Neolithic spread dynamics: archaeology, genetics and linguistics

Joaquim Fort
Universitat de Girona
(Catalonia, Spain)
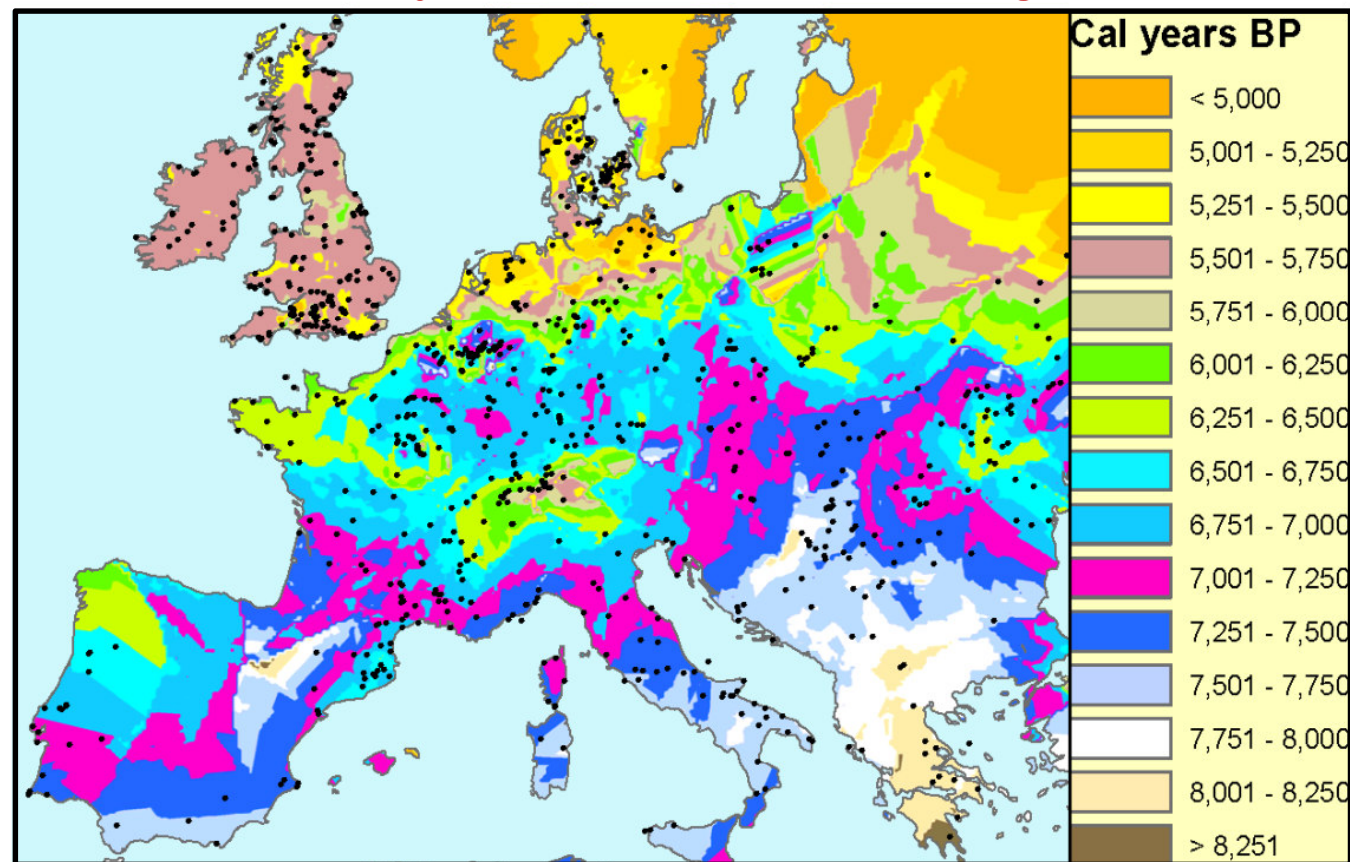
Fudan University
Shanghai
May 8th, 2024

