

# AdFiT (Admixture Files Tool)

Version 1.0.5

Un programa creado por Géraud GOURJON  
Post-doctoral researcher – Université Aix-Marseille II

Con la participación de Anna DEGIOANNI  
« Maître de Conférences » – Université Aix-Marseille II

UMR6578, Unité d'Anthropologie Bioculturelle  
Aix-Marseille Université / CNRS / EFS  
Faculté de Médecine, Avenue Pierre Dramard, Marseille

ggourjon@hotmail.com

Enero 2009

Esta documentación describe las características y la utilización de AdFiT, un programa destinado a la creación simplificada de archivos de entrada para software de estimación de mezcla genética.

Si usted encuentra algún problema o dificultad en la utilización de este programa, contáctenos vía email.

## Referencia:

Gourjon G, Degioanni A (2009) AdFiT v1.7 (Admixture File Tool): input files creating tool for genetic admixture estimation software. *Bulletins et Mémoires de la Société d'Anthropologie de Paris* **21**, 223-229.

## El Programa

Este programa de utilización sencilla se concibió para ayudar a los investigadores en la creación de archivos de entrada para varios software de estimación de mezcla genética. El programa entrega archivos de entrada para los siguientes softwares:

Software	Creador	Año	Métodos implantados	Publicaciones (métodos)	Sitios de descarga
ADMIX	Long	1991	Weighted Squares	Least (Long 1991)	<a href="http://www.genetica.fmed.edu.uy/">http://www.genetica.fmed.edu.uy/</a>
ADMIX95	Bertoni	1995	Gene Identity	(Chakraborty 1985)	Disponible solicitándolo al Pr. S.J. Long ( <a href="mailto:longjc@umich.edu">longjc@umich.edu</a> )
Mistura	Cabello Krieger	1997	Maximum Likelihood	(Krieger et al. 1965) (Cabello and Krieger 1997) (Krieger and Cabello 1997)	<a href="http://web.unife.it/progetti/genetica/Isabelle/admix2_0.html">http://web.unife.it/progetti/genetica/Isabelle/admix2_0.html</a>
Admix 2.0	Bertorelle Dupanloup	1998 2001	Coalescence-based	(Bertorelle and Excoffier 1998) (Dupanloup and Bertorelle 2001)	Disponible solicitándolo al Pr. H. Krieger ( <a href="mailto:hkrieger@icb.usp.br">hkrieger@icb.usp.br</a> )
LEA	Beaumont Langella	2001	Coalescent-based maximum Likelihood	(Chikhi et al. 2001) (Langella et al. 2001)	<a href="http://www.rubic.rdg.ac.uk/~mab/software.html">http://www.rubic.rdg.ac.uk/~mab/software.html</a>
LEADMIX	Wang	2003	Maximum likelihood	(Wang 2003)	<a href="http://www.zoo.cam.ac.uk/ioz/people/wang.htm">http://www.zoo.cam.ac.uk/ioz/people/wang.htm</a>
			Least-Square	(Roberts and Hiorns 1965)	
			Weighted Squares	Least (Long 1991) (Chakraborty and Srinivasan 1992)	
			Coalescent-based	(Bertorelle and Excoffier 1998)	

Antes de cualquier utilización de AdFiT y para cumplir correctamente con los parámetros específicos de cada software de estimación de mezcla genética (como por ejemplo el tamaño de la muestra para ADMIX, la tasa de mutación y el tiempo después de la mezcla para Admix 2.0 o la Divergencia genética para LEADMIX), se recomienda leer atentamente la documentación del software correspondiente y la publicación de referencia del método.

Toda la información se entrega para la versión francesa, diríjase a la documentación en inglés o en español si usted utiliza una lengua diferente

Desde la apertura de AdFiT, usted tiene la opción de seleccionar la lengua con la cual desea trabajar. Todos los parámetros, campos y mensajes de AdFiT estarán en la lengua de su elección.

