A new composite likelihood method to characterize demographic expansions: preliminary results from Y-chromosome STR data

Miguel Navascués, Concetta Burgarella

To cite this version:

Miguel Navascués, Concetta Burgarella. A new composite likelihood method to characterize demographic expansions: preliminary results from Y-chromosome STR data. DNA in Forensics, Università Politecnica delle Marche, May 2008, Ancona, Italy. hal-02755469

HAL Id: hal-02755469
https://hal.inrae.fr/hal-02755469v1

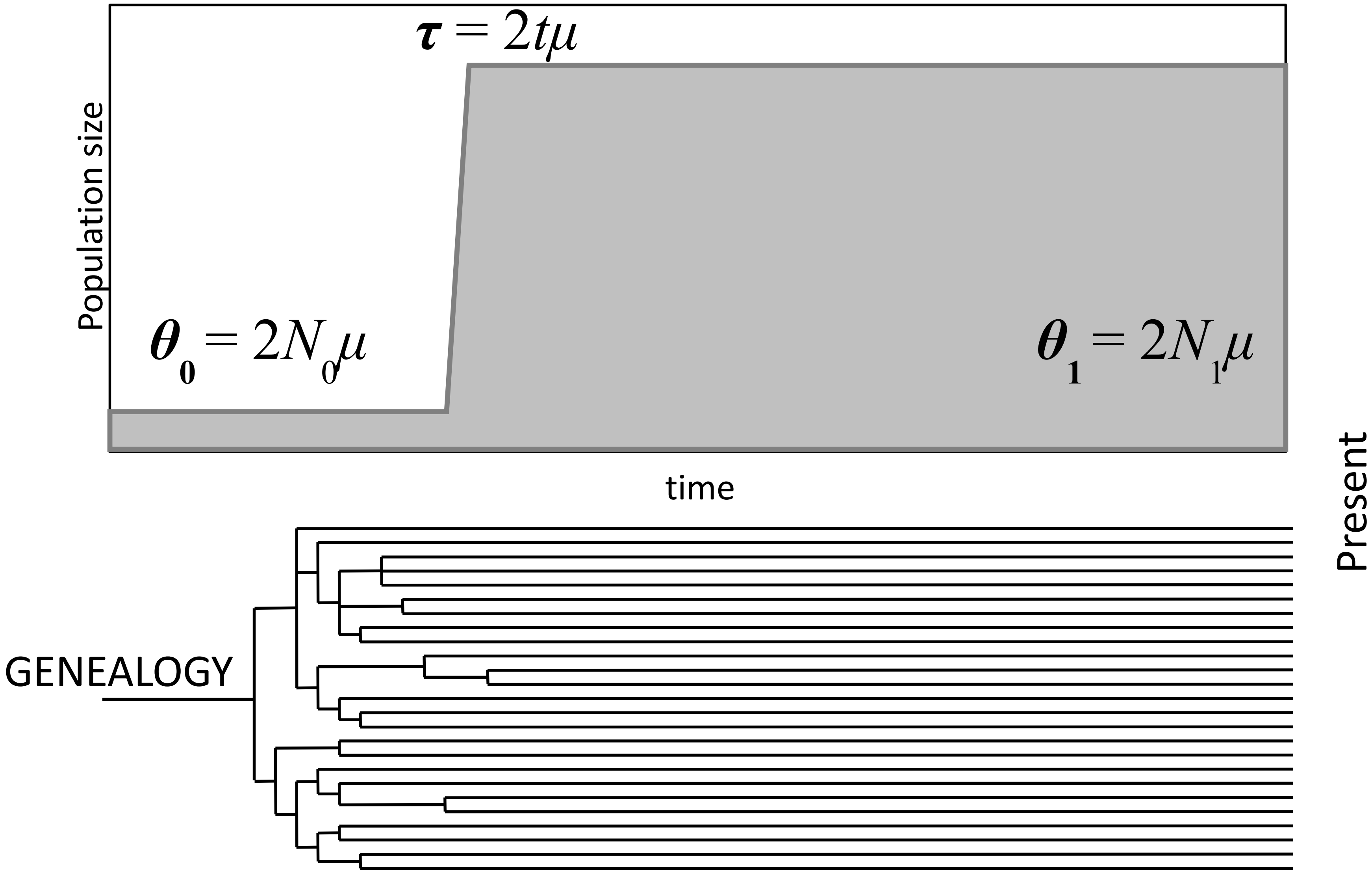
Submitted on 1 Mar 2021

A new composite likelihood method to characterize demographic expansions

Miguel Navascués & Concetta Burgarella

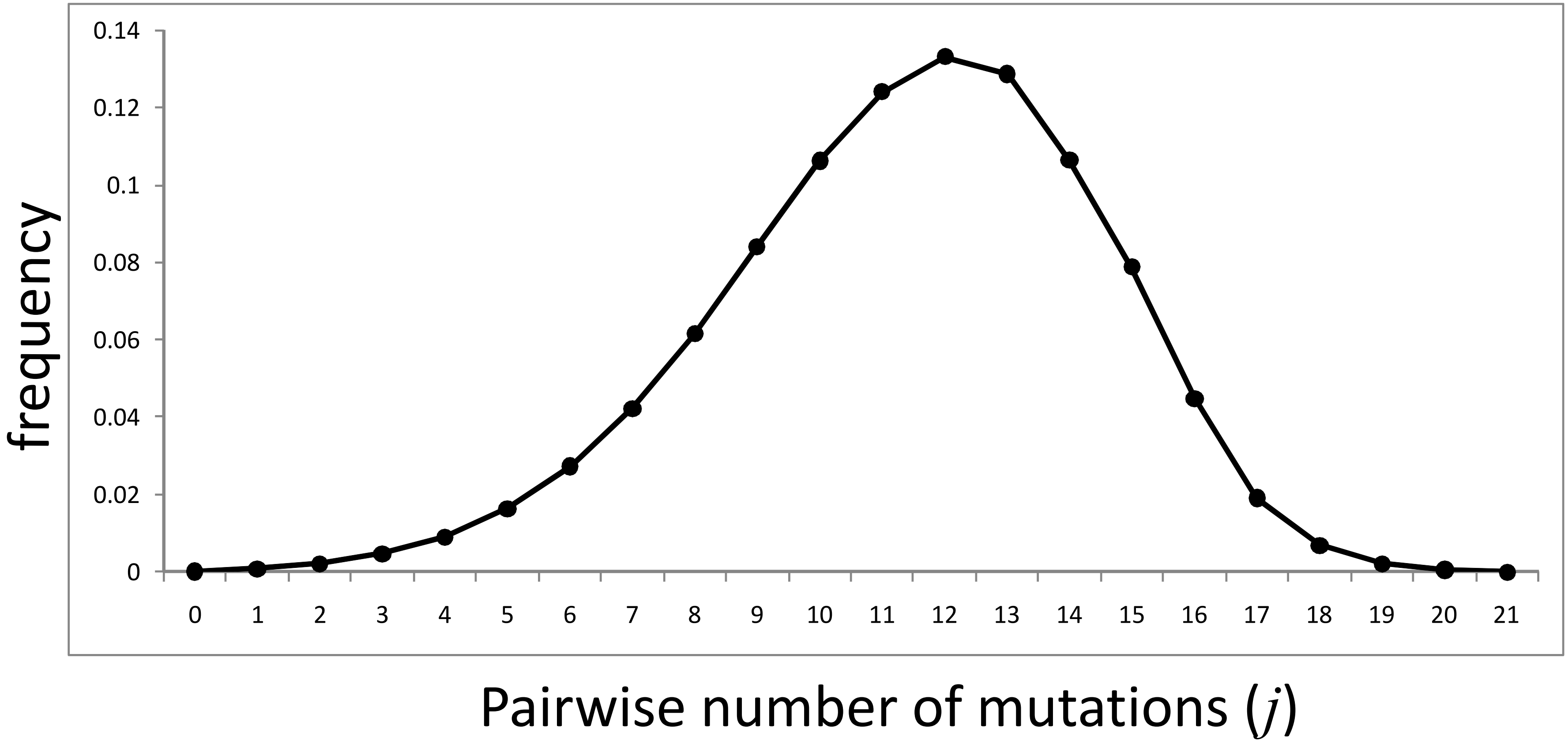
CNRS UMR 7625 Écologie et Évolution
Université Pierre et Marie Curie / École Normale Supérieure