MINISTERIO DE CIENCIA E INNOVACIÓN

AGENCIA ESTATAL DE INVESTIGACIÓN

ICREA
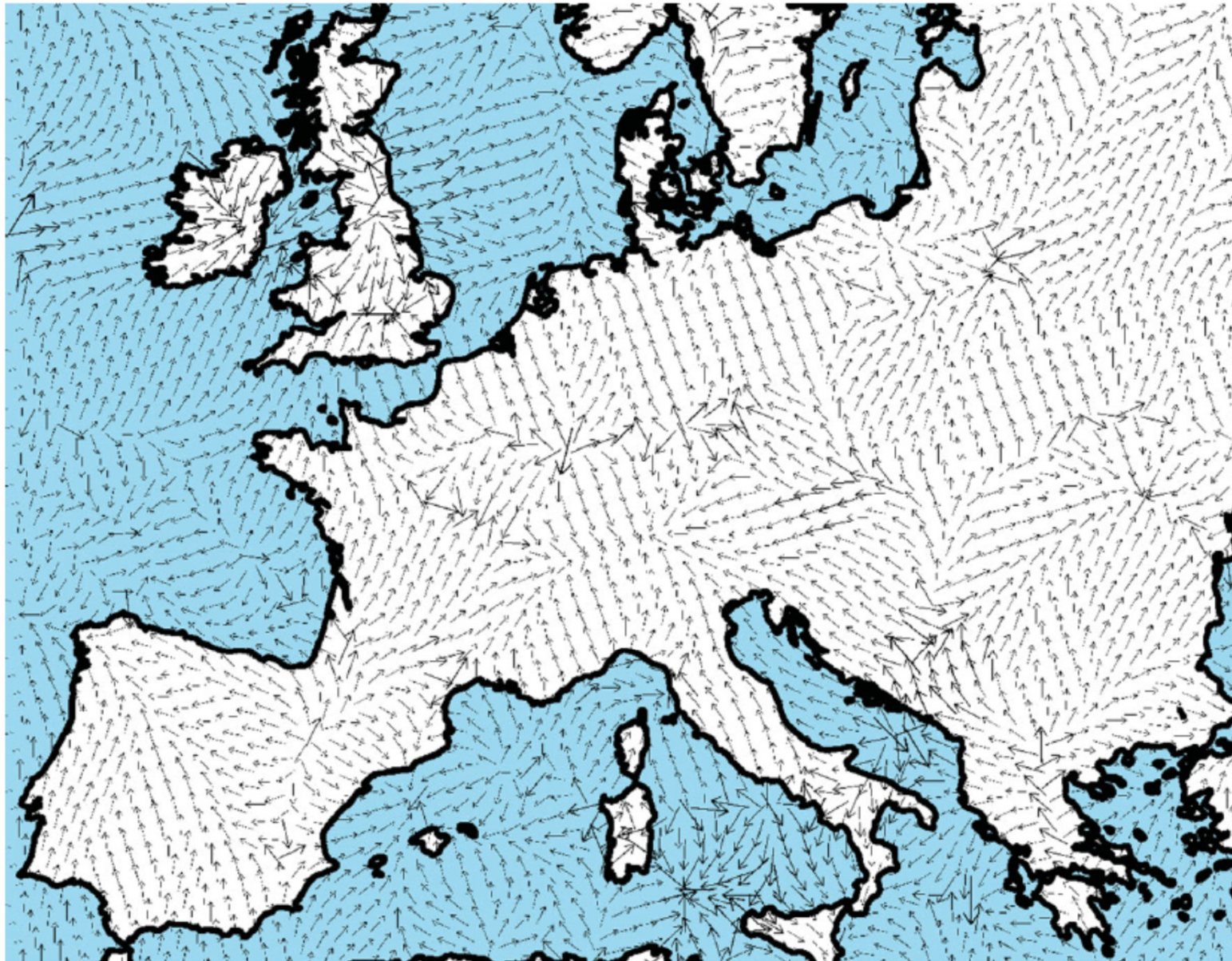
Universitat de Girona

1

Neolithic transition =
hunting-gathering →
farming and stockbreeding

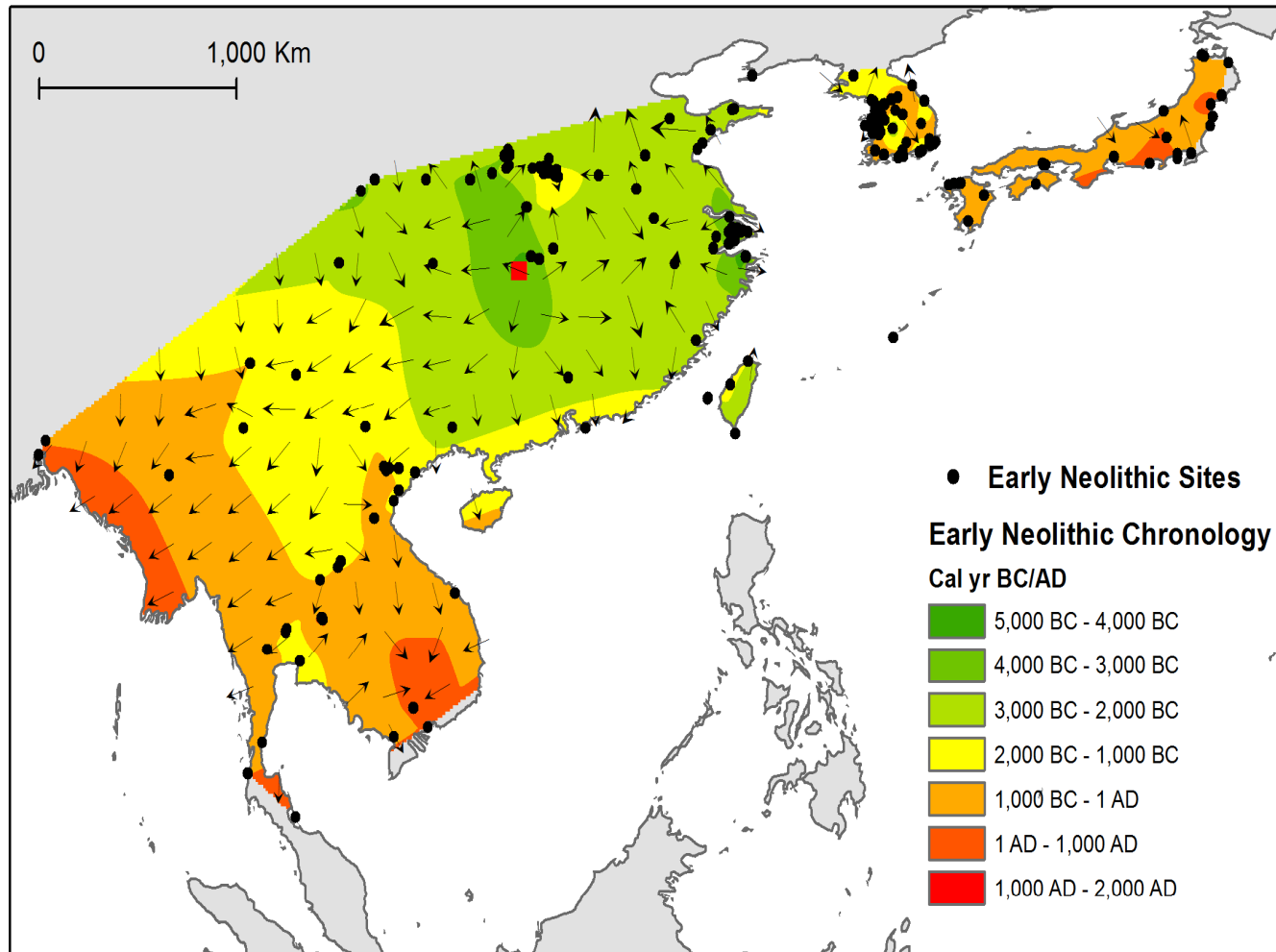Europe, early Neolithic archaeological sites:

Fort,
J. Roy.
Soc.
Interface
(2015)

