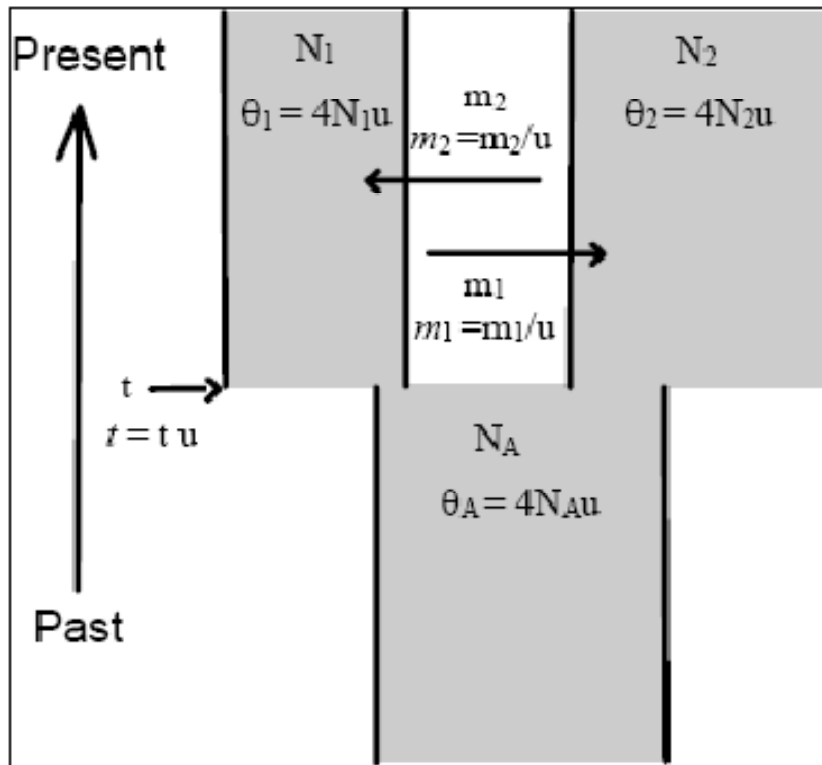


Estimativa de taxas de migração,
efectivos populacionais e tempo
de divergência

Modelo: “Isolation with migration model” (Hey & Nielsen 2004).



Este modelo inclui 6 parâmetros:

$$\theta_1 = 4N_1\mu$$

$$\theta_2 = 4N_2\mu$$

$$\theta_A = 4N_A\mu$$

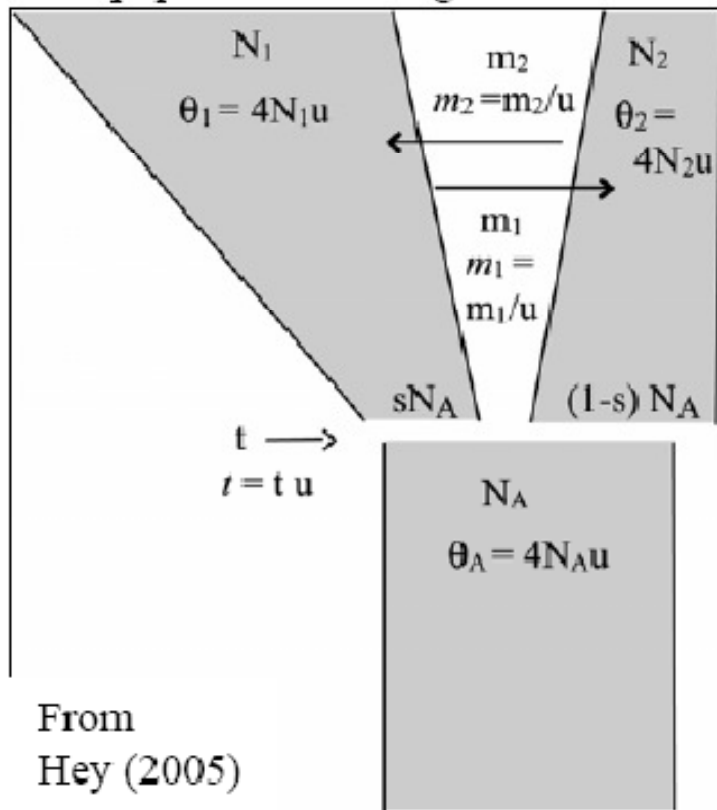
$$m_1 = m_1/\mu$$

$$m_2 = m_2/\mu$$

$$t = t u$$

Podemos estimar estes parâmetros usando os softwares IM e IMA.