Demographic expansion



Mutation distribution

$$P(j;\theta_0,\theta_1,\tau)=P(j;\theta_1)+\mathrm{e}^{-\tau\frac{\theta_1+1}{\theta_1}}\times\sum_{j'=0}^{j}\frac{\tau^{j'}}{j'!}\left[P(j-j';\theta_0)-P(j-j';\theta_1)\right]$$



Li (1977) *Genetics // *Slatkin & Hudson (1991) *Genetics*

Mutations (j) ≥ Genetic differences (i)

$$P(i;\theta_0,\theta_1,\tau)=\sum_{j=i}^{\infty}P(j;\theta_0,\theta_1,\tau)\times P(i;j)$$

Schneider & Excoffier (1999) *Genetics*

Mutations (j) ≥ Genetic differences (i)

$$P(i;\theta_0,\theta_1,\tau)=\sum_{j=i}^{\infty}P(j;\theta_0,\theta_1,\tau)\times P(i;j)$$

Distribution of genetic differences

Schneider & Excoffier (1999) *Genetics*

Mutations (j) ≥ Genetic differences (i)

Distribution of mutations

$$P(i;\theta_0,\theta_1,\tau)=\sum_{j=i}^{\infty}P(j;\theta_0,\theta_1,\tau)\times P(i;j)$$

Distribution of genetic differences

Schneider & Excoffier (1999) *Genetics*

Mutations (j) ≥ Genetic differences (i)

Distribution of mutations

$$P(i;\theta_0,\theta_1,\tau)=\sum_{j=i}^{\infty}P(j;\theta_0,\theta_1,\tau)\times P(i;j)$$

Distribution of genetic differences

Mutation model

Schneider & Excoffier (1999) *Genetics*

Mutational model, $P(i;j)$

Stepwise Mutation Model

$$P(\delta;j)=\begin{cases}\binom{j}{\frac{j}{2}}\left(\frac{1}{2}\right)^{j} & \delta=0\wedge j\;even\\ \binom{j}{\frac{j+2}{2}}\left(\frac{1}{2}\right)^{j}+\binom{j}{\frac{j-2}{2}}\left(\frac{1}{2}\right)^{j} & \delta\neq0\wedge j\;even\\ 0 & otherwise\end{cases}$$

Linked loci with different mutation rates

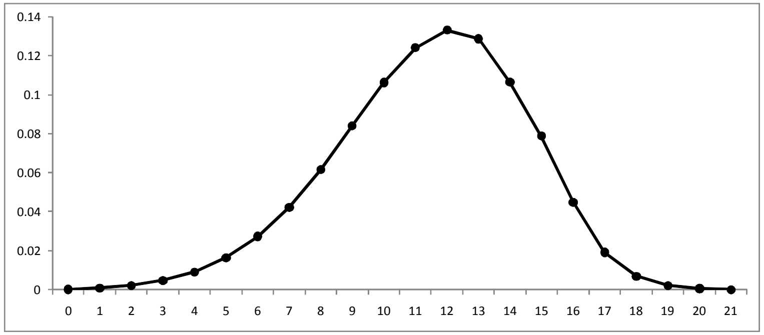
$$P(\{k_{1,}k_{2,}\ldots,k_L\};j)=\begin{cases}\frac{j!}{k_1!k_2!\ldots k_L!}p_1^2p_2^2\ldots p_L^2 & \sum_{l=1}^{L}k_l=j\\ 0 & otherwise\end{cases}$$

