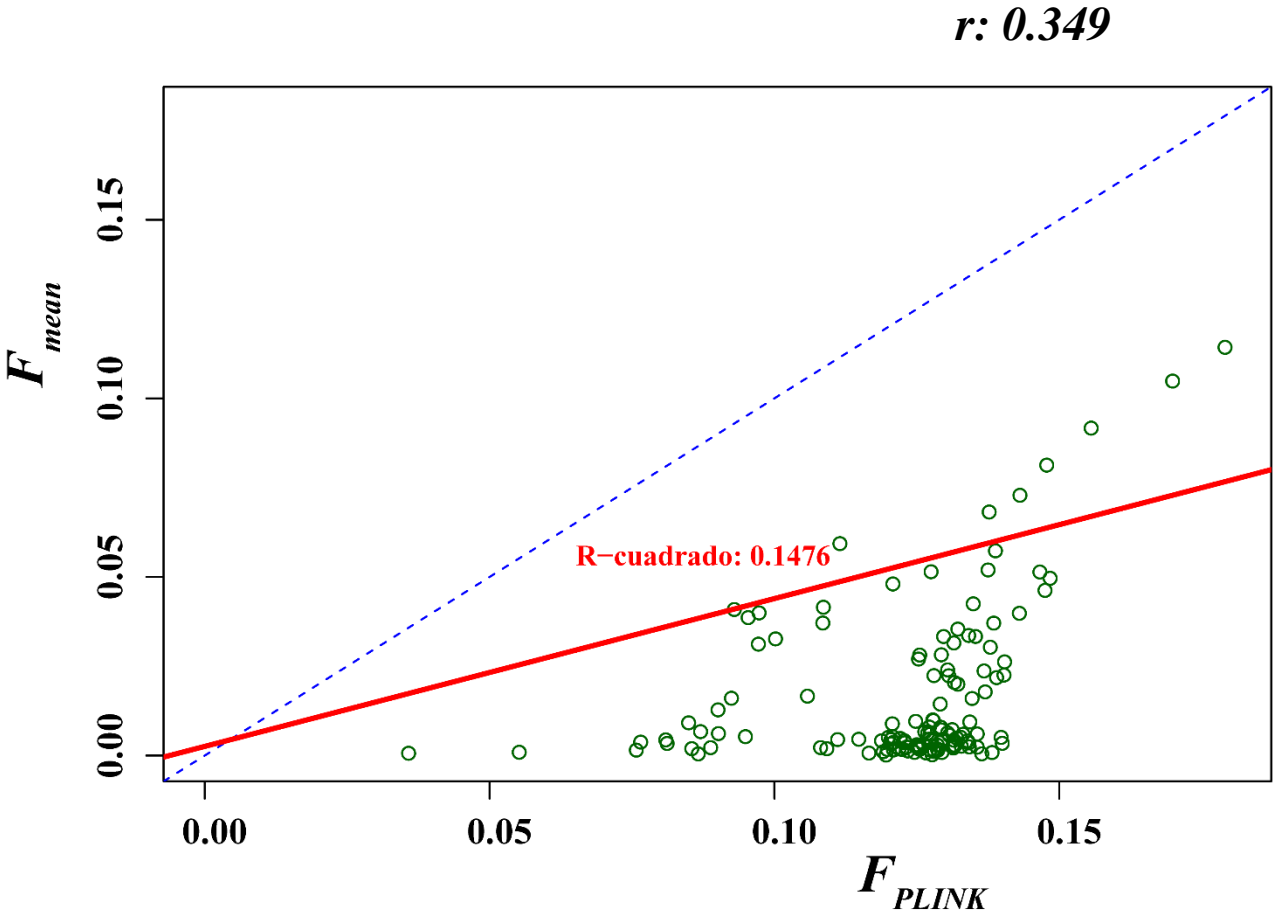
Scatter plots



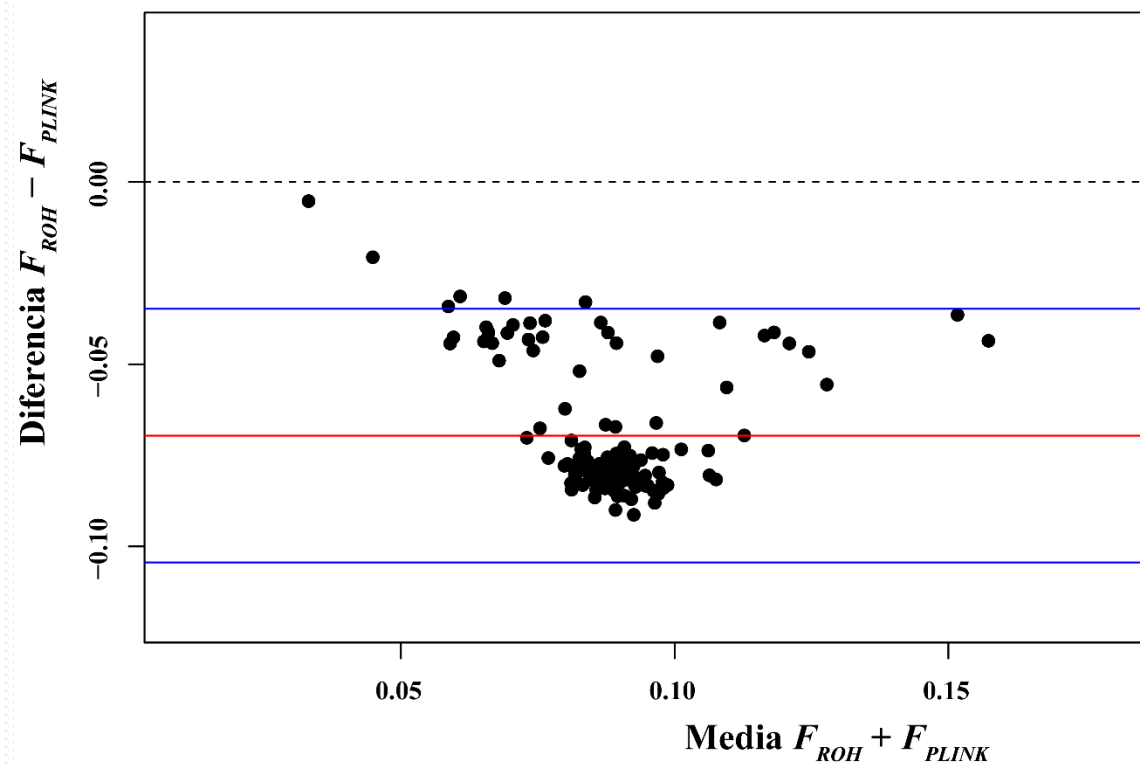
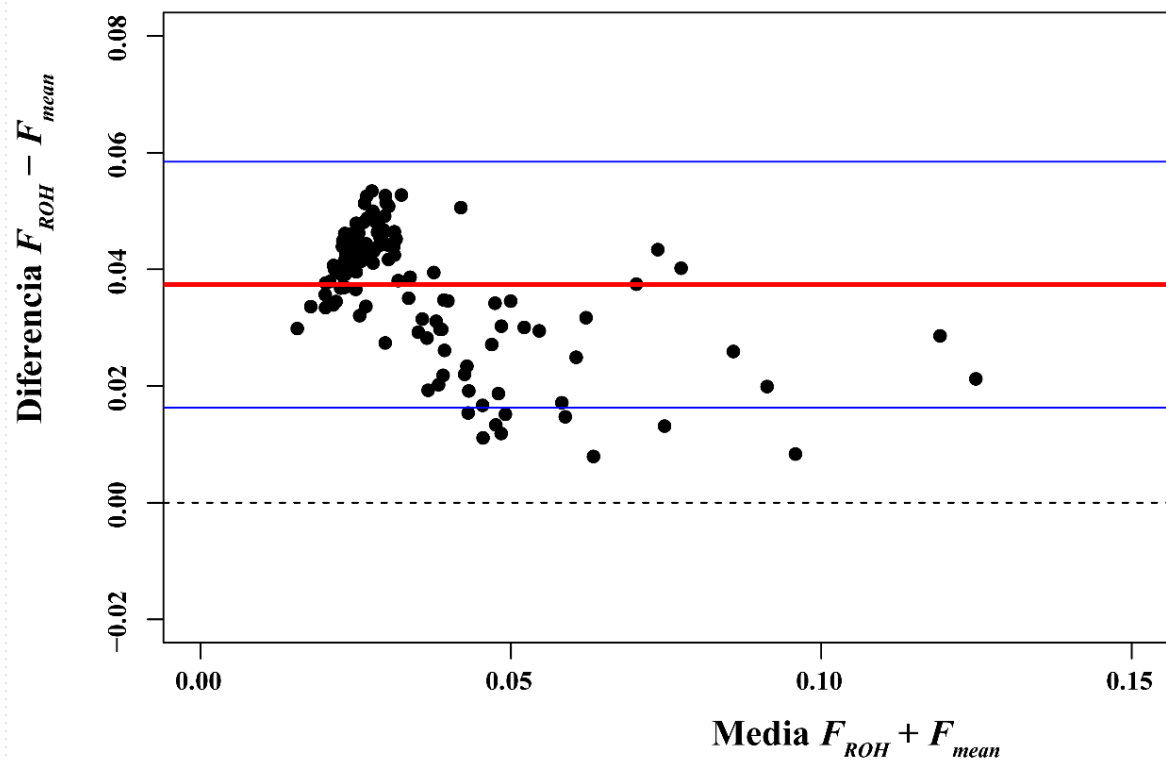
Scatter plots