AdFiT es un ejecutable sin instalación. Por ello, desde la primera utilización del programa se crean varios archivos en la misma carpeta donde se encuentra el ejecutable (dll necesarios para el funcionamiento del programa). Estos archivos dll deben estar presentes para que el programa funcione. Lo ideal es ubicar el ejecutable en una carpeta especialmente creada para ello.

## Creación de los Archivos de entrada

### 1. CREACIÓN DE UN ARCHIVO COMÚN SOBRE EXCEL PREVIAMENTE A LA EJECUCIÓN DE AdFiT

La primera manipulación necesita el agrupamiento de todos los datos en un archivo común (figura 1) en el formato Excel. AdFiT necesita las librerías Microsoft® Office para funcionar correctamente.

Primera línea: Nombres de los Loci (comenzando en la segunda columna)

Segunda línea: Nombres de los Alelos para cada Locus (comenzando en la segunda columna)

Tercera línea:

Primera columna: Nombre de la población mezclada (población híbrida, hija)

Siguientes columnas: Frecuencias alélicas para cada Locus (el total de las frecuencias alélicas para cada locus debe ser igual a 1)

Este archivo de datos será leído directamente por AdFiT.

Lo único que hay que hacer es seleccionar en AdFiT el software de estimación de mezcla para el cual el archivo de entrada debe ser creado y llenar eventualmente los parámetros específicos del software (ver a continuación).

Figura 1 Archivo de datos común (formato Excel)

	A	B	C	D	E	F	G
1		ABO			Duffy		
2		A	B		O FYA	FYB	FYO
3	Admixed Population	0,141	0,069	0,79	0,061	0,009	0,93
4	Parent Population 1	0,153	0,127	0,72	0,052	0	0,948
5	Parent Population 2	0,144	0,123	0,733	0,162	0,15	0,688
6	Parent Population 3	0,221	0,184	0,595	0,483	0,393	0,124
7	Parent Population 4	0,358	0	0,642	0,842	0,156	0,002

Nombres de los loci

Nombres de los alelos

Nombre de la población mezclada

Nombres de las poblaciones paternas

Frecuencia de los alelos para cada locus (para todas las poblaciones)

## 2. PROCEDIMIENTO DE EJECUCIÓN DE ADFIT

Una vez que el archivo Excel ha sido creado, debe abrirse punteando sobre el botón “Abrir un archivo de datos Excel” en la ventana de AdFit. La tabla se llena automáticamente en AdFit. La primera línea de la tabla de datos tiene que ver con la población mezclada que ha sido seleccionada. La creación del archivo de entrada se hace inmediatamente en tres etapas:

- 1) Elegir en la tabla de datos las poblaciones parentales. Para ello, solamente hay que puntear en la columna «Sel» de la tabla de Datos: las poblaciones parentales seleccionadas se subrayan en azul cielo.
- 2) Seleccionar los loci. Se puede utilizar el conjunto de informaciones punteando sobre «seleccionar todos los loci» o bien, elegir algunos loci en la última «Sel» de la tabla de los loci: los loci seleccionados se subrayan en azul cielo en la tabla de los Loci y en naranja en la tabla de los Datos.
- 3) Seleccionar a continuación el software de estimación de mezcla y llenar los parámetros específicos si es necesario (para ello diríjase a la documentación de cada software) en seguida, púntee sobre el botón “Crear”. El archivo de entrada para el software deseado ya ha sido generado sobre el Escritorio de su computador.

**Abrir un archivo de datos**      **Tabla de datos**      **Tabla de los loci**

AdFIT - Creado por Gérard GOURJO - JMR, 6578, Marseilles, France

Abrir un archivo de datos (Excel)      Elegir sus loci:

No Columnas: 0      Nombre del estudio: \_\_\_\_\_

No Lineas: 0

Seleccione las poblaciones parentales: 0      Selección pobl      Selección Loci

Sel COL1

Crear un archivo para:

Admix     LEADMIX

ADMIX 95     Admix 2.0

LEA 1.0     Mistura

HEADMIX     Miztur

**Crear**

ADMIX: Muestra: 1.000

Muestra LEA:  50  100  200  500  1000

LEADMIX:  Muestra faltante     Div. Gen.