Estimation of time of expansion

1000 coalescent simulations of demographic expansions

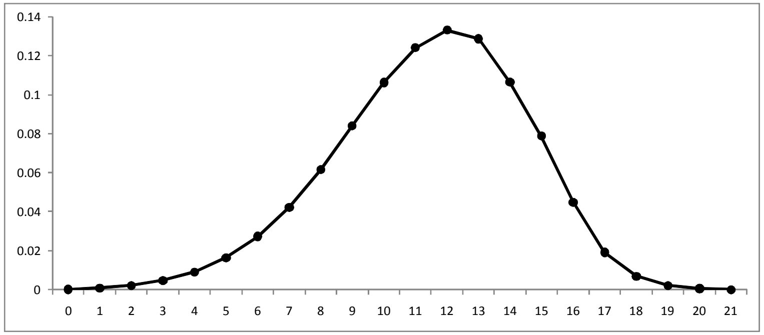
Estimation of time of expansion

1000 coalescent simulations of demographic expansions



Estimation of time of expansion

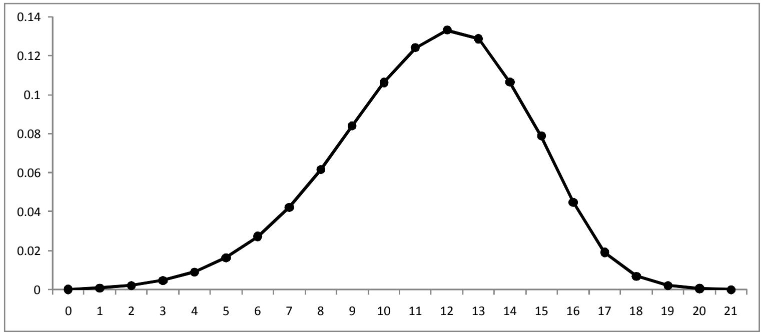
1000 coalescent simulations of demographic expansions