A taxa de migração é estimada na mesma perspectiva da divergência, por isso conseguimos avaliar correctamente a causa para a partilha de polimorfismos.

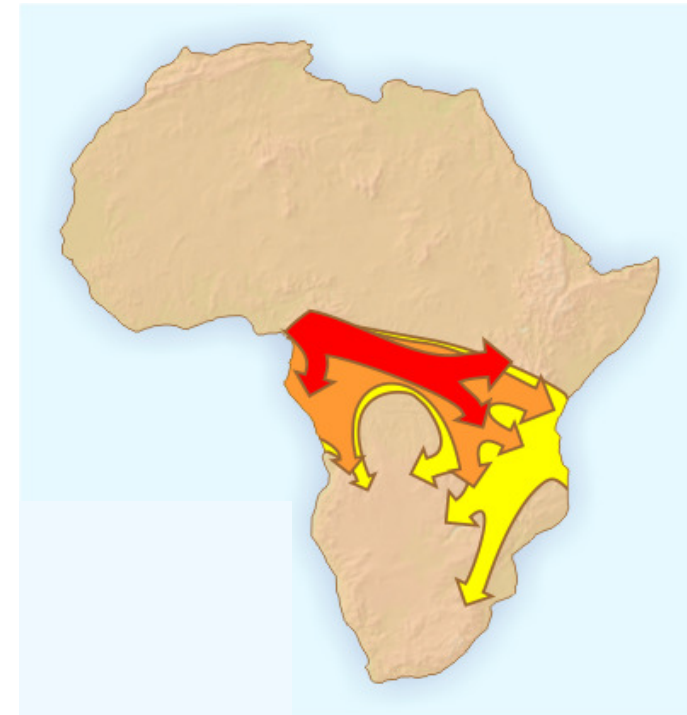


Este modelo foi aumentado para incorporar alterações no efectivo populacional.

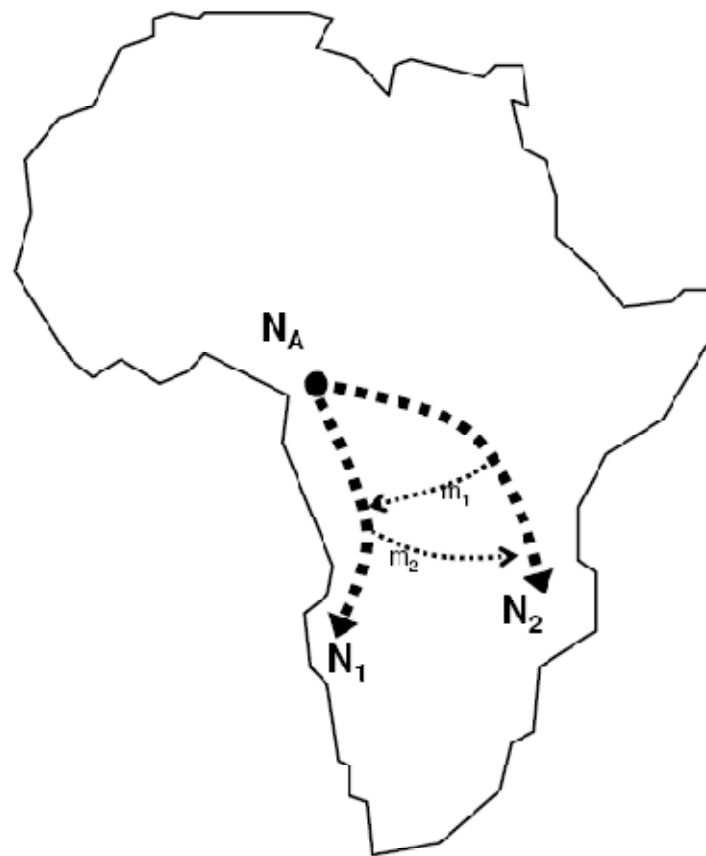
(mas nós vamos usar a vertente “clássica”)