Archaeology

Fort,
J. Roy.
Soc.
Interface
(2015)
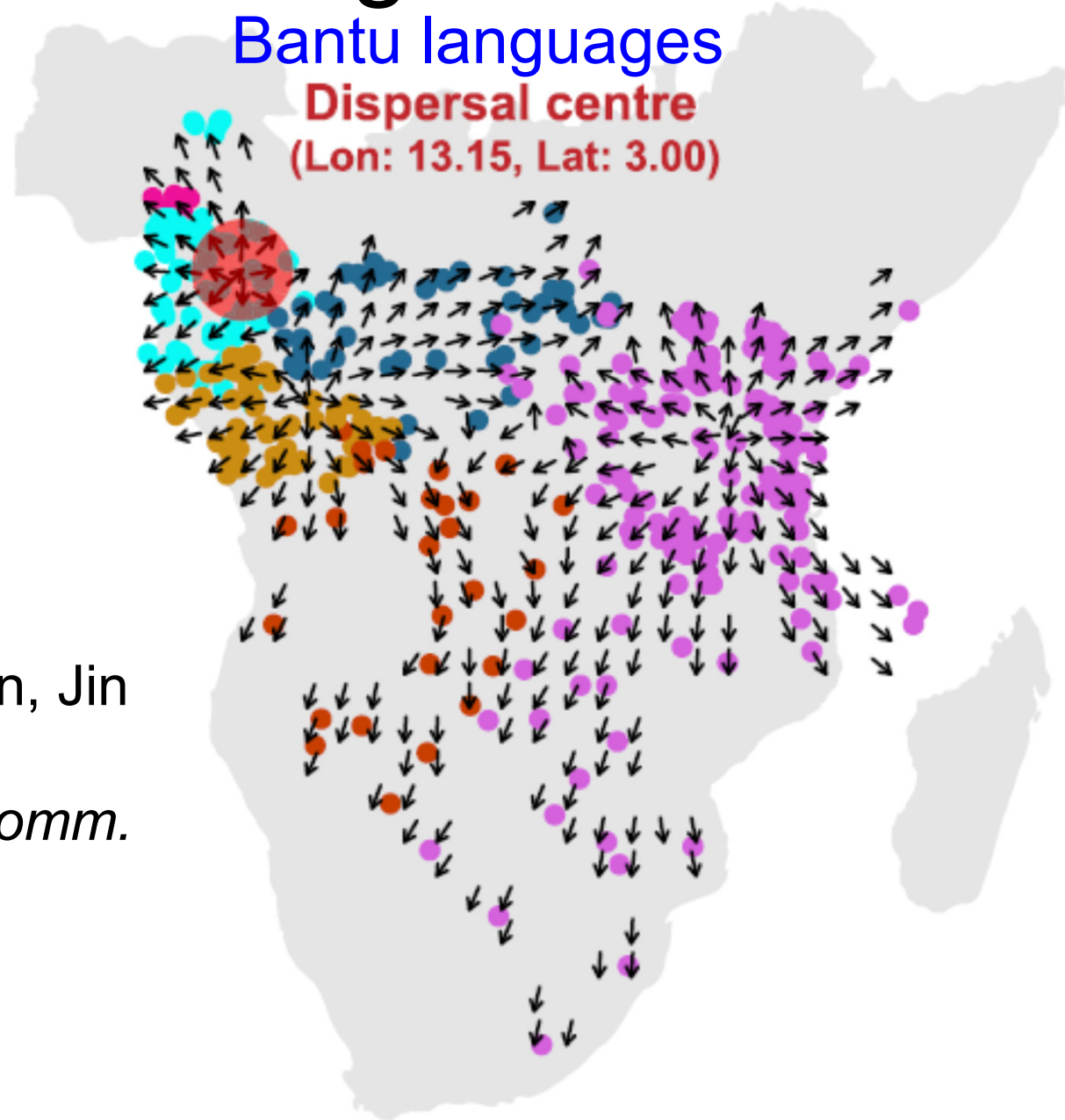
3

The spread of domesticated rice

Cobo, Fort & Isern, *J. Arch. Sci.* (2019)

4

Linguistics
Bantu languages

Dispersal centre
(Lon: 13.15, Lat: 3.00)

Yang, Sun, Jin
& Zhang,
Nature Comm.
(2024)

Archaeology

Bantu spread of farming and pastoralism
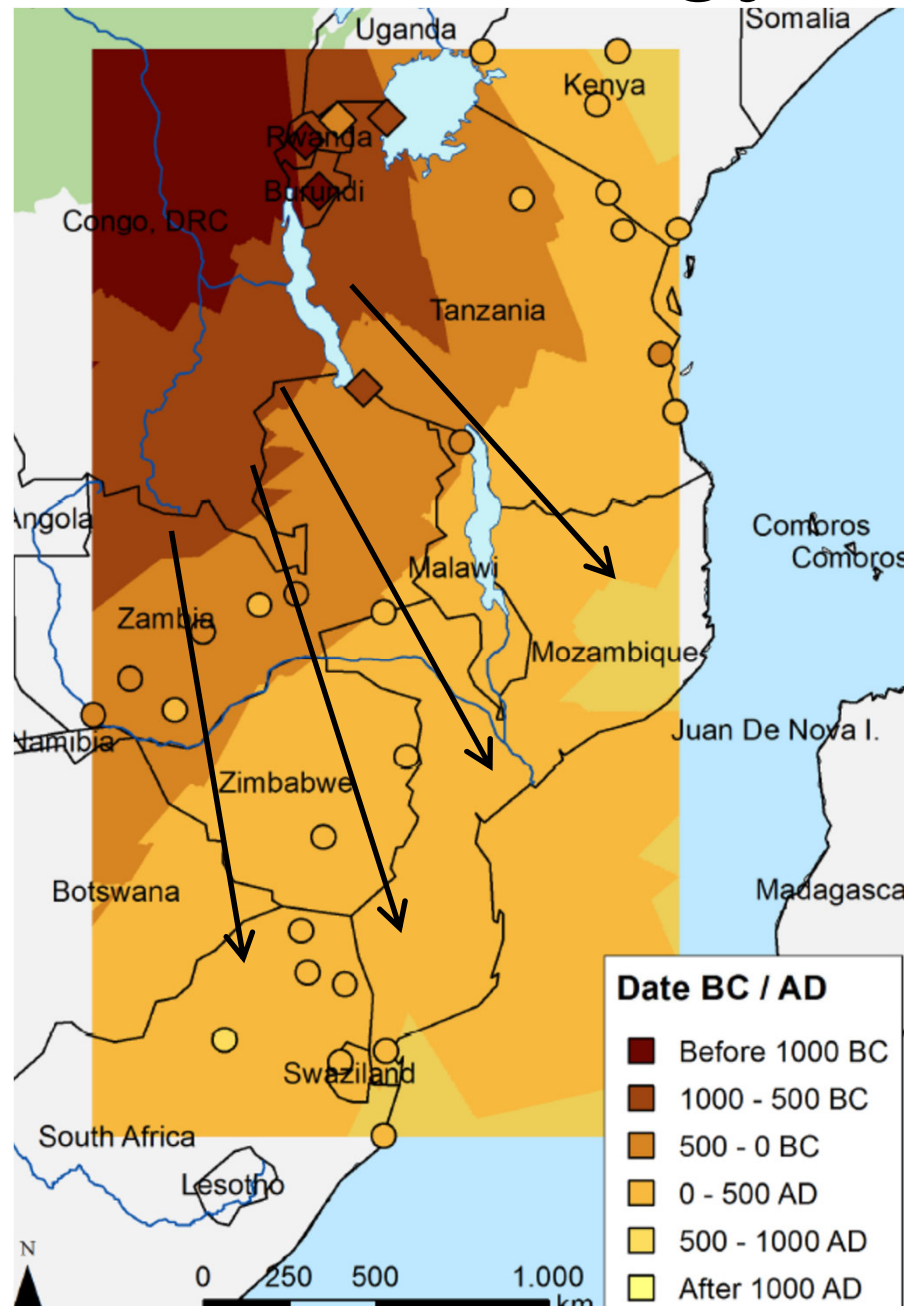
Isern & Fort, *PLoS One* (2019)

6

Models of the spread of farming

- Demic diffusion = Farming populations spread = dispersal + net reproduction

- Cultural diffusion = spread of ideas = transmission of plants, animals and knowledge from farmers to hunter-gatherers (acculturation).