# Scatter plots



# Bland-Altman plots



# Coeficiente de correlación intraclase

```
library(psych)
```

```
A<-matrix(c(datos$Froh,datos$Fmean),ncol=2,byrow=F)
```

```
FROHvsFmean<-ICC(A,missing=TRUE,alpha=0.05,lmer=FALSE,check.keys=FALSE)
```

```
FROHvsFmean$results
```

```
> FROHvsFmean$results
```

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.06087064	0.885244	138	139	7.628415e-01	-0.22424861	0.1059365
single_random_raters	ICC2	0.28794025	11.573190	138	138	6.325186e-39	-0.03666546	0.6571287
single_fixed_raters	ICC3	0.84093138	11.573190	138	138	6.325186e-39	0.78444034	0.8835838
Average_raters_absolute	ICC1k	-0.12963206	0.885244	138	139	7.628415e-01	-0.57814555	0.1915779
Average_random_raters	ICC2k	0.44713294	11.573190	138	138	6.325186e-39	-0.07612196	0.7930931
Average_fixed_raters	ICC3k	0.91359340	11.573190	138	138	6.325186e-39	0.87920041	0.9381943

```
> FROHvsFPLINK$results
```

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.69085951	0.1828304	138	139	1.0000e+00	-0.76855072	-0.5930235
single_random_raters	ICC2	0.05740032	2.9953105	138	138	1.6807e-10	-0.02667689	0.2205237
single_fixed_raters	ICC3	0.49941312	2.9953105	138	138	1.6807e-10	0.36356650	0.6144618
Average_raters_absolute	ICC1k	-4.46955044	0.1828304	138	139	1.0000e+00	-6.64120197	-2.9142881
Average_random_raters	ICC2k	0.10856876	2.9953105	138	138	1.6807e-10	-0.05481611	0.3613591
Average_fixed_raters	ICC3k	0.66614479	2.9953105	138	138	1.6807e-10	0.53325818	0.7611971

```
> FmeanvsFPLINK$results
```

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.82001004	0.09889504	138	139	1.000000e+00	-0.86778198	-0.7571775
single_random_raters	ICC2	0.02659129	2.23442966	138	138	1.638532e-06	-0.01750914	0.1123122
single_fixed_raters	ICC3	0.38165296	2.23442966	138	138	1.638532e-06	0.23025485	0.5150121
Average_raters_absolute	ICC1k	-9.11173070	0.09889504	138	139	1.000000e+00	-13.12653150	-6.2364681
Average_random_raters	ICC2k	0.05180502	2.23442966	138	138	1.638532e-06	-0.03564235	0.2019437
Average_fixed_raters	ICC3k	0.55245850	2.23442966	138	138	1.638532e-06	0.37432058	0.6798786

# Coeficiente de correlación intraclass

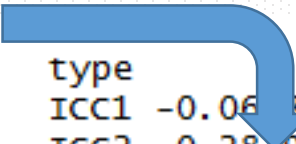
*library(psych)*

*A<-matrix(c(datos\$Froh,datos\$Fmean),ncol=2,byrow=F)*

*FROHvsFmean<-ICC(A,missing=TRUE,alpha=0.05,lmer=FALSE,check.keys=FALSE)*

*FROHvsFmean\$results*

*> FROHvsFmean\$results*



	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.0687064	0.885244	138	139	7.628415e-01	-0.22424861	0.1059365
single_random_raters	ICC2	0.28794025	11.573190	138	138	6.325186e-39	-0.03666546	0.6571287
single_fixed_raters	ICC3	0.84093138	11.573190	138	138	6.325186e-39	0.78444034	0.8835838
Average_raters_absolute	ICC1k	-0.12963206	0.885244	138	139	7.628415e-01	-0.57814555	0.1915779
Average_random_raters	ICC2k	0.44713294	11.573190	138	138	6.325186e-39	-0.07612196	0.7930931
Average_fixed_raters	ICC3k	0.91359340	11.573190	138	138	6.325186e-39	0.87920041	0.9381943