EXPANSÃO BANTU

- Evidências arqueológicas e linguísticas sugerem que a expansão terá começado há cerca de 4000 mil anos a partir da região dos Camerões-Nigeria.
- Os primeiros movimentos terão incluído emigração para a região do Uganda e para as áreas costeiras húmidas das florestas tropicais da África Central (2500)
- Mais tarde os Bantu terão ultrapassado as fronteiras sul da floresta húmida e entrado em territórios mais secos, tendo permitido a disseminação das línguas Bantu até ao Sul de África (1500)



EXPANSÃO BANTU



Antes de usar o IMA temos de verificar se os dados preenchem uma série de requisitos (caso contrário não podíamos aplicar o modelo):

-ausência de recombinação intra-locus (OK)

-ausência de ligação entre os dois loci (OK)

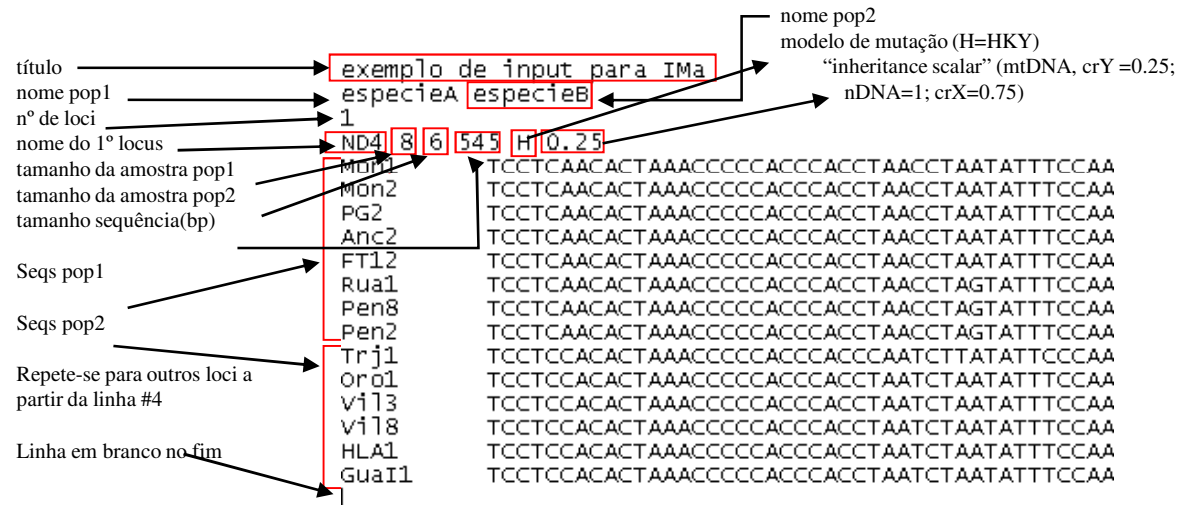
-ausência de selecção natural (OK)

-ausência de subestruturação dentro de cada população (vamos assumir que sim...)

-ausência de outras populações que possam interagir com as espécies em estudo e complicar o cenário (não, mas vamos fazer de conta que sim).