Muestra LEADMIX:  100  1000  10000

Deriva: 0,00001

Integrated Points: 0 30 50 100

Initial Points: 20

Monitoring: 5

ADMIX 2.0: Tiempo desde la mezcla: 0

Índice de mutación: 0

Divergenca molecular:  100  1000  10000

Muestra Admix 2.0:  100  1000  10000

Archivo Mistur2: \_\_\_\_\_

Mistura: Muestra: 100

Archivo de Ferropos: \_\_\_\_\_

Crear    Archivo

Archivo (Mistura): \_\_\_\_\_

**Cerrar**

**Botón "Crear"**      **Parámetros específicos de los software**      **Botón "Cerrar"**

**Zona de selección del software de estimación**

Los archivos de entrada son nombrados automáticamente. Si usted desea crear varios archivos de entrada del mismo tipo (con marcadores diferentes, y según el software de poblaciones parentales diferentes, ver para cada software el nombre dado) es aconsejable cambiarles el nombre, borrarlos o moverlos a otra carpeta para no perderlos.

Una de nuestras prioridades era concebir un programa lo más intuitivo posible. La mayoría de errores posibles cuando el archivo es creado, ha sido anticipada y tomada en cuenta. Mensajes de error o de información fueron introducidos en AdFiT para explicar los potenciales problemas. Estos riesgos pueden prevenirse en algunos casos. Por ejemplo, LEA (Langella et al. 2001) considera un evento de mezcla que implica únicamente dos poblaciones parentales. Por lo tanto, será imposible generar un archivo de entrada cuando más de dos poblaciones parentales son seleccionadas (un mensaje de error aparecerá en la pantalla). Inversamente, cuando para un alelo dado, las frecuencias para todas las poblaciones son iguales a 0, la mayoría de software de estimación no funcionará. Si bien, es posible crear estos archivos de entrada con AdFiT, un mensaje informativo aparecerá para señalar el riesgo potencial de rechazo del archivo de entrada para el software de estimación, y los alelos para los cuales las frecuencias son iguales a 0 para todas las poblaciones serán subrayados en rojo en la tabla de datos.

### **3. PARÁMETROS ESPECÍFICOS**

#### *ADMIX95 input file*

Ningún parámetro específico debe ser indicado para este programa. El archivo de entrada es creado automáticamente. Se le asigna el nombre a partir de las dos letras de las poblaciones parentales seleccionadas (la extensión del archivo para ADMIX95 es .inp)

#### *ADMIX input file*

El tamaño de la muestra debe ser indicado (ver la documentación de ADMIX). Esto influye únicamente el valor de  $F_{st}$ . Se le asigna el nombre a partir de las dos letras de las poblaciones parentales seleccionadas (la extensión del archivo para ADMIX es .dat)

#### *LEA input file*

El tamaño de la muestra debe ser indicado (ver la documentación de LEA). Ello puede tener un peso en la estimación de la mezcla para LEA. Aumentar el tamaño de la muestra aumenta la precisión de los resultados pero el tiempo de simulación puede alargarse demasiado. El archivo creado para AdFiT es llamado "infile" (sin extensión). Todos los archivos son llamados de la misma manera (formato LEA) así que es necesario moverlos a otra carpeta antes de crear un nuevo archivo de entrada (no les cambie el nombre, LEA necesita un archivo de entrada exclusivamente llamado "infile").

#### *Admix 2.0 input file*

Para la creación del archivo de entrada, un archivo Excel intermediario el cual contiene las matrices de divergencias moleculares entre los alelos para cada locus debe ser creado. AdFiT