- Demic-cultural models

Cultural transmission

Population numbers after *(P')* and before *(P)* cultural transmission (during 1 generation):

$$
\begin{cases}
\text{farmers } (F): \quad P_F' = P_F + f\,\dfrac{P_F P_H}{P_F + \gamma P_H} \\[2em]
\text{hunter} - \text{gatherers } (H): \quad P_H' = P_H - f\,\dfrac{P_F P_H}{P_F + \gamma P_H}
\end{cases}
$$

f = intensity of cultural transmission

γ = preference of *H*s to copy *F*s rather than *H*s (if $\gamma < 1$)

A different model: Lotka-Volterra equations

$$P_F' = P_F + \eta P_F P_H \quad (1)$$
$$P_H' = P_H - \eta P_F P_H \quad (2)$$

They have 2 problems:

1) They are not derived from cultural transmission theory

2) Number of HGs converted per farmer according to Eq. (1): $\dfrac{P_F' - P_F}{P_F} = \eta P_H \to \infty$! No maximum!

Fort (*PNAS* 2012)   if $P_H \to \infty$

$$\begin{cases} P'_F = P_F + f \dfrac{P_F P_H}{P_F + \gamma P_H} \approx P_F + C\, P_F \\[2em] P'_H = P_H - f \dfrac{P_F P_H}{P_F + \gamma P_H} \approx P_H - C\, P_F \end{cases}$$

$$C = \dfrac{f}{\gamma}$$

intensity of acculturation

if $P_H \gg P_F$:

$\dfrac{P'_F - P_F}{P_F} = C$ is the number of *H*s converted by farmer

$\dfrac{P'_F - P_F}{P_F}$ is not ∞, in contrast to Lotka-Volterra eqs.

The front speed does not depend on f and γ separately, but only on the number of HGs converted by farmer, $C = \dfrac{f}{\gamma}$.

10

How to simulate a spread of farming?
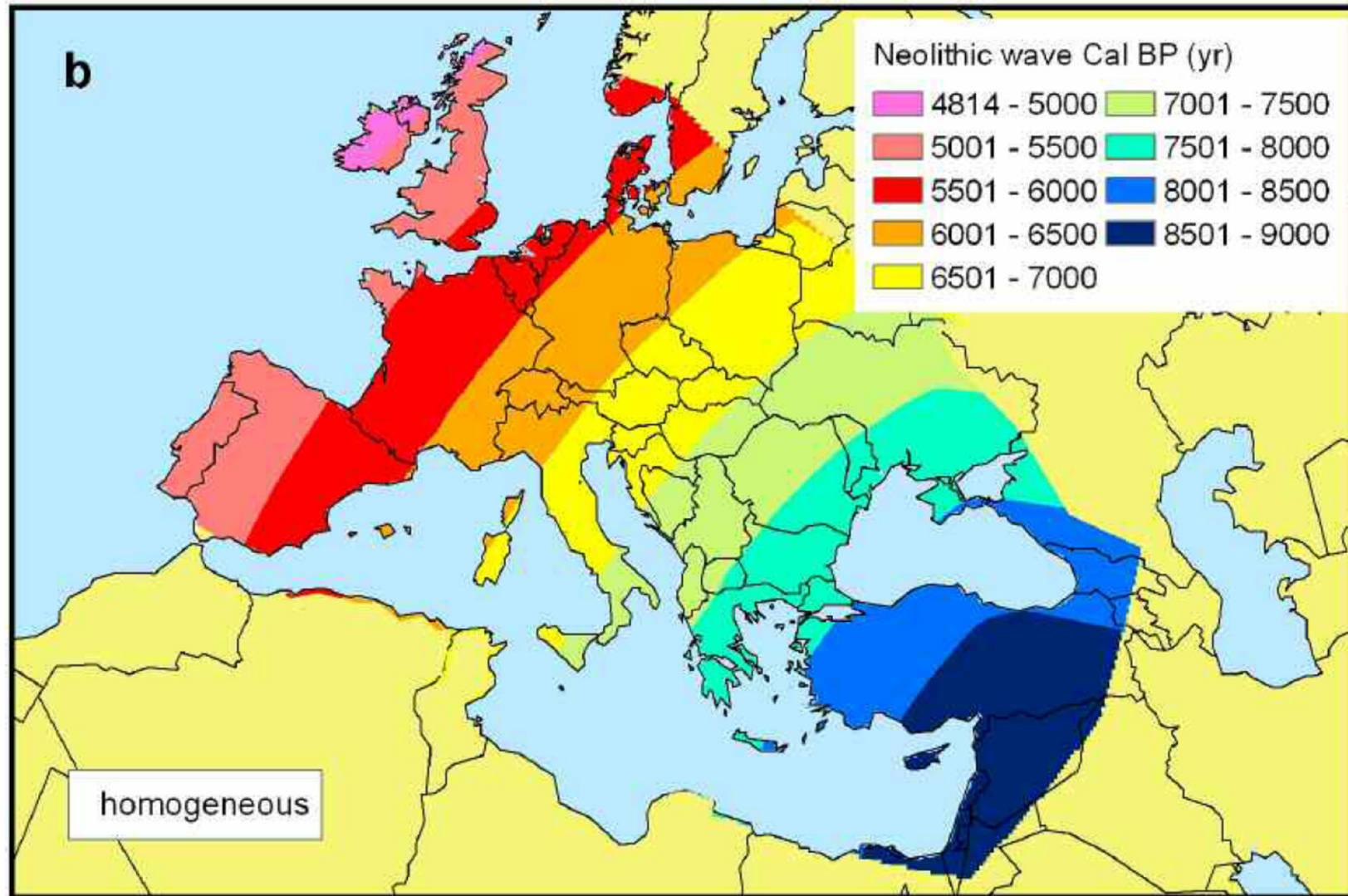
3 steps (every generation):