$$P(mutation\,;\theta_0,\theta_1,\tau)$$

Estimation of time of expansion

1000 coalescent simulations of demographic expansions

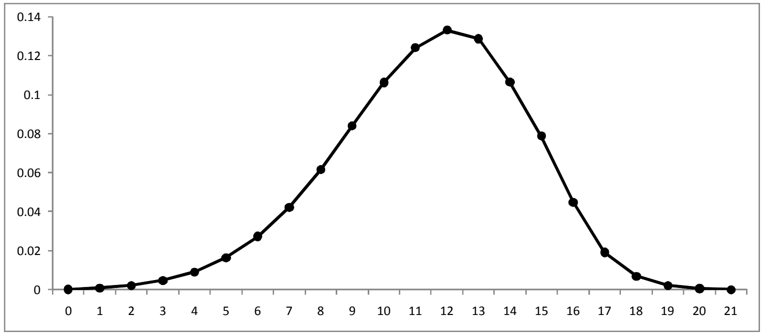


$$\boxed{P(mutation\,;\theta_0,\theta_1,\tau)}$$

$$\boxed{P(genetic\ differences\,;\theta_0,\theta_1,\tau)}$$

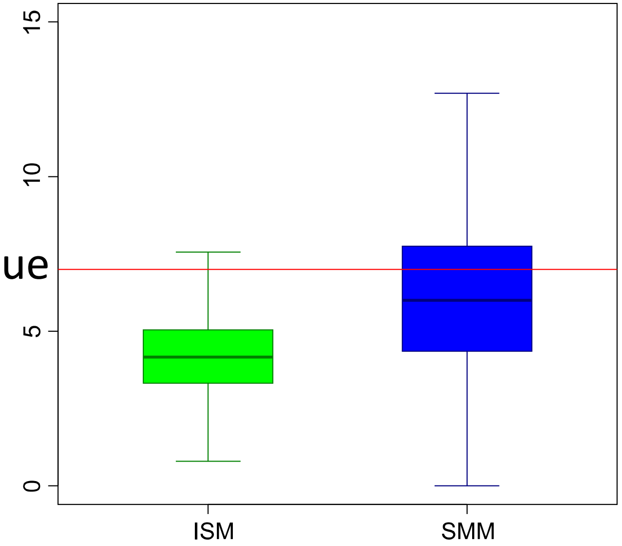
Estimation of time of expansion

1000 coalescent simulations of demographic expansions



$$P(mutation\,;\theta_0,\theta_1,\tau)$$

$$P(genetic\ differences\,;\theta_0,\theta_1,\tau)$$

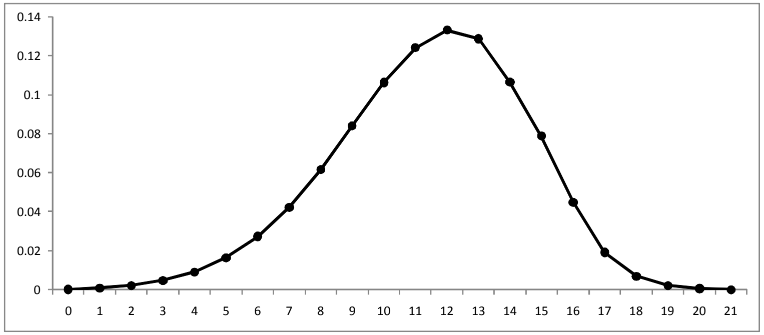


True (simulated) value

1000 estimates of the time of expansion

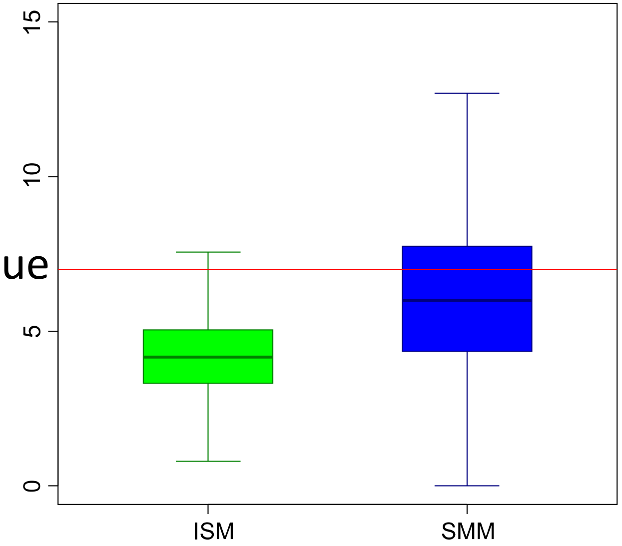
Estimation of time of expansion

1000 coalescent simulations of demographic expansions



$$P(mutation\,;\theta_0,\theta_1,\tau)$$

$$P(genetic\ differences\,;\theta_0,\theta_1,\tau)$$



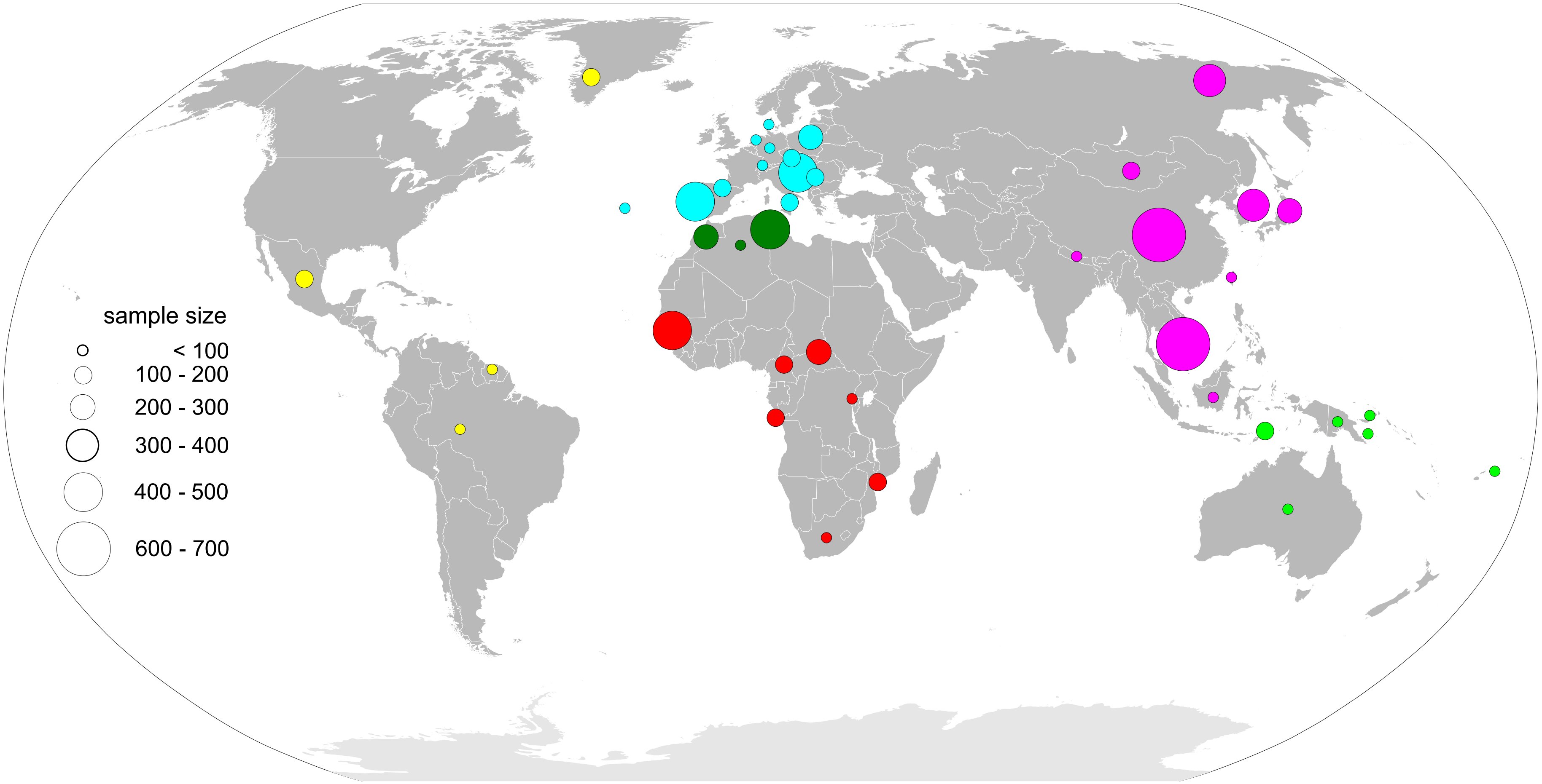
True (simulated) value