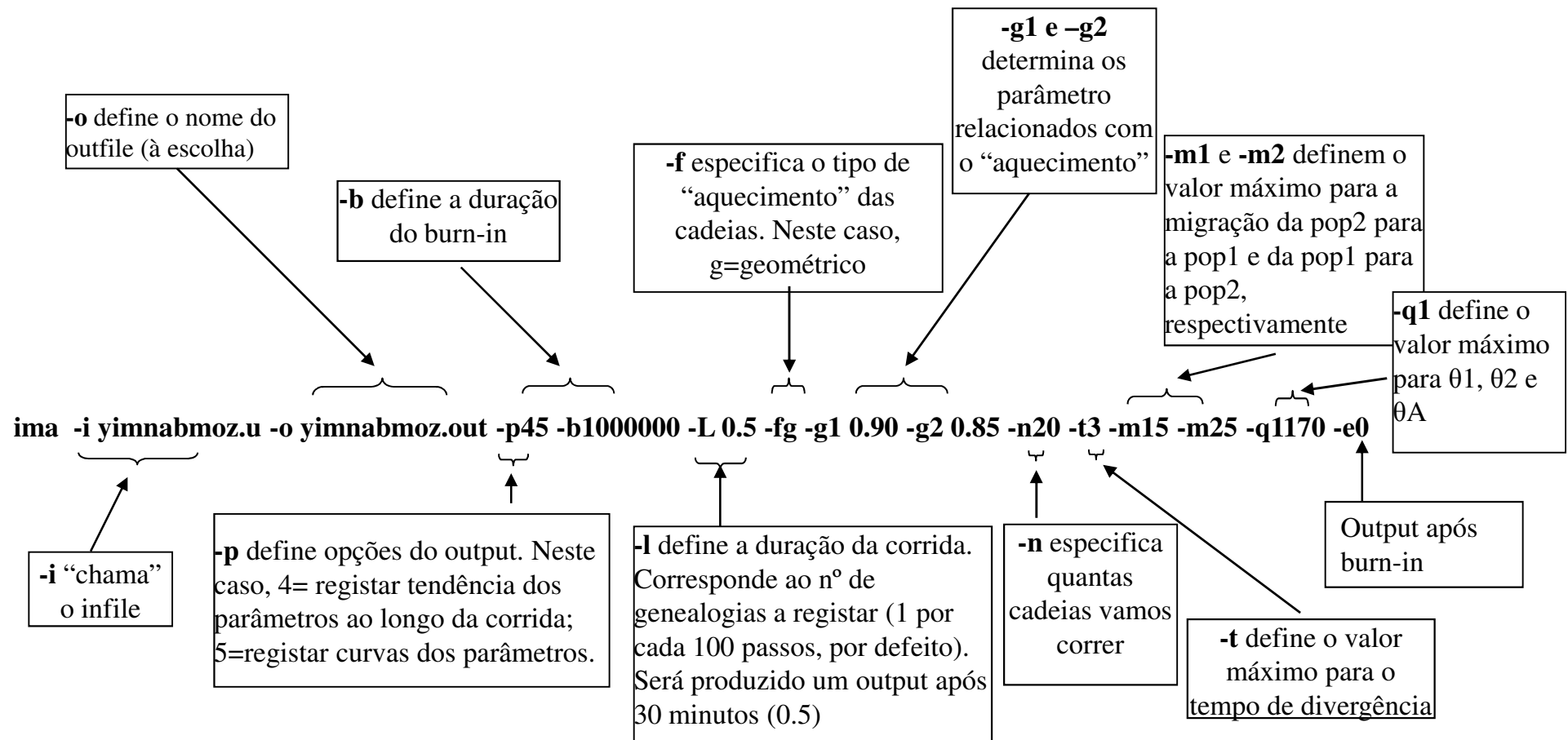
Preparação do input file para o IMA

Este deverá obedecer ao seguinte formato:



1. Abra a linha de comandos (**Menu Iniciar>Executar>cmd**)

2. O **IMa** funciona por linha de comandos. Normalmente, é necessário fazer experiências com as definições até optimizarmos os valores apropriados para a nossa corrida. Vamos usar directamente o seguinte comando (que foi testado anteriormente e já se sabe que resulta):



**ima -i yimnabmoz.u -o yimnabmoz.out
-p45 -b1000000 -L 0.5 -fg -g1 0.90 -g2
0.85 -n20 -t3 -m15 -m25 -q1170 -e0**

Antes de olhar para os resultados, é importante verificar se tudo correu bem. Por exemplo:

- **Os valores de ESS são todos >50?**
- **As estimativas dos parâmetros estabilizaram ao longo do tempo?**
- **As curvas dos parâmetros parecem-lhe suficientemente definidas para tirar conclusões?**

Autocorrelations and **Effective Sample Size Estimates**

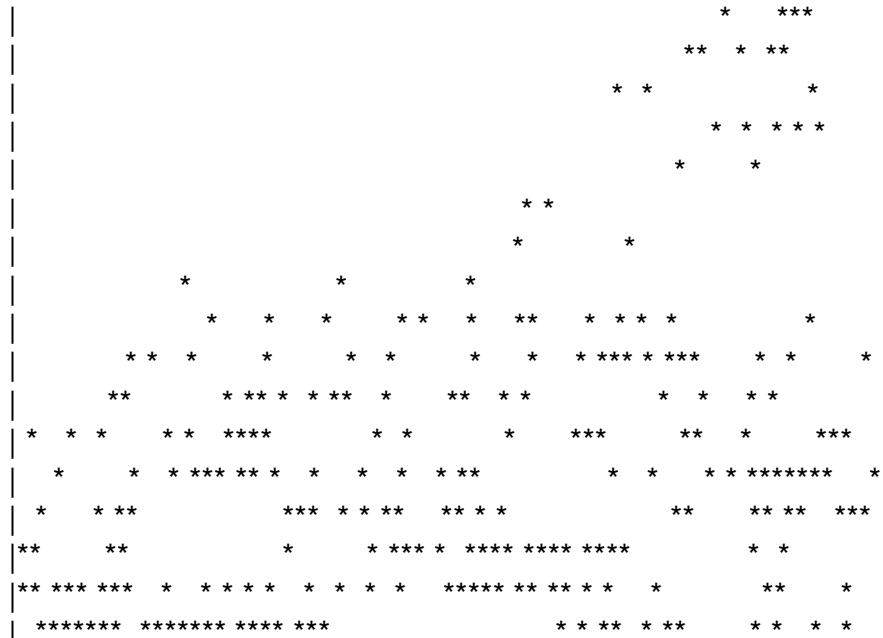
Steps Between Values and Autocorrelation Estimates