1. Reproduction (logistic)
2. Cultural transmission (acculturation)
3. Dispersal (distances r_j with probabilities p_j)

Initially there are farmers in only a limited region

The order of steps does not change the spread rate

This cycle is repeated many times (once per generation)

Examples of demic simulations

b

Neolithic wave Cal BP (yr)
- 4814 – 5000
- 5001 – 5500
- 5501 – 6000
- 6001 – 6500
- 6501 – 7000
- 7001 – 7500
- 7501 – 8000
- 8001 – 8500
- 8501 – 9000

homogeneous

demic simulations: no acculturation

Fort, Pujol & vander Linden, Amer. Antiq. (2012)

12

Neolithic wave Cal BP (yr)
- 5644 - 6000
- 6001 - 6500
- 6501 - 7000
- 7001 - 7500
- 7501 - 8000
- 8001 - 8500
- 8501 - 9000
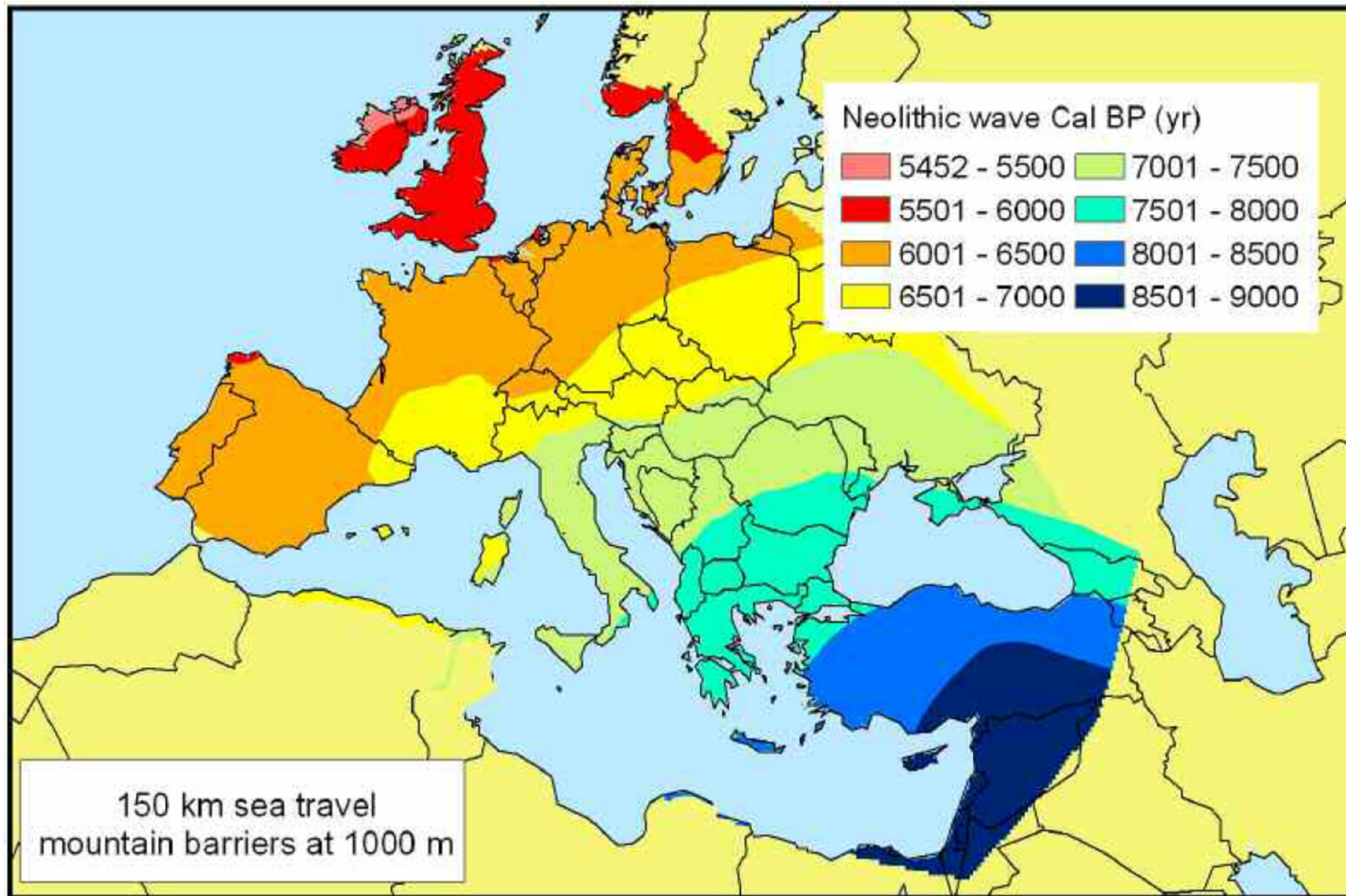
150 km sea travel
no mountains

demic simulations: no acculturation

Fort, Pujol & vander Linden, Amer. Antiq. (2012)

13

demic simulations: no acculturation

Fort, Pujol & vander Linden, Amer. Antiq. (2012)