*> FROHvsFPLINK\$results*

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.69085951	0.1828304	138	139	1.0000e+00	-0.76855072	-0.5930235
single_random_raters	ICC2	0.05740032	2.9953105	138	138	1.6807e-10	-0.02667689	0.2205237
single_fixed_raters	ICC3	0.49941312	2.9953105	138	138	1.6807e-10	0.36356650	0.6144618
Average_raters_absolute	ICC1k	-4.46955044	0.1828304	138	139	1.0000e+00	-6.64120197	-2.9142881
Average_random_raters	ICC2k	0.10856876	2.9953105	138	138	1.6807e-10	-0.05481611	0.3613591
Average_fixed_raters	ICC3k	0.66614479	2.9953105	138	138	1.6807e-10	0.53325818	0.7611971

*> FmeanvsFPLINK\$results*

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.82001004	0.09889504	138	139	1.000000e+00	-0.86778198	-0.7571775
single_random_raters	ICC2	0.02659129	2.23442966	138	138	1.638532e-06	-0.01750914	0.1123122
single_fixed_raters	ICC3	0.38165296	2.23442966	138	138	1.638532e-06	0.23025485	0.5150121
Average_raters_absolute	ICC1k	-9.11173070	0.09889504	138	139	1.000000e+00	-13.12653150	-6.2364681
Average_random_raters	ICC2k	0.05180502	2.23442966	138	138	1.638532e-06	-0.03564235	0.2019437
Average_fixed_raters	ICC3k	0.55245850	2.23442966	138	138	1.638532e-06	0.37432058	0.6798786



# Coeficiente de correlación intraclase

```
library(psych)
```

```
A<-matrix(c(datos$Froh,datos$Fmean),ncol=2,byrow=F)
```

```
FROHvsFmean<-ICC(A,missing=TRUE,alpha=0.05,lmer=FALSE,check.keys=FALSE)
```

```
FROHvsFmean$results
```

```
> FROHvsFmean$results
```

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.06087064	0.885244	138	139	7.628415e-01	-0.22424861	0.1059365
single_random_raters	ICC2	0.28794025	11.573190	138	138	6.325186e-39	-0.03666546	0.6571287
single_fixed_raters	ICC3	0.84093138	11.573190	138	138	6.325186e-39	0.78444034	0.8835838
Average_raters_absolute	ICC1k	-0.12963206	0.885244	138	139	7.628415e-01	-0.57814555	0.1915779
Average_random_raters	ICC2k	0.44713294	11.573190	138	138	6.325186e-39	-0.07612196	0.7930931
Average_fixed_raters	ICC3k	0.91359340	11.573190	138	138	6.325186e-39	0.87920041	0.9381943

```
> FROHvsFPLINK$results
```

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.69835951	0.1828304	138	139	1.0000e+00	-0.76855072	-0.5930235
single_random_raters	ICC2	0.05740032	2.9953105	138	138	1.6807e-10	-0.02667689	0.2205237
single_fixed_raters	ICC3	0.49941312	2.9953105	138	138	1.6807e-10	0.36356650	0.6144618
Average_raters_absolute	ICC1k	-4.46955044	0.1828304	138	139	1.0000e+00	-6.64120197	-2.9142881
Average_random_raters	ICC2k	0.10856876	2.9953105	138	138	1.6807e-10	-0.05481611	0.3613591
Average_fixed_raters	ICC3k	0.66614479	2.9953105	138	138	1.6807e-10	0.53325818	0.7611971

```
> FmeanvsFPLINK$results
```

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.82001004	0.09889504	138	139	1.000000e+00	-0.86778198	-0.7571775
single_random_raters	ICC2	0.02659129	2.23442966	138	138	1.638532e-06	-0.01750914	0.1123122
single_fixed_raters	ICC3	0.38165296	2.23442966	138	138	1.638532e-06	0.23025485	0.5150121
Average_raters_absolute	ICC1k	-9.11173070	0.09889504	138	139	1.000000e+00	-13.12653150	-6.2364681
Average_random_raters	ICC2k	0.05180502	2.23442966	138	138	1.638532e-06	-0.03564235	0.2019437
Average_fixed_raters	ICC3k	0.55245850	2.23442966	138	138	1.638532e-06	0.37432058	0.6798786

# Coeficiente de correlación intraclase

```
library(psych)
```

```
A<-matrix(c(datos$Froh,datos$Fmean),ncol=2,byrow=F)
```

```
FROHvsFmean<-ICC(A,missing=TRUE,alpha=0.05,lmer=FALSE,check.keys=FALSE)
```

```
FROHvsFmean$results
```

```
> FROHvsFmean$results
```

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.06087064	0.885244	138	139	7.628415e-01	-0.22424861	0.1059365
single_random_raters	ICC2	0.28794025	11.573190	138	138	6.325186e-39	-0.03666546	0.6571287
single_fixed_raters	ICC3	0.84093138	11.573190	138	138	6.325186e-39	0.78444034	0.8835838
Average_raters_absolute	ICC1k	-0.12963206	0.885244	138	139	7.628415e-01	-0.57814555	0.1915779
Average_random_raters	ICC2k	0.44713294	11.573190	138	138	6.325186e-39	-0.07612196	0.7930931
Average_fixed_raters	ICC3k	0.91359340	11.573190	138	138	6.325186e-39	0.87920041	0.9381943

```
> FROHvsFPLINK$results
```