1000 estimates of the time of expansion

Using a mutational model yields less biased estimates

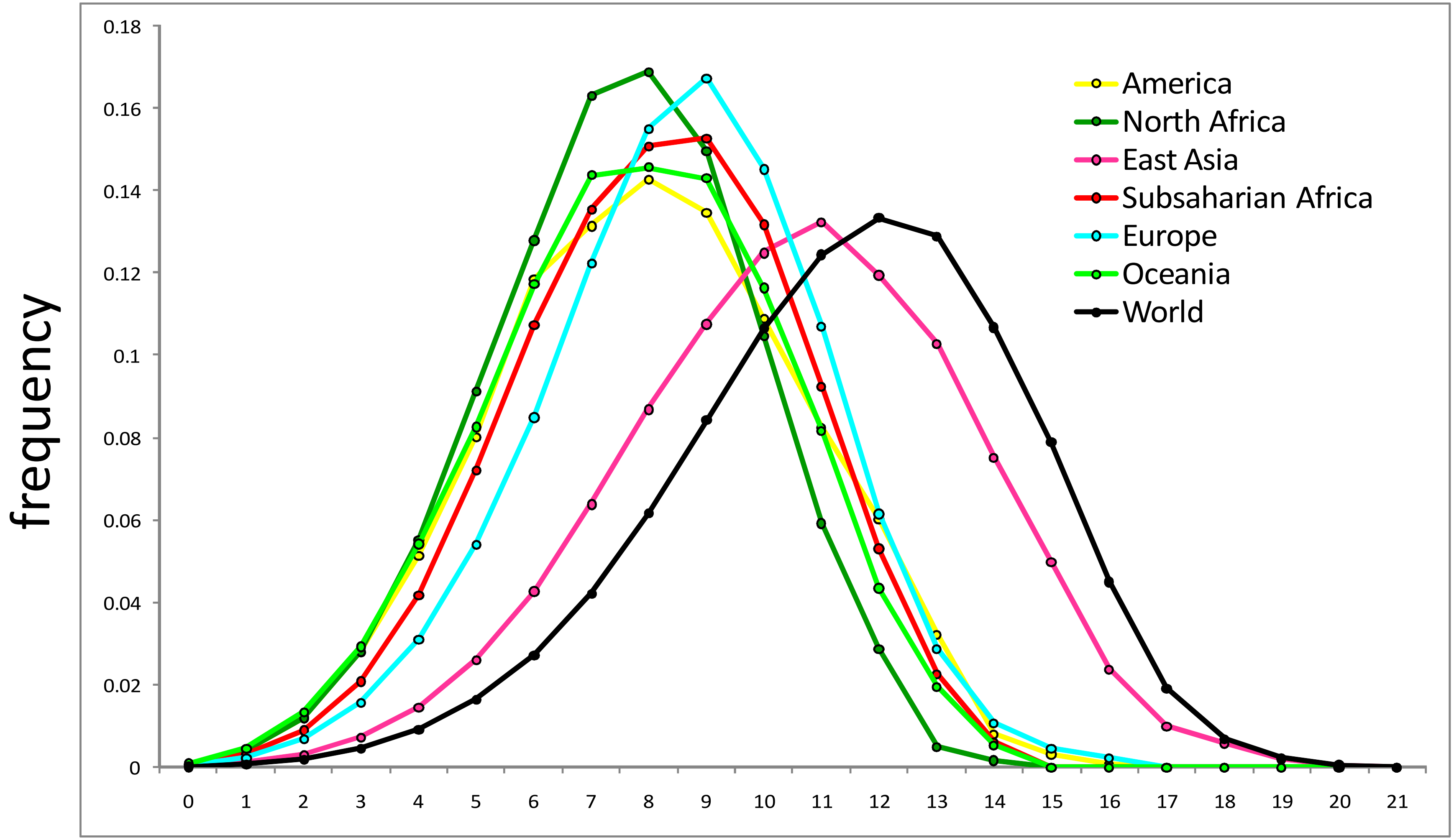
Y-STRs population samples

DYS19, DYS390, DYS391, DYS392 and DYS393



groups following Li *et al*. (2008) *Science*

Genetic distance distribution



Pairwise difference in number of repeats

Time of expansion estimates

Group	# individuals	# populations	$\hat{\tau}$ (95%CI)	$\hat{t}$ years (95%CI)[b]
Sub-Saharan-Africa	1183	23	12.53 (1.29-26.18)	17612 (2672-36811)
N-Africa	694	13	13.69 (0.10-24.51)	19247 (441-34463)
Europe	2150	14	9.40 (5.00-12.81)	13213 (7030-18012)
East-Asia	2621	16	11.51 (8.18-15.75)	16182 (11502-22146)
Oceania	293	7	12.04 (6.76-17.45)	16924 (9505-23552)
America	422	8	8.70 (5.10-15.42)	12231 (7171-21682)
World[a]	9182	94	12.85 (9.25-14.41)	18061 (13006-20262)

[a] Includes samples not classified in previous groups

[b] mutation rate per generation over the five loci, $\mu = 0.89 \times 10^{-2}$ (www.ystr.org); generation time: 25 years

Time of expansion estimates

Group	# individuals	# populations	$\hat{\tau}$ (95%CI)	$\hat{t}$ years (95%CI)^{b}
Sub-Saharan-Africa	1183	23	12.53 (1.29-26.18)	17612 (2672-36811)
N-Africa	694	13	13.69 (0.10-24.51)	19247 (441-34463)
Europe	2150	14	9.40 (5.00-12.81)	13213 (7030-18012)
East-Asia	2621	16	11.51 (8.18-15.75)	16182 (11502-22146)
Oceania	293	7	12.04 (6.76-17.45)	16924 (9505-23552)
America	422	8	8.70 (5.10-15.42)	12231 (7171-21682)
World[a]	9182	94	12.85 (9.25-14.41)	18061 (13006-20262)

[a] Includes samples not classified in previous groups

[b] mutation rate per generation over the five loci, $\mu = 0.89 \text{ x}10^{-2}$ (www.ystr.org); generation time: 25 years

Approximate Bayesian Computation:
Pritchard *et al.* (1999) *Mol Biol Evol* (sample size 445) t=18,000 95%HPD (7,000-41,000)