14
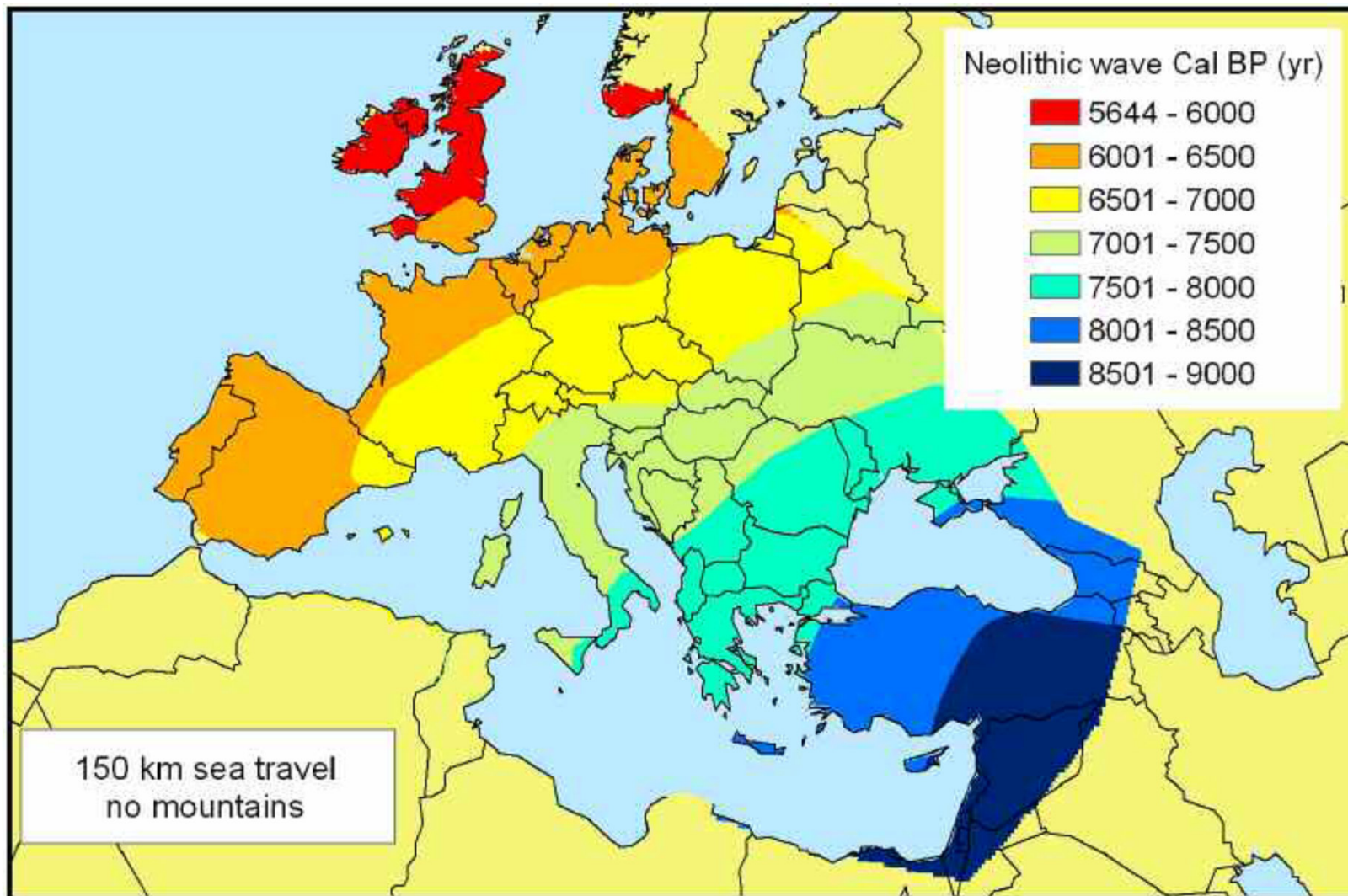
Mathematical models

They are useful to make quantitative estimations (e.g., of the intensity of acculturation)

1) Fisher's model

Applied by Ammerman & Cavalli-Sforza (1973, 1984), etc.

$$\frac{\partial N}{\partial t} = D_N \nabla^2 N + a_N N \left(1 - \frac{N}{K_N}\right) \quad \rightarrow \quad \text{speed} = 2\sqrt{a_N D_N}$$

N = population density of the Neolithic population

This model causes an error of up to <u>30%</u> [1]
It is more precise to use a cohabitation model (next slide).
<u>This is a purely demic model: no acculturation</u>

[1] Isern, Fort & Pérez-Losada, *J. Stat. Mechs. Theor. & Exp.* (2008)

2) Cohabitation models
2a) Purely demic model

$$N(x, y, t+T)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_T\left[N(x + \Delta_x, y + \Delta_y, t)\right] \, \phi_N\left(\Delta_x, \Delta_y\right) d\Delta_x \, d\Delta_y$$

Logistic reproduction: $R_T[N(x, y, t)] = \dfrac{e^{a_N T} K_N N(x,y,t)}{K_N + \left(e^{a_N T} - 1\right) N(x,y,t)}$

$\phi_N\left(\Delta_x, \Delta_y\right)$ is a set of probabilities p_j for farmers to disperse at distances r_j during a generation time T.

$$\text{speed} = \min_{\lambda > 0} \frac{a_N T + \ln\left[\sum_{j=1}^{M} p_j I_0(\lambda r_j)\right]}{T\lambda}$$

$I_0(\lambda r_j) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \, \exp[-\lambda r_j \cos\theta]$ is the modified Bessel function of the first kind and order zero

2b) Demic-cultural model

$$\begin{cases} N(x,y,t+T) = \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{N}(x+\Delta_x,y+\Delta_y,t)\phi_N(\Delta_x,\Delta_y)d\Delta_x\,d\Delta_y \\[2em] P(x,y,t+T) = \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{P}(x+\Delta_x,y+\Delta_y,t)\phi_P(\Delta_x,\Delta_y)d\Delta_x\,d\Delta_y \end{cases},$$

where

$$\tilde{N}(x,y,t) \equiv R_T[N(x,y,t)] + f\frac{R_T[N(x,y,t)]R_T[P(x,y,t)]}{R_T[N(x,y,t)]+\gamma R_T[P(x,y,t)]}$$