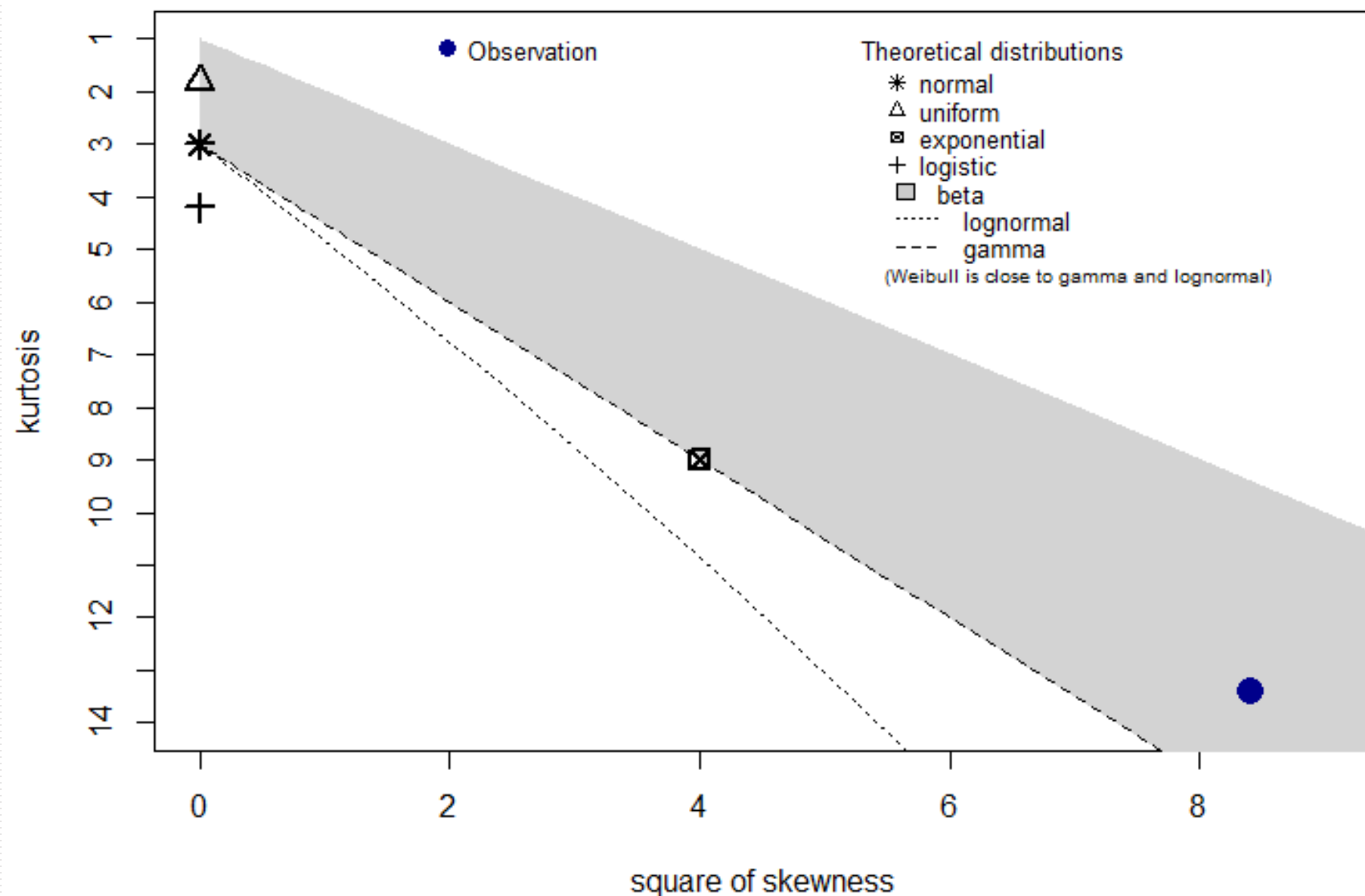
	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.69085951	0.1828304	138	139	1.0000e+00	-0.76855072	-0.5930235
single_random_raters	ICC2	0.05740032	2.9953105	138	138	1.6807e-10	-0.02667689	0.2205237
single_fixed_raters	ICC3	0.49941312	2.9953105	138	138	1.6807e-10	0.36356650	0.6144618
Average_raters_absolute	ICC1k	-4.46955044	0.1828304	138	139	1.0000e+00	-6.64120197	-2.9142881
Average_random_raters	ICC2k	0.10856876	2.9953105	138	138	1.6807e-10	-0.05481611	0.3613591
Average_fixed_raters	ICC3k	0.66614479	2.9953105	138	138	1.6807e-10	0.53325818	0.7611971

```
> FmeanvsFPLINK$results
```

	type	ICC	F	df1	df2	p	lower bound	upper bound
single_raters_absolute	ICC1	-0.820004	0.09889504	138	139	1.000000e+00	-0.86778198	-0.7571775
single_random_raters	ICC2	0.0269129	2.23442966	138	138	1.638532e-06	-0.01750914	0.1123122
single_fixed_raters	ICC3	0.38165296	2.23442966	138	138	1.638532e-06	0.23025485	0.5150121
Average_raters_absolute	ICC1k	-9.11173070	0.09889504	138	139	1.000000e+00	-13.12653150	-6.2364681
Average_random_raters	ICC2k	0.05180502	2.23442966	138	138	1.638532e-06	-0.03564235	0.2019437
Average_fixed_raters	ICC3k	0.55245850	2.23442966	138	138	1.638532e-06	0.37432058	0.6798786