genera ese archivo parcialmente pre-diligenciado con los loci y los alelos. Para ello, una vez que el archivo de datos sea seleccionado (y los datos mostrados en la tabla de datos), puntee sobre “Seleccionar todo” (al lado de la tabla de los Loci), seleccione Admix 2.0 en la zona de selección del software y puntee sobre “Crear” en la zona de los parámetros específicos de Admix 2.0 debajo de “Matrices de divergencia molecular”. Ello crea sobre el Escritorio el archivo Excel intermediario con las matrices pre-formateadas para todos los loci. Abra el archivo a partir de Excel y remplace los «0» con los valores de divergencia molecular entre los alelos para cada locus y guarde el archivo (utilice “guardar como” en el menú y no “guardar”) en formato Excel 97-2003. Tenga cuidado de no grabar en formato texto (txt). Este archivo es necesario para crear el archivo de entrada. Debe llenarse correctamente si la diferenciación molecular es tomada en cuenta. En el caso contrario, AdFiT llena automáticamente las matrices para todos los loci con 0 (considerando ninguna diferencia molecular entre los alelos). Diríjase a la documentación del software Admix 2.0 para llenar correctamente las matrices. Sin embargo, si usted utiliza Office 2007, es **obligatorio** realizar un proceso ya que existen problemas de compatibilidad aunque usted no utilice la diferenciación (ver abajo, la parte **IMPORTANTE** página 7).

**Columna de índices de los loci**

	A	B	C	D
1	1	ABO		
2	1		3	
3	1	0		
4	1	0	0	
5	1	0	0	0
6	2	Duffy		
7	2		3	
8	2		0	
9	2		0	0
10	2		0	0
11	3	Kell		
12	3		2	
13	3		0	
14	3		0	0
15	4	Kidd		

Si un archivo de matrices ha sido previamente creado, usted puede seleccionarlo con el botón “Archivo” situado al lado del botón de creación de matrices. Atención: Seleccione un archivo que contenga las matrices para todos los loci del archivo de datos, los loci deben estar en el mismo orden que en la tabla de los Loci.

El nombre del archivo seleccionado para la creación de un archivo de entrada aparece bajo los dos botones en la casilla gris.

Igualmente, los otros parámetros deben ser indicados correctamente (diríjase a la documentación de Admix 2.0): tiempo desde la mezcla (ingresar 0 si el tiempo es desconocido); tasa de mutación (debe ser diferente de 0 si el tiempo desde la mezcla ha sido previamente especificado); y el tamaño de la muestra.

Cuando los parámetros y el archivo de matrices son diligenciados, seleccione las poblaciones parentales y los loci para el archivo de entrada y puntee sobre el botón “Creación” (bajo la zona de selección de software de estimación).

Se le da nombre al archivo a partir de las dos primeras letras de las poblaciones parentales seleccionadas (la extensión es .txt)

**IMPORTANTE:** Algunos problemas de compatibilidad aparecen con Office versión 2007. Para evitar esos problemas, abra el archivo de Matrices y guárdelo con “grabar todos” en formato Excel 97-2003 antes de crear el archivo de entrada. Si esta etapa no se realiza, Office 2007 considera el archivo Matriz como un archivo txt en lugar de un archivo xls y AdFiT no podrá crear el archivo de entrada.

NOTA : Cuando usted abre el archivo Matriz por primera vez con Office 2007, un mensaje de error aparece, puntee sobre « Si » y remplace tranquilamente los « 0 » con los valores de divergencia ☺

### *LEADMIX input file*

Diríjase a la documentación de LEADMIX para diligenciar sin errores estos parámetros.

#### Parámetros específicos:

*Muestras faltantes:* Señale esta casilla si los datos son faltantes para una población parental.

*Div. Gen.:* Señale esta casilla si la diferenciación genética debe ser tomada en cuenta.

*Tamaño de la muestra:* Seleccione un tamaño de la muestra (esto no modifica los resultados de los estimadores puntuales).

*Deriva:* Indique la tasa de deriva genética.

*Initial points et Int. points:* Diríjase a la documentación de LEADMIX para indicar estos parámetros.

*Monitoreo:* Indique un número entero entre 1 y 5. Esto no afecta los resultados pero permite mostrar más detalles a la hora de la simulación con LEADMIX.