$$\tilde{P}(x,y,t) \equiv R_T[P(x,y,t)] - f\frac{R_T[N(x,y,t)]R_T[P(x,y,t)]}{R_T[N(x,y,t)]+\gamma R_T[P(x,y,t)]},$$
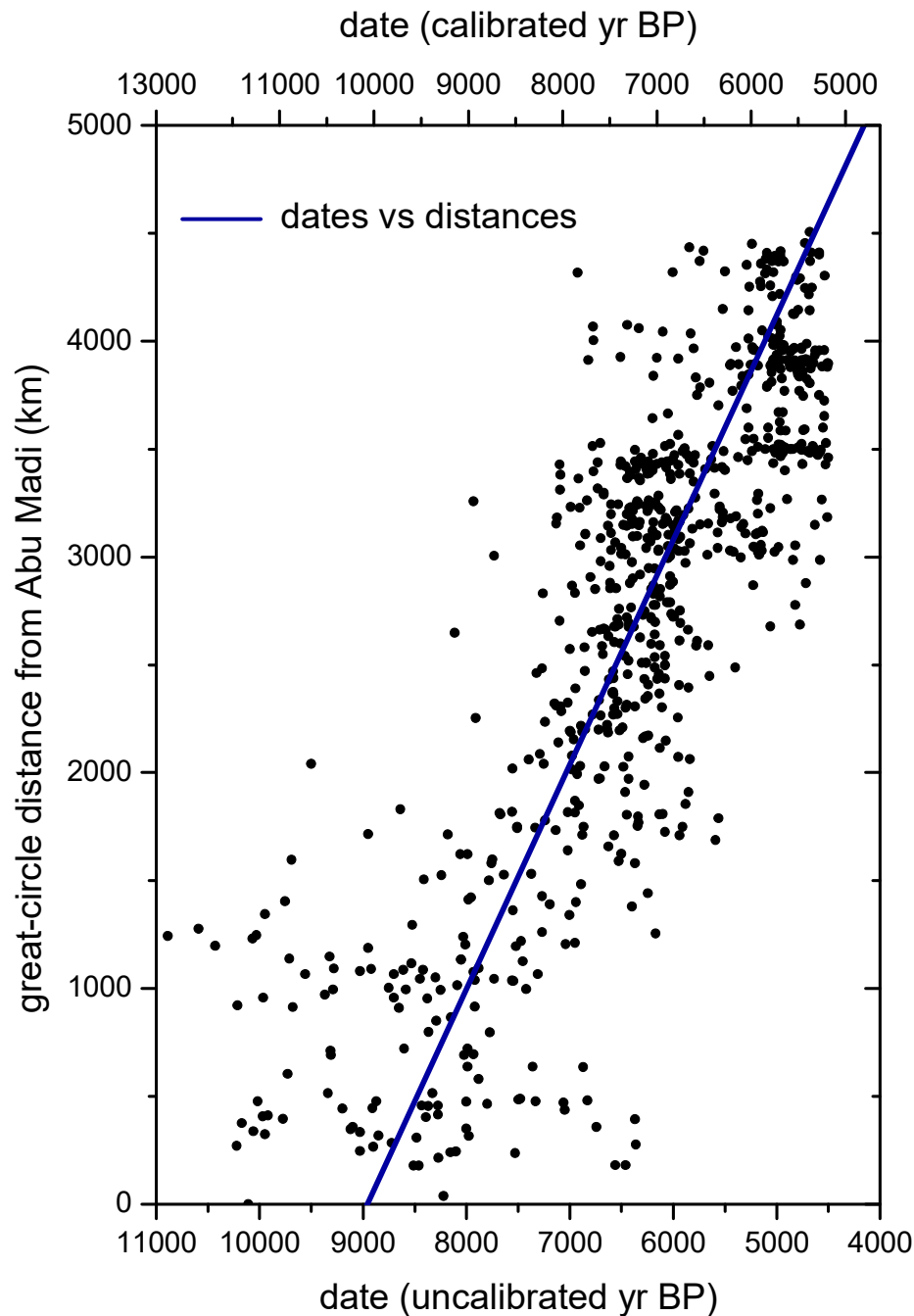
$$\text{speed} = \min_{\lambda>0}\frac{a_N T + \ln\left[(1+C)\left(\sum_{j=1}^{M}p_j I_0(\lambda r_j)\right)\right]}{T\lambda}$$

$$C = \frac{f}{\gamma}$$

intensity of acculturation

We will use this model
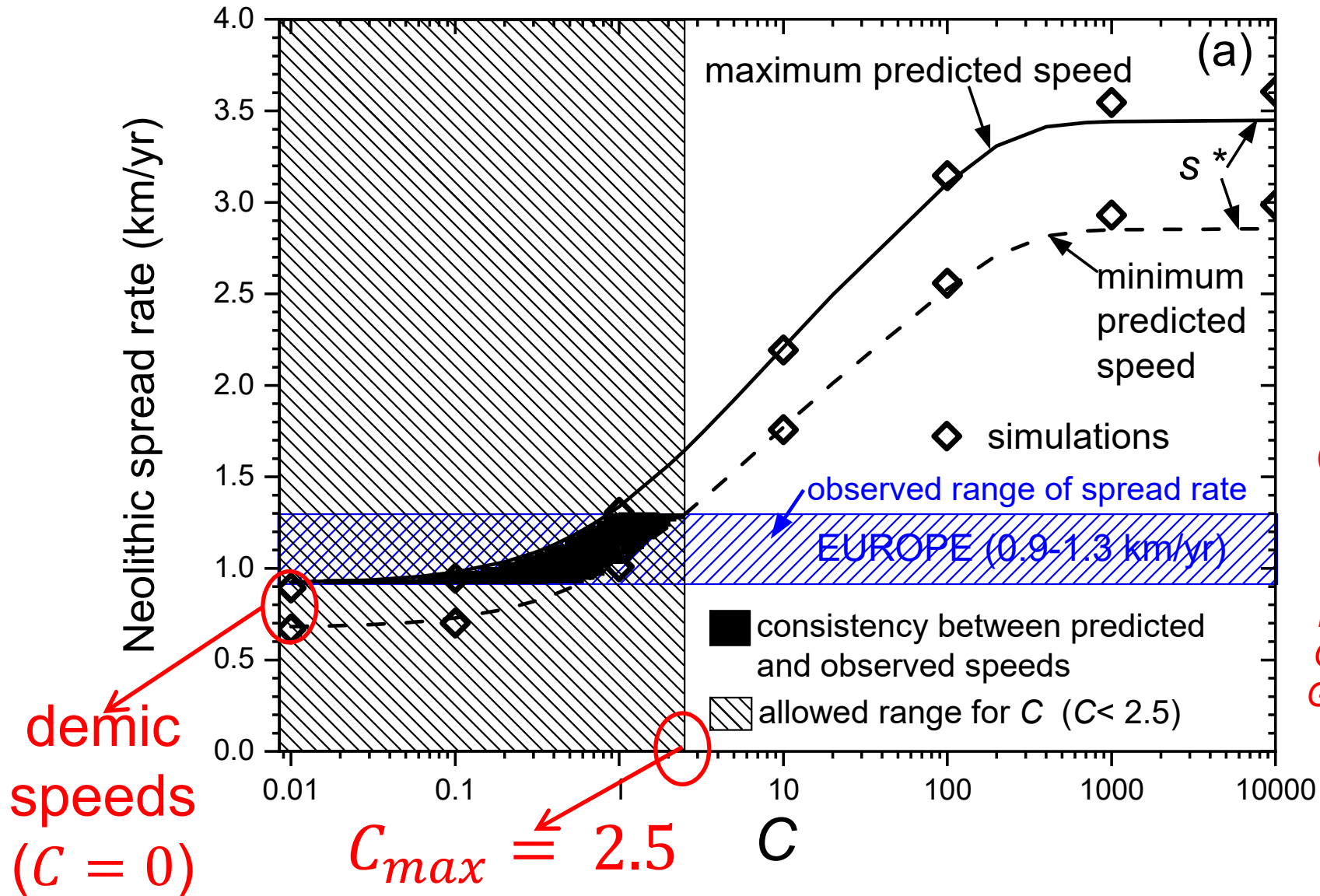
To apply our model we need the observed speed