# Modelización del coeficiente de inbreeding $F_{ROH}$

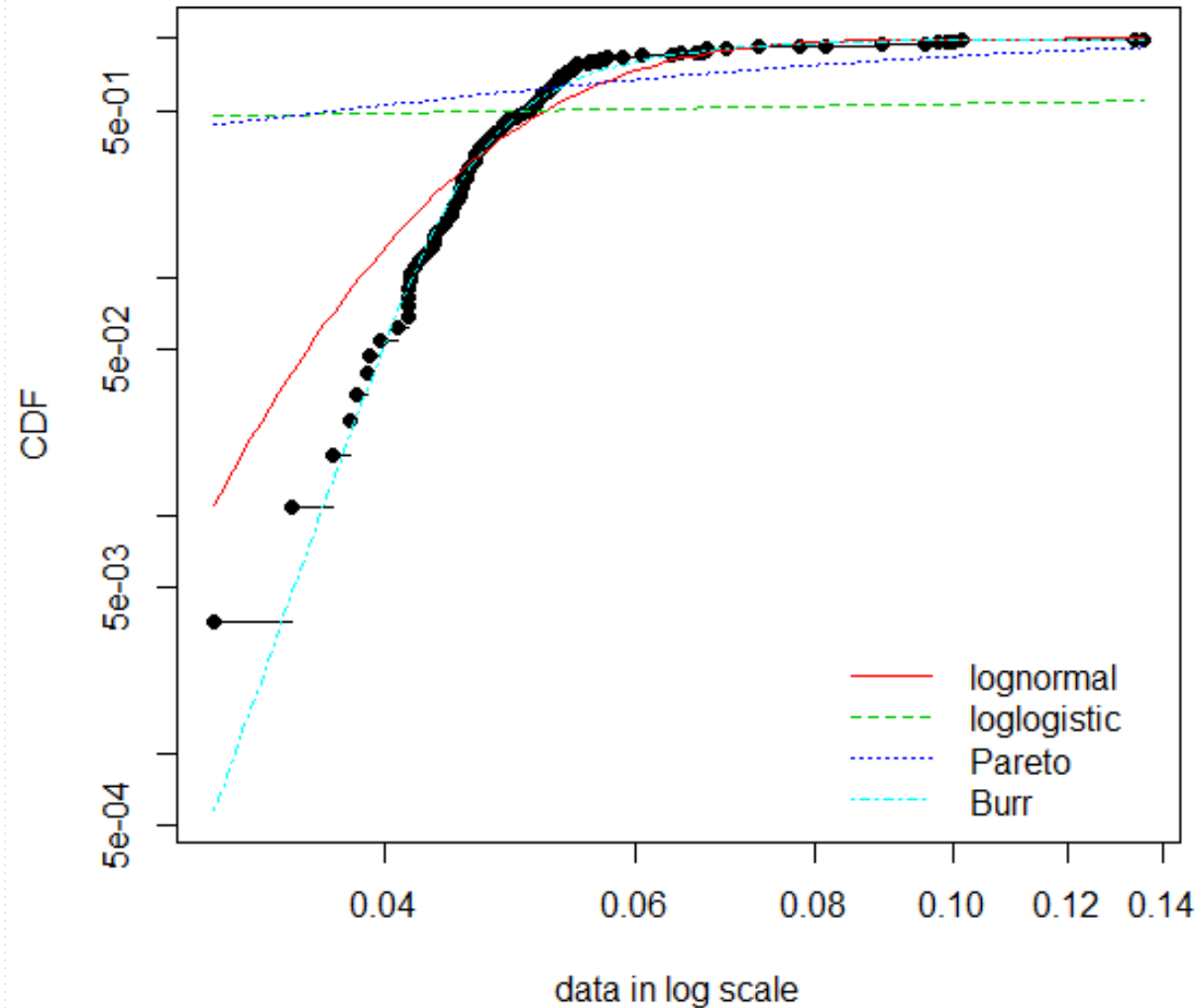
Cullen & Frey graph



## summary statistics

-----  
min: 0.03048681 max: 0.1355066  
median: 0.05056894  
mean: 0.0535173  
estimated sd: 0.01579288  
estimated skewness: 2.898423  
estimated kurtosis: 13.41897

## Empirical and theoretical CDFs

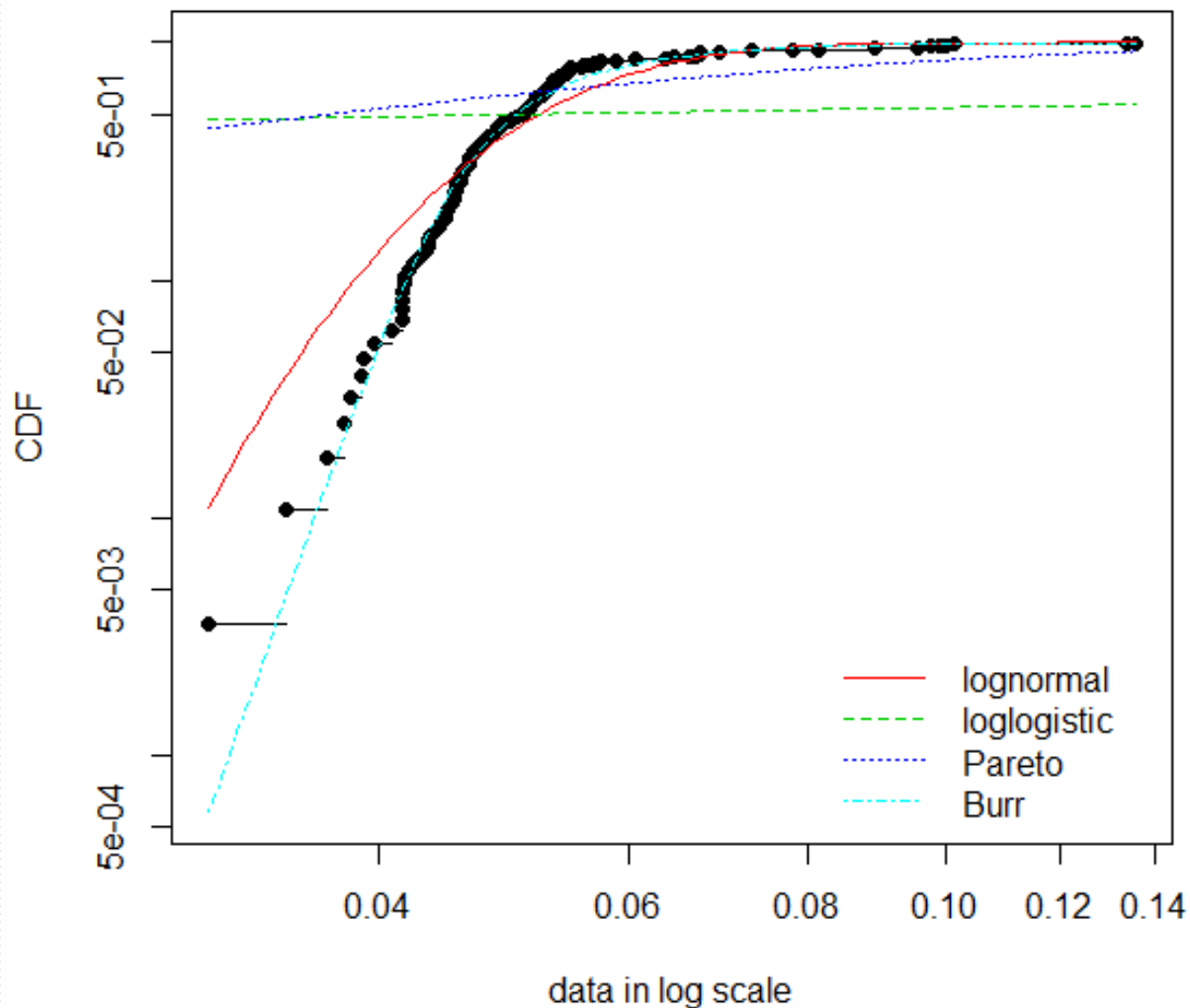


```
library(actuar)
FROH <- Datos_F$Froh
fendo.ln <- fitdist(FROH, "lnorm")
fendo.ll <- fitdist(FROH, "llogis", start = list(shape = 1, scale = 500))
fendo.P <- fitdist(FROH, "pareto", start = list(shape = 1, scale = 500))
fendo.B <- fitdist(FROH, "burr", start = list(shape1 = 0.3, shape2 = 1,
rate = 1))

cdfcomp(list(fendo.ln, fendo.ll, fendo.P, fendo.B), xlogscale = TRUE,
          ylogscale = TRUE, legendtext = c("lognormal", "loglogistic",
"Pareto", "Burr"))

gofstat(list(fendo.ln, fendo.ll, fendo.P, fendo.B), fitnames = c("lnorm",
"llogis", "Pareto", "Burr"))
```