### *Mistura input file*

Para la realización de un archivo de entrada para Mistura, se debe crear un archivo intermediario. Se trata de un archivo que sirve para establecer la correspondencia existente entre genotipos y fenotipos para cada locus. El utilizador debe indicar qué genotipos corresponden a cada fenotipo. Este archivo intermediario es parcialmente diligenciado. Para ello, una vez que el archivo de datos seleccionado (y los datos mostrados en la tabla de datos), puntee sobre “Seleccionar todo” (al lado de la tabla de los Loci), seleccione Mistura en la zona de selección del software y puntee sobre “Crear” en la zona de parámetros específicos de Mistura, bajo “Archivo de fenotipos”. Esto crea sobre el Escritorio el archivo Excel intermediario con las matrices pre-formateadas para todos los loci (ver abajo).

	Indice de los loci		Indice de los alelos (1 para el primer alelo, 2 para el segundo, etc)		
	A	B	C	D	E
1	ID	ALLELES	ID ALLELES	FREQ.PHENC	ID PHENO
2	1	ABO			
3	1		3		
4	1	A			
5	1	B			
6	1	O			
7	1	A A	1 1	0,019881	
8	1	A B	1 2	0,019458	
9	1	A O	1 3	0,22278	
10	1	B B	2 2	0,004761	
11	1	B O	2 3	0,10902	
12	1	O O	3 3	0,6241	
13	2	Duffy			
14	2		3		
15	2	FYA			
16	2	FYB			
17	2	FYO			
18	2	FYA FYA	1 1	0,003721	
19	2	FYA FYB	1 2	0,001098	
20	2	FYA FYO	1 3	0,11346	
21	2	FYB FYB	2 2	0,000081	

Nombre del locus  
 Número de alelos de ese locus  
 Nombre de los Alelos de ese locus  
 Frecuencias de los fenotipos (calculadas con Hardy-Weinberg)  
 Genotipos de ese locus

Abrir este archivo con Excel (si se utiliza Office 2007 aparecerá un mensaje de error, púntee sobre Si). La columna ID PHENO debe llenarse de la siguiente manera: ingrese la cifra 1 frente a cada Genotipo correspondiente al Fenotipo 1; ingrese la cifra 2 frente a cada Genotipo correspondiente al Fenotipo 2, etc. Las cifras utilizadas deben ser enteras y sucesivas. Por ejemplo, para 4 fenotipos diferentes, serán nombrados 1, 2, 3, 4 y no 1, 2, 4, 6. Ver el ejemplo para el Grupo Sanguíneo ABO.

Inmediatamente, guarde el archivo (utilizar “guardar como” en el menú y no “guardar”) en formato Excel 97-2003. Tenga cuidado de no grabar en formato texto (txt). Esta operación es importante y obligatoria ya que existen problemas de compatibilidad con Office 2007 (ver abajo la parte **IMPORTANTE** página 9). Contrariamente a la creación del archivo para Admix 2.0, esta operación es una etapa estrictamente obligatoria dada la necesidad de indicar la columna ID PHENO.

	A	B	C	D	E
1	ID	ALLELES	ID ALLELES	FREQ.PHENC	ID PHENO
2	1	ABO			
3	1		3		
4	1	A			
5	1	B			
6	1	O			
7	1	A A	1 1	0,019881	1
8	1	A B	1 2	0,019458	3
9	1	A O	1 3	0,22278	1
10	1	B B	2 2	0,004761	2
11	1	B O	2 3	0,10902	2
12	1	O O	3 3	0,6241	4
13	2	Duffy			
14	2		3		
15	2	FYA			
16	2	FYB			
17	2	FYO			

Nombre del locus  
 Número de alelos de ese locus  
 Nombre de los alelos de ese locus  
 El número 1 corresponde al primer fenotipo (aquí A)  
 El número 3 corresponde al tercer fenotipo (aquí AB)  
 El número 2 corresponde al segundo fenotipo (aquí B)  
 El número 4 corresponde al cuarto fenotipo (aquí O)  
 Génotypes présents à ce locus

Si un archivo de Fenotipos ya ha sido creado, usted puede seleccionar con el botón “Archivo” ubicado al lado del botón de creación. Atención: Seleccione un archivo que contenga los fenotipos para todos los loci del archivo de datos, los loci debe estar en el mismo orden que en la tabla de los Loci. El nombre del archivo seleccionado para la creación del archivo de entrada aparece bajo esos dos botones en la casilla gris.