For the spread of the Neolithic in Europe:

0.9-1.3 km/yr

735 sites in Europe & Near East

$r = 0.83$ (highest-r origins, great circles & shortest paths)

Pinhasi, Fort & Ammerman, *PLoS Biol.* (2005)

18

Effect of acculturation intensity C on the front speed in Europe

Effect of cultural diffusion in Europe

Effect (%) = (speed – demic speed) /speed · 100

- – – effect for the minimum predicted speed
- —— effect for the maximum predicted speed

cultural effect (%)

(b)

ALLOWED RANGE

0-48% cultural effect

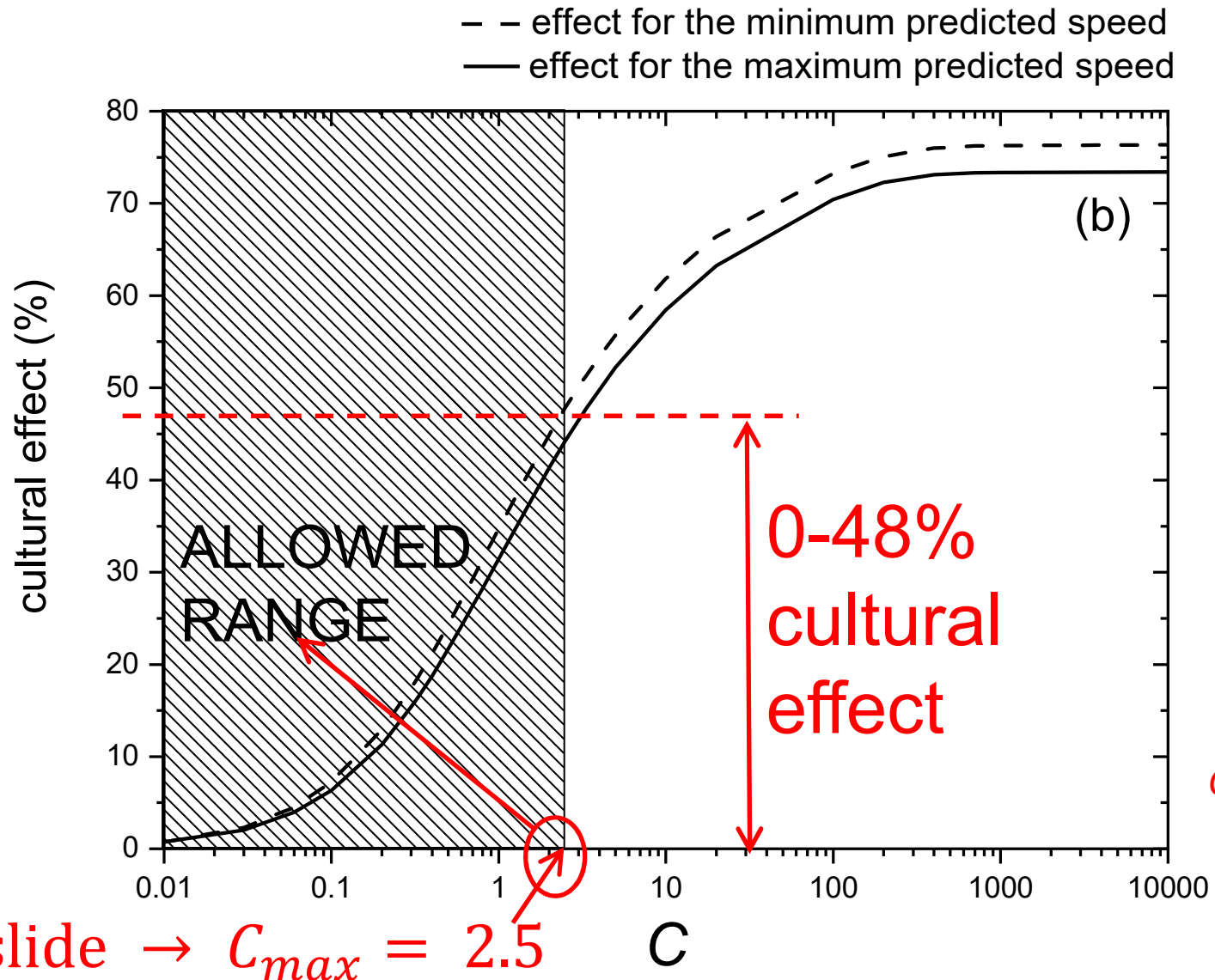Fort, *PNAS* (2012)

Fort, *Hum. Popul. Gen. & Genom.* 2022

C

previuos slide → $C_{max} = 2.5$

The spread of domesticated rice

$C_{max} = 0.5$

0-24% cultural effect

Cobo, Fort & Isern, *J. Arch. Sci.* (2019)

21
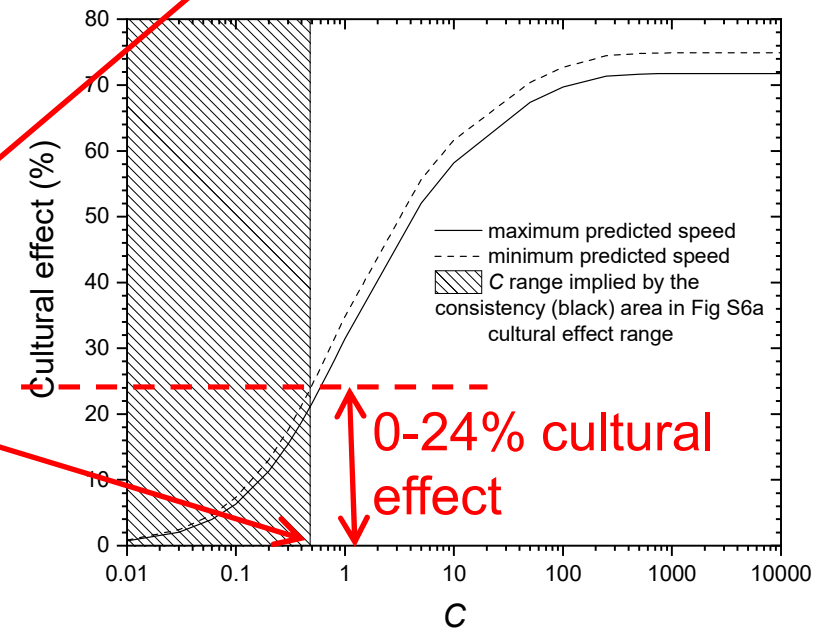