Steps	L[P]	t	tmrca0
1	0.3663	0.2558	0.3322
10	0.3034	0.1939	0.3161
50	0.2798	0.1929	0.2617
100	0.2776	0.1853	0.2831
500	0.2606	0.1784	0.2900
1000	0.2440	0.1762	0.2721
5000	0.2275	0.1600	0.2622
10000	0.2220	0.1639	0.2470
50000	0.1738	0.1213	0.0919
100000	0.1366	0.0947	0.0399
500000	0.0385	0.0050	-0.0126
1000000	0.0044	-0.0034	-0.0062
ESS	150	297	456

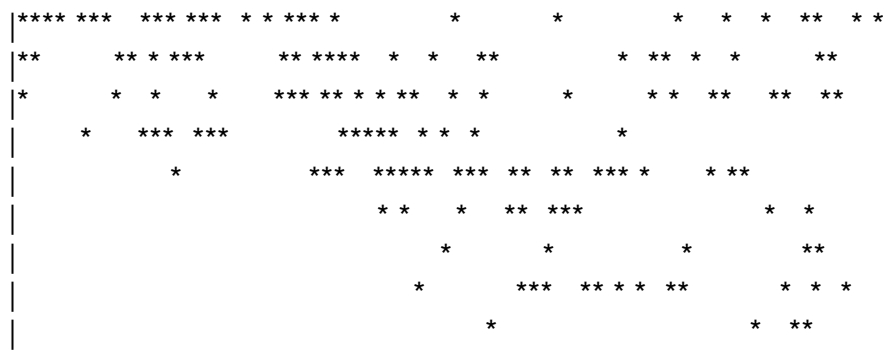
ASCII Plots of Parameter Trends

=====
Log[P(G)+P(D|G)] trend

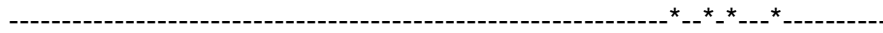
-951.3 |

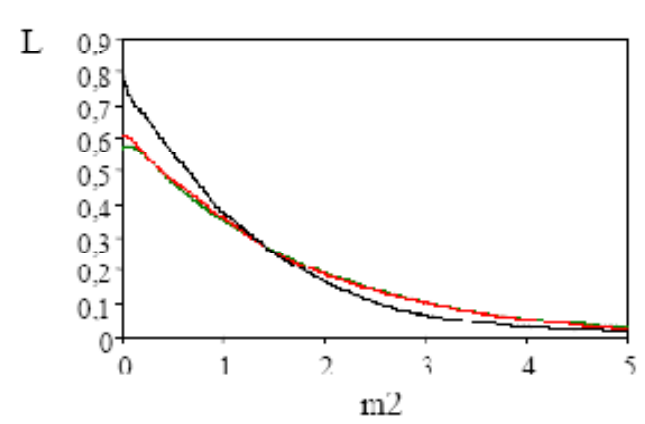
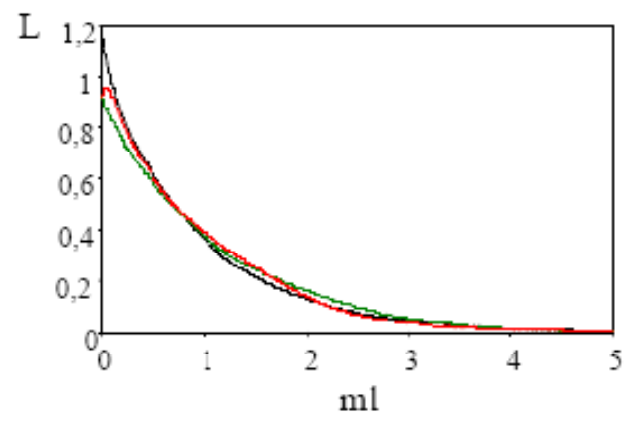
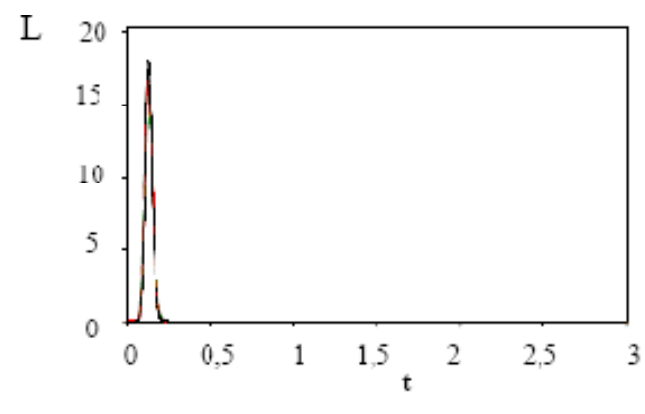
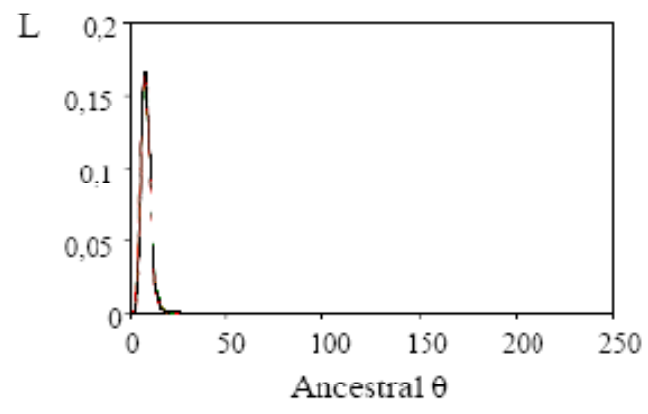
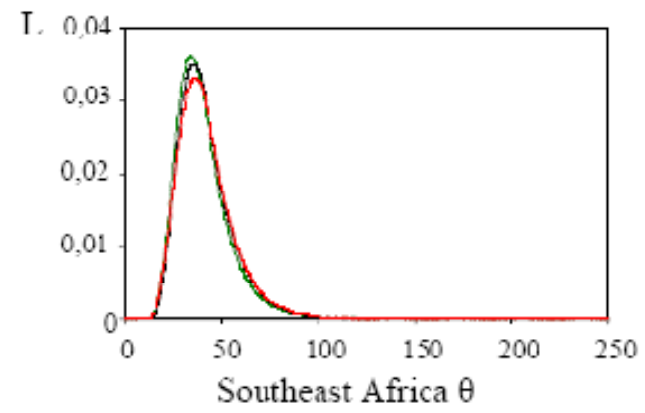
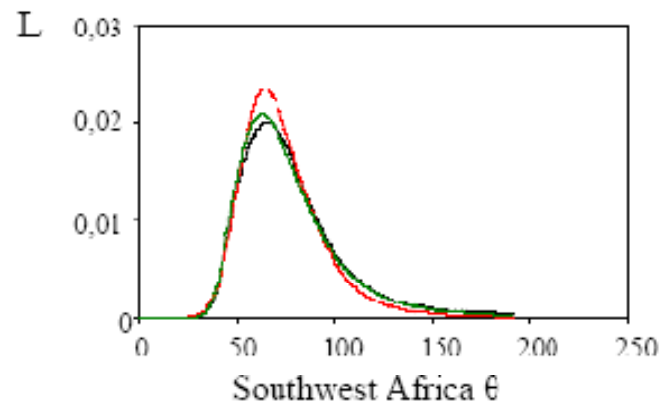


*



-5160.6366|





- Registrar o valor “HiPT” correspondente aos parâmetros que queremos estimar: θ_1 , θ_2 , θ_a , m_1 , m_2 e t . Registrar também os valores HPD90Lo e HPD90Hi, que correspondem ao limite inferior e superior, respectivamente, do intervalo contendo 90% da probabilidade posterior e que funcionam como intervalo de confiança.
- Os valores que foram estimados (θ_1 , θ_2 , θ_a , m_1 , m_2 e t) não correspondem exactamente aos parâmetros que nos interessam. Estão todos “escalados” pela taxa de mutação μ .
- Para fazermos os cálculos, é preciso usarmos um valor para a taxa de mutação (no caso de múltiplos loci, usa-se a média geométrica da taxa de mutação de cada um) por geração

Para estimar o efectivo populacional

$$\theta_1 = 4 * N_1 * \mu g$$

N_1 é o efectivo da população 1.

$$\text{Logo, } N_1 = \theta_1 / (4 * \mu g)$$

Procede-se de modo análogo para calcular N_2 e N_a .