## Empirical and theoretical CDFs



```
library(actuar)
FROH <- Datos_F$Froh
fendo.ln <- fitdist(FROH, "lnorm")
fendo.ll <- fitdist(FROH, "llogis", start = list(shape = 1, scale = 500))
fendo.P <- fitdist(FROH, "pareto", start = list(shape = 1, scale = 500))
fendo.B <- fitdist(FROH, "burr", start = list(shape1 = 0.3, shape2 = 1,
rate = 1))
```

```
cdfcomp(list(fendo.ln, fendo.ll, fendo.P, fendo.B), xlogscale = TRUE,
ylogscale = TRUE, legendtext = c("lognormal", "loglogistic",
"Pareto", "Burr"))
```

```
gofstat(list(fendo.ln, fendo.ll, fendo.P, fendo.B), fitnames = c("lnorm",
"llogis", "Pareto", "Burr"))
```

### Goodness-of-fit criteria

	lnorm	llogis	Pareto	Burr
Akaike's Information Criterion	-829.8173	11.99614	-531.9146	-883.6232
Bayesian Information Criterion	-823.9483	17.86509	-526.0456	-874.8198



❖ Mediante *Bootstrap* se estiman las medias y los intervalos de confianza:

```
bendo.B <- bootdist(fendo.B, niter = 1001)  
summary(bendo.B)
```

```
quantile(bendo.B, probs = c(0.05,0.95))
```

❖ Mediante *Bootstrap* se estiman las medias y los intervalos de confianza:

```
bendo.B <- bootdist(fendo.B, niter = 1001)
summary(bendo.B)
```

```
quantile(bendo.B, probs = c(0.05,0.95))
```

$$F(x) = [1 + (x/b)^g]^{-s}$$

```
Parametric bootstrap medians and 95% percentile CI
      Median      2.5%      97.5%
shape1 0.3143209 0.1902101 0.5406003
shape2 17.3643821 12.1940783 25.1086202
rate    22.5818605 21.2957323 23.6129058
```

```
(original) estimated quantiles for each specified
probability (non-censored data)
```

```
      p=0.05      p=0.95
estimate 0.03991963 0.07706692
Median of bootstrap estimates
      p=0.05      p=0.95
estimate 0.04000877 0.07678897
```

```
two-sided 95 % CI of each quantile
```

```
      p=0.05      p=0.95
2.5 % 0.03826478 0.07003586
97.5 % 0.04153261 0.08561908
```

❖ Mediante *Bootstrap* se estiman las medias y los intervalos de confianza:

```
bendo.B <- bootdist(fendo.B, niter = 1001)
summary(bendo.B)
```

```
quantile(bendo.B, probs = c(0.05,0.95))
```

$$F(x) = [1 + (x/b)^g]^{-s}$$

```
Parametric bootstrap medians and 95% percentile CI
      Median      2.5%      97.5%
shape1  0.3143209  0.1902101  0.5406003
shape2 17.3643821 12.1940783 25.1086202
rate    22.5818605 21.2957323 23.6129058
```

```
(original) estimated quantiles for each specified
probability (non-censored data)
```

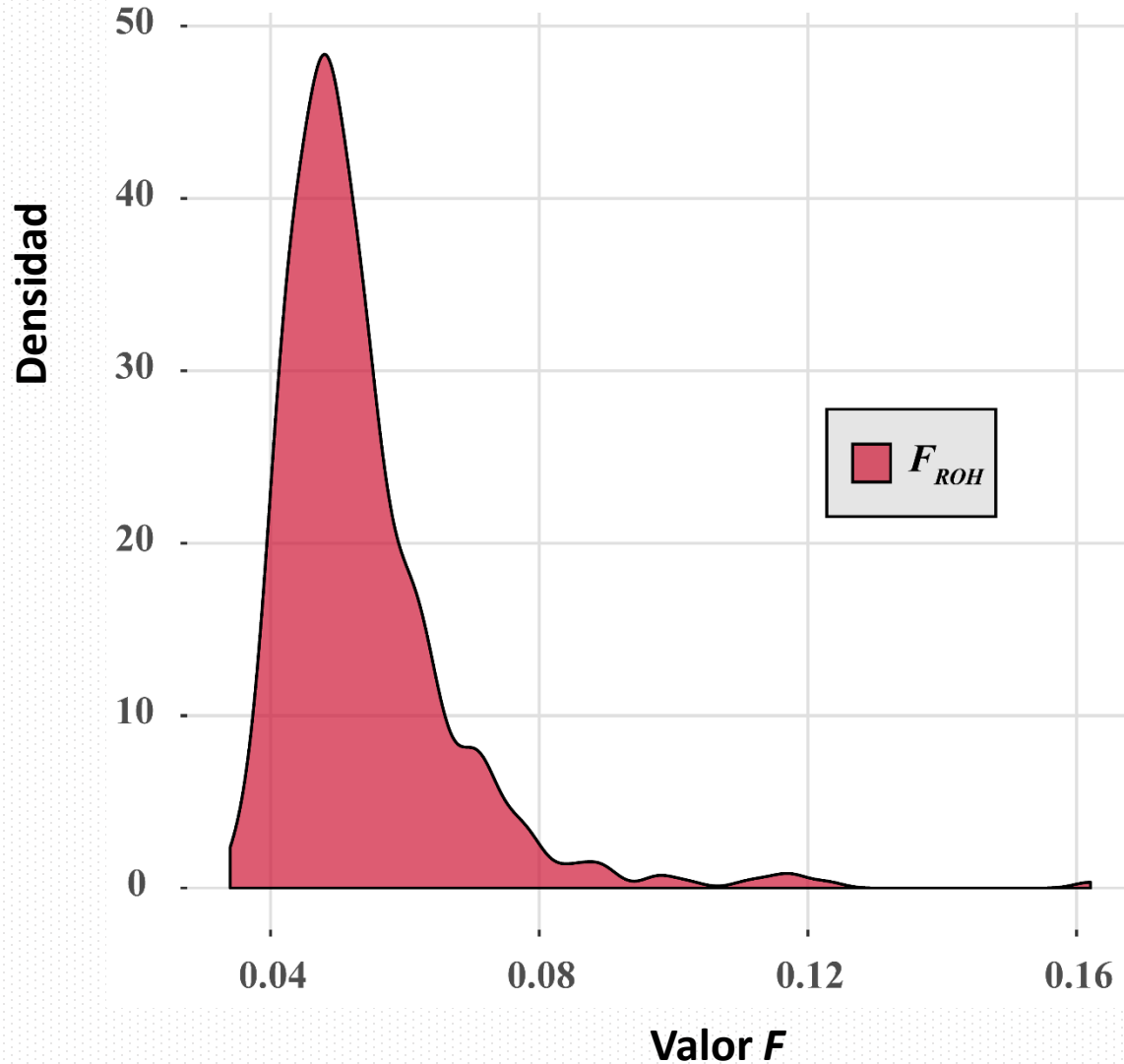
```
      p=0.05      p=0.95
estimate 0.03991963 0.07706692
Median of bootstrap estimates
      p=0.05      p=0.95
estimate 0.04000877 0.07678897
```

```
two-sided 95 % CI of each quantile
```

```
      p=0.05      p=0.95
2.5 %  0.03826478 0.07003586
97.5 % 0.04153261 0.08561908
```