Bayesian Coalescent MCMC:
Wilson *et al.* (2003) *J Roy Stat Soc A* (sample size 115) t=14,000 90%HPD (6,800-28,000)

Conclusion

- Analytical description of mutational process for linked microsatellites
- This model can applied in statistical analysis:
  - **Estimation of demographic expansion**
  - *Neutrality test*
  - *Identity by descent/identity in state probabilities*
- Example data set: estimates comparable to those of previous studies (using other statistics and data)

Thanks

Olivier Hardy (Universite Libre de Bruxelles)
Frantz Depaulis (CNRS)

Funding: EUROPEAN SCIENCE FOUNDATION MAIRIE DE PARIS

Data:

Alves et al. 2003. *Forensic Sci. Int.* **134**(2-3): 126-133.
Alves et al. 2007. *Forensic Sci. Int.* **171**(2-3): 250-255.
Ayadi et al. 2006. *Forensic Sci. Int.* **164**(2-3): 249-253.
Barac, et al. 2003. *Forensic Sci. Int.* **138**(1-3): 127-133.
Barrot et al. 2007. *Forensic Sci. Int.* **168**(1), e10-e12.
Berger et al. 2003. *Forensic Sci. Int.* **137**(2-3): 221-230.
Bosch et al. 2000. *Int. J. Legal Med.* **114**(1): 36-40.
Bosch et al. 2003. *Forensic Sci. Int.* **132**(3): 228-232.
Brandt-Casadevall et al. 2003. *Forensic Sci. Int.* **135**(3): 247-250.
Cagliá et al. 2003. *Hum. Biol.* **75**(3): 313-330.
Carvalho et al. 2003. *Forensic Sci. Int.* **134**(1): 29-35.
Chang et al. 2007. *Forensic Sci. Int.* **167**(1): 70-76.
Cherni et al. 2005. *Forensic Sci. Int.* **152**(1): 95-99.
Coia et al. 2004. *Am. J. Hum. Genet.* **16**(1): 57-67.
Destro Bisol et al. 2004. *Mol. Biol. Evol.* **21**(9): 1673-1682.
Frigi et al. 2006. *Forensic Sci. Int.* **160**(1): 80-83.
Foster et al. 1998. *Mol. Biol. Evol.* **15**(9): 1108-1114.
Garcia et al. 2004. *Forensic Sci. Int.* **145**(1): 65-68.
Hu 2006. *Forensic Sci. Int.* **158**(1): 80-85.
Kayser et al. 2001. *Am. J. Hum. Genet.* **68**(4): 990-1018.
Khodjet el Khil et al. 2005. *Forensic Sci. Int.* **148**(2-3): 211-218.
Kumagai et al. 2007. *Forensic Sci. Int.* **172**(1): 72-78.
Leat et al. 2007. *Forensic Sci. Int.* **168**(2-3): 154-161.
Lecerf et al. 2007. *Forensic Sci. Int.* **171**(2-3): 212-215.
Lessig et al. 2006. *Forensic Sci. Int.* **159**(1): 71-76.
Li et al. 2007. *Forensic Sci. Int.* **172**(1): 79-83.
Park et al. 2005. *Forensic Sci. Int.* **152**(2-3): 133-147.
Pereira et al. 2002. *Ann. Hum. Genet*. **66**: 369-378
Quintana-Murci et al. 2004. *Forensic Sci. Int.* **140**(1): 113-115.
Rosa et al. 2007. *BMC Evolutionary Biology* **7**(1): 124.
Soltyszewski et al. 2007. *Forensic Sci. Int.* **168:** 61-67.
Souto et al. 2006. *Forensic Sci. Int.* **156**: 261-265.
Trovoada et al. 2001. *Ann. Hum. Genet.* **65**(3): 271-283.
Veselinovic et al. 2008. *Forensic Sci. Int.* **176**(2-3): 23-28.
Woźniak et al. 2006. *Forensic Sci. Int.* **164**: 271-275.
Xin et al. 2008. *Forensic Sci. Int.* **174**(2-3): 244-248.
Zhang et al. 2008. *Forensic Sci. Int.* **175**(2-3): 244-249.
Zhu et al. 2005. *Forensic Sci. Int.* **153**: 260-263.