Cuando el tamaño de la muestra y el archivo de fenotipos son diligenciados; púntee sobre el botón “Creación” (bajo la zona de selección de software de estimación). El archivo es llamado “Mis” + la primera letra de las poblaciones parentales seleccionadas (la extensión es .txt).

Mistura necesita que el nombre del estudio esté presente en el archivo de entrada. Este nombre corresponde por defecto al nombre de la población mezclada pero puede ser modificado en el campo “Nombre del estudio” de la ventana de AdFiT.

**IMPORTANTE:** Algunos problemas de compatibilidad aparecen con Office version 2007. Para evitar esos problemas, abra el archivo de Fenotipos y guárdelo con “grabar todos” en formato Excel 97-2003 antes de crear el archivo de entrada. Si esta etapa no se realiza, Office 2007 considera el archivo Fenotipo como un archivo txt en lugar de un archivo xls y AdFiT no podrá crear el archivo de entrada.

NOTA: Cuando usted abre el archivo Fenotipo por primera vez con Office 2007, un mensaje de error aparece, púntee sobre « Si » ☺

### Características Técnicas y prerrequisitos

<b>Nombre del Proyecto</b>	AdFiT (Admixture Files Tool)
<b>Versión</b>	1.0.5
<b>Lenguas</b>	Francés, Inglés, Español
<b>Sitio Web</b>	<a href="http://www.anthropologie-biologique.cnrs.fr/recherche/axe1equipe3.php">http://www.anthropologie-biologique.cnrs.fr/recherche/axe1equipe3.php</a>
<b>Sistema de Explotación</b>	Dependiente, necesita Windows (9X/XP/Vista)
<b>Otros prerrequisitos</b>	Microsoft© Office Libraries
<b>Licencia</b>	El ejecutable es de libre acceso para un uso personal (citando la publicación de referencia)

## Bibliografia

- Bertorelle G, Excoffier L. 1998. Inferring admixture proportions from molecular data. *Molecular Biology and Evolution* 15(10):1298-1311.
- Cabello PH, Krieger H. 1997. GENIOC: Sistema para análises de dados de genética. Rio de Janeiro: FIOCRUZ.
- Chakraborty R. 1985. Gene identity in racial hybrids and estimation of admixture rates. JV AYaN, editor. Delhi: Indian Anthropological Association.
- Chakraborty R, Srinivasan MR. 1992. A modified best-maximum likelihood estimator of line regression with errors in both variables - An application for estimating genetic admixture. *Biometrical Journal* 34(5):567-576.
- Chikhi L, Bruford MW, Beaumont MA. 2001. Estimation of admixture proportions: A likelihood-based approach using Markov chain Monte Carlo. *Genetics* 158(3):1347-1362.
- Dupanloup I, Bertorelle G. 2001. Inferring admixture proportions from molecular data: Extension to any number of parental populations. *Molecular Biology and Evolution* 18(4):672-675.
- Krieger H, Cabello P. 1997. *Mistura Racial*. Rio de Janeiro: PH Cabello & H Krieger.
- Krieger H, Morton NE, Mi MP, Azevedo E, Freirema.A, Yasuda N. 1965. Racial admixture in North-Eastern Brazil. *Annals of Human Genetics* 29:113-125.
- Langella O, Chikhi L, Beaumont MA. 2001. LEA (likelihood-based estimation of admixture): a program to estimate simultaneously admixture and time since the admixture event. *Molecular Ecology Notes* 1(4):357-358.
- Long JC. 1991. The Genetic Structure of Admixed Populations. *Genetics* 127(2):417-428.
- Roberts DF, Hiorns RW. 1965. Methods of analysis of genetic composition of a hybrid population. *Human Biology* 37(1):38-&.
- Wang JL. 2003. Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics* 164(2):747-765.