❖ Simulación de datos y representación de su densidad mediante *ggplot2*:



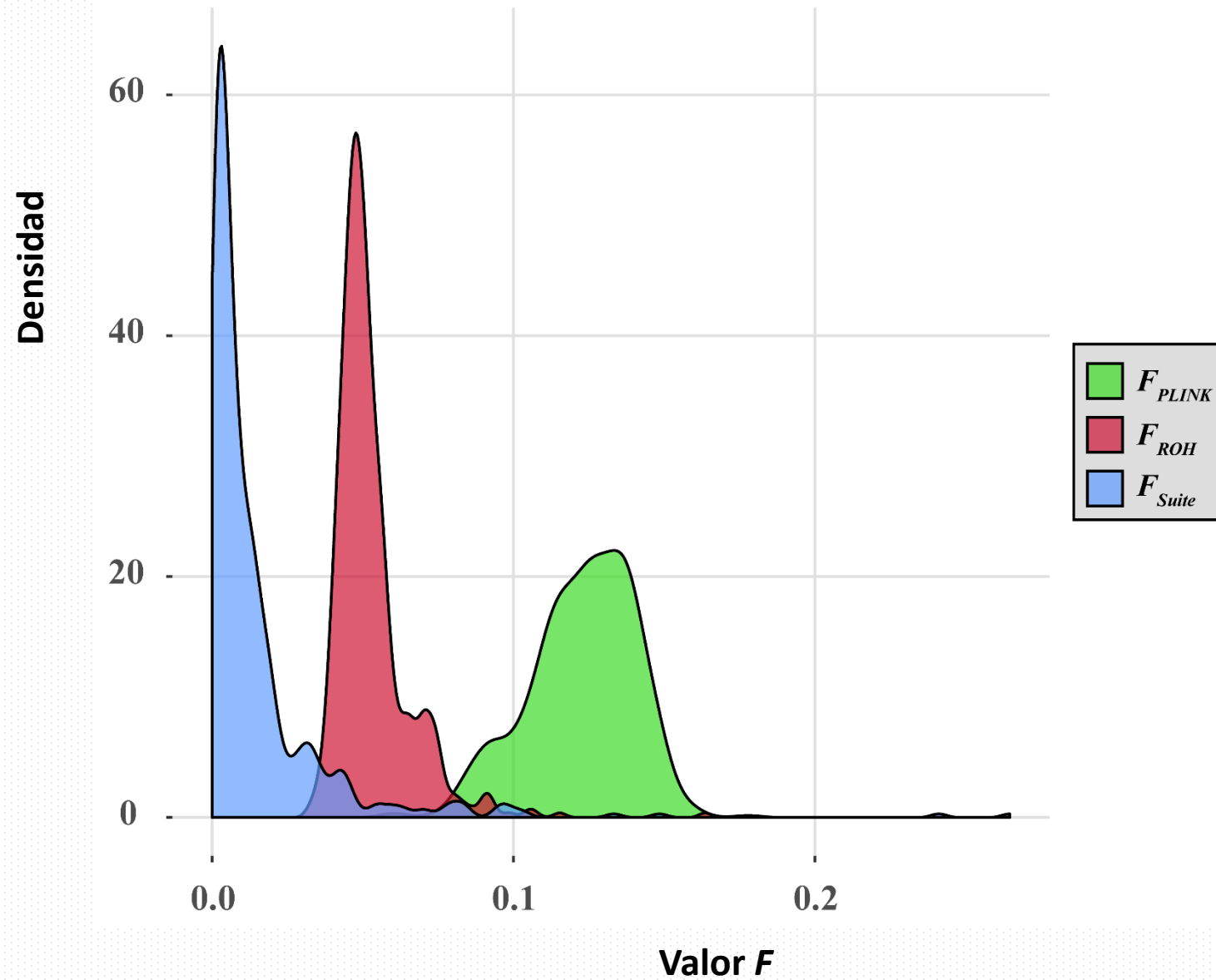
```
x<-rburr(n=500, shape1 = 0.3147055, shape2 =  
17.2429319, rate = 22.5877735, scale = 0.0442717)
```

```
library(ggplot2)
```

```
df = data.frame(datos,Estimador.F=as.factor(grupo))
```

```
ggplot(df, aes(datos, fill = Estimador.F)) + geom_density  
(alpha = 0.2)
```

❖ Simulación de datos y representación de su densidad mediante *ggplot2*:



## ❖ Modelo de regresión logística empleando el valor $F$ . Aplicación en un caso real.

$F_{ROH}$

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
    data = ModlogisFROH)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7262	-1.0530	-0.3375	1.1551	2.1615

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.89436	0.07101	-40.76	<2e-16 ***
datos	52.80186	1.31821	40.06	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27726 on 19999 degrees of freedom  
Residual deviance: 24832 on 19998 degrees of freedom  
AIC: 24836

Number of Fisher Scoring iterations: 7

$F_{PLINK}$

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
    data = Modlogis)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6712	-0.9573	-0.1556	0.8503	5.4491

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	6.0799	0.1109	54.83	<2e-16 ***
datos	-51.3698	0.9131	-56.26	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27726 on 19999 degrees of freedom  
Residual deviance: 22853 on 19998 degrees of freedom  
AIC: 22857

Number of Fisher Scoring iterations: 4

## ❖ Modelo de regresión logística empleando el valor $F$ . Aplicación en un caso real.

$F_{ROH}$

0 = Iberia  
1 = Marruecos

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
    data = ModlogisFROH)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7262  -1.0530  -0.3375   1.1551   2.1615
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.89436    0.07101  -40.76  <2e-16 ***
datos       52.80186    1.31821   40.06  <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 27726  on 19999  degrees of freedom
Residual deviance: 24832  on 19998  degrees of freedom
AIC: 24836
```

```
Number of Fisher Scoring iterations: 7
```

0.528 → 0.01 unidad de  $F$

$IC_{95\%} : (e^{0.502}; e^{0.553}) = 1.65; 1.74$

$F_{PLINK}$

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
    data = Modlogis)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6712  -0.9573  -0.1556   0.8503   5.4491
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   6.0799    0.1109   54.83  <2e-16 ***
datos       -51.3698    0.9131  -56.26  <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 27726  on 19999  degrees of freedom
Residual deviance: 22853  on 19998  degrees of freedom
AIC: 22857
```

```
Number of Fisher Scoring iterations: 4
```