- Tempo de geração no Homem = 25 anos
- $\mu_{DYS19} = 0.0017$; $\mu_{DYS389I} = 0.0019$; $\mu_{DYS389II} = 0.0023$;
- $\mu_{DYS390} = 0.0023$; $\mu_{DYS391} = 0.0035$; $\mu_{DYS392} = 0.0006$;
- $\mu_{DYS393} = 0.0007$.

$$U \text{ (média geométrica)} = 0.001585 = \mu g$$

Para estimar o tempo de divergência (em anos)

$t = t^* \mu a$, em que t é o tempo de divergência.

Logo, $t = t^* / \mu a$. (Nota: $\mu g = \mu a * \text{tempo de geração}$)

Para estimar as taxas de migração

$m1 = m1/\mu g$, em que $m1$ é a taxa de migração, por geração, da população 2 para a população 1.

Logo, $m1 = m1 * \mu g$. Forma análoga para $m2$.

Para estimar as taxas de migração populacional

(número efectivo de genes que migra para uma população, por geração)

$$2.N1.m1 = 2 * \theta 1 / (4 * \mu g) * m1 * \mu g$$

$$= (\theta 1 * m1) / 2$$

Table 4: Estimates of demographic parameters in the Southwest and Southeast edges of the Bantu expansions

	N_1	N_2	N_A	m_1	m_2	$2N_1m_1$	$2N_2m_2$	t (years)
Y chrom	10020 (6684–21557)	5510 (3313–12057)	1195 (647–2372)	0 ($0-4 \times 10^{-3}$)	0 ($0-6 \times 10^{-3}$)	0 (0–80.2)	0 (0–66.1)	1950 (1388–2940)
mtDNA								
A	33212 (22680–44905)	31885 (23592–46315)	7090 (3441–16129)	2.6×10^{-4} ($9.0 \times 10^{-5}-4.5 \times 10^{-3}$)	5.5×10^{-4} ($2.8 \times 10^{-4}-4.7 \times 10^{-3}$)	17 (6–297)	35 (18–298)	25410 (14612–39135)
B	35700 (24256–46066)	30558 (23055–46315)	7173 (3607–17041)	2.4×10^{-4} ($5.0 \times 10^{-5}-3.9 \times 10^{-3}$)	3.0×10^{-3} ($4.5 \times 10^{-4}-4.8 \times 10^{-3}$)	17 (4–275)	184 (21–296)	24978 (13749–38319)
Joint								
A	6894 (4631–9571)	6150 (4403–8619)	1394 (343–2246)	2×10^{-3} ($4.7 \times 10^{-4}-1.6 \times 10^{-2}$)	7×10^{-3} ($1.4 \times 10^{-3}-1.6 \times 10^{-2}$)	28 (6–215)	86 (17–191)	4133 (3071–13384)
B	7558 (5417–9929)	6274 (4884–8578)	1372 (933–2433)	1×10^{-3} ($4.5 \times 10^{-4}-1.1 \times 10^{-2}$)	9×10^{-3} ($4.5 \times 10^{-4}-1.5 \times 10^{-2}$)	15 (7–211)	113 (6–185)	3981 (3147–6332)

N_1 -Current effective population size in the Southwest edge; N_2 -Current population size in the Southeast edge; N_A -Ancestral effective population size; m_1 -Probability of migration from Southeast to Southwest Africa, per gene copy per generation; m_2 -Probability of migration from Southwest to Southeast Africa, per gene copy per generation; $2N_1m_1$ -Effective number of genes migrating into Southwest Africa, per generation; $2N_2m_2$ -Effective number of genes migrating into Southeast Africa, per generation; t-time since divergence from a common ancestor. 95% credibility intervals are given in parentheses. A and B show the outcome of runs with different probability density peaks for m_2 .