## ❖ Modelo de regresión logística empleando el valor $F$ . Aplicación en un caso real.

$F_{ROH}$

0 = Iberia  
1 = Marruecos

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
     data = ModlogisFROH)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7262  -1.0530  -0.3375   1.1551   2.1615
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.89436    0.07101  -40.76  <2e-16 ***
datos       52.80186    1.31821   40.06  <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 27726  on 19999  degrees of freedom
Residual deviance: 24832  on 19998  degrees of freedom
AIC: 24836
```

```
Number of Fisher Scoring iterations: 7
```

0.528 → 0.01 unidad de  $F$

$IC_{95\%} : (e^{0.502}; e^{0.553}) = 1.65; 1.74$

$F_{PLINK}$

0 = Iberia  
1 = Marruecos

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
     data = Modlogis)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6712  -0.9573  -0.1556   0.8503   5.4491
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   6.0799    0.1109   54.83  <2e-16 ***
datos       -51.3698    0.9131  -56.26  <2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 27726  on 19999  degrees of freedom
Residual deviance: 22853  on 19998  degrees of freedom
AIC: 22857
```

```
Number of Fisher Scoring iterations: 4
```

-0.513 → 0.01 unidad de  $F$

$IC_{95\%} : (e^{-0.531}; e^{-0.495}) = 0.58; 0.61$

## ❖ Modelo de regresión logística empleando el valor $F$ . Aplicación en un caso real.

$F_{ROH}$

0 = Iberia  
1 = Marruecos

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
    data = ModlogisFROH)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7262	-1.0530	-0.3375	1.1551	2.1615

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.89436	0.07101	-40.76	<2e-16 ***
datos	52.80186	1.31821	40.06	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27726 on 19999 degrees of freedom  
Residual deviance: 24832 on 19998 degrees of freedom  
AIC: 24836

Number of Fisher Scoring iterations: 7

0.528 → 0.01 unidad de  $F$

$IC_{95\%} : (e^{0.502}; e^{0.553}) = 1.65; 1.74$

$F_{Suite}$

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
    data = ModlogisFsuite)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-8.4904	-0.9132	-0.4091	1.0725	1.5872

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.92599	0.02182	-42.44	<2e-16 ***
datos	51.97062	1.05514	49.26	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27726 on 19999 degrees of freedom  
Residual deviance: 23413 on 19998 degrees of freedom  
AIC: 23417

Number of Fisher Scoring iterations: 6

## ❖ Modelo de regresión logística empleando el valor $F$ . Aplicación en un caso real.

$F_{ROH}$

0 = Iberia  
1 = Marruecos

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
    data = ModlogisFROH)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7262	-1.0530	-0.3375	1.1551	2.1615

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.89436	0.07101	-40.76	<2e-16 ***
datos	52.80186	1.31821	40.06	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27726 on 19999 degrees of freedom  
Residual deviance: 24832 on 19998 degrees of freedom  
AIC: 24836

Number of Fisher Scoring iterations: 7

0.528 → 0.01 unidad de  $F$

$IC_{95\%} : (e^{0.502}; e^{0.553}) = 1.65; 1.74$

$F_{Suite}$

0 = Iberia  
1 = Marruecos

```
Call:
glm(formula = clase ~ datos, family = binomial(link = "logit"),
    data = ModlogisFSuite)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-8.4904	-0.9132	-0.4091	1.0725	1.5872

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.92599	0.02182	-42.44	<2e-16 ***
datos	51.97062	1.05514	49.26	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27726 on 19999 degrees of freedom  
Residual deviance: 23413 on 19998 degrees of freedom  
AIC: 23417

Number of Fisher Scoring iterations: 6

0.519 → 0.01 unidad de  $F$

$IC_{95\%} : (e^{0.499}; e^{0.540}) = 1.64; 1.71$

# Conclusiones más relevantes

- Los valores de los coeficientes de inbreeding genómico hallados en las poblaciones asentadas en el Mediterráneo occidental son, en promedio, más altos en las poblaciones Bereberes de Marruecos frente a las poblaciones Ibéricas, salvo en la metodología  $F_{PLINK}$  donde esta tendencia se revierte.



# Conclusiones más relevantes

- Los valores de los coeficientes de inbreeding genómico hallados en las poblaciones asentadas en el Mediterráneo occidental son, en promedio, más altos en las poblaciones Bereberes de Marruecos frente a las poblaciones Ibéricas, salvo en la metodología  $F_{PLINK}$  donde esta tendencia se revierte.
- El valor de la medida del inbreeding genómico aportada por la metodología  $F_{suite}$  subestima el valor aportado por  $F_{ROH}$  y a su vez, la medida de  $F_{PLINK}$  sobreestima el valor de las dos anteriores.

# Conclusiones más relevantes

- Los valores de los coeficientes de inbreeding genómico hallados en las poblaciones asentadas en el Mediterráneo occidental son, en promedio, más altos en las poblaciones Bereberes de Marruecos frente a las poblaciones Ibéricas, salvo en la metodología  $F_{PLINK}$  donde esta tendencia se revierte.
- El valor de la medida del inbreeding genómico aportada por la metodología  $F_{suite}$  subestima el valor aportado por  $F_{ROH}$  y a su vez, la medida de  $F_{PLINK}$  sobreestima el valor de las dos anteriores.
- Las metodologías  $F_{ROH}$  y  $F_{suite}$  presentaron una concordancia buena entre sus valores hallados de inbreeding genómico ( $ICC : 0.84$ ), siendo prácticamente idénticos los resultados de ambas en el modelo de regresión logística. Mientras que, tanto  $F_{ROH}$  como  $F_{suite}$ , presentaron con la medida de  $F_{PLINK}$  una concordancia pobre ( $ICC < 0.5$ ).

## ➤Referencias Bibliográficas Citadas:

- Gazal S, Sahbatou M, Perdry H, Letort S, Génin E, Leutenegger AL. Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap III. Hum Hered. 2014; 77(1-4): 49-62.
- Ghani M, Reitz C, Cheng R, Vardarajan BN, Jun G, Sato C, et al. Association of Long Runs of Homozygosity With Alzheimer Disease Among African American Individuals. JAMA Neurol. 2015;72(11):1313-23.
- Keller MC, Visscher PM, Goddard ME. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 2011; 189(1): 237-249.
- Keller MC, Simonson MA, Ripke S, Neale BM, Gejman PV, Howrigan DP et al. Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. PLoS Genet. 2012; 8(4):e1002656.
- Simón-Sánchez J, Kilarski LL, Nalls MA, Martinez M, Schulte C, Holmans P, et al. Cooperative Genome-Wide Analysis Shows Increased Homozygosity in Early Onset Parkinson's Disease. PLoS One. 2012;7(3):e28787.

# MUCHAS GRACIAS POR SU ATENCIÓN



## AGRADECIMIENTOS



Proyecto de Investigación  
CGL-2014-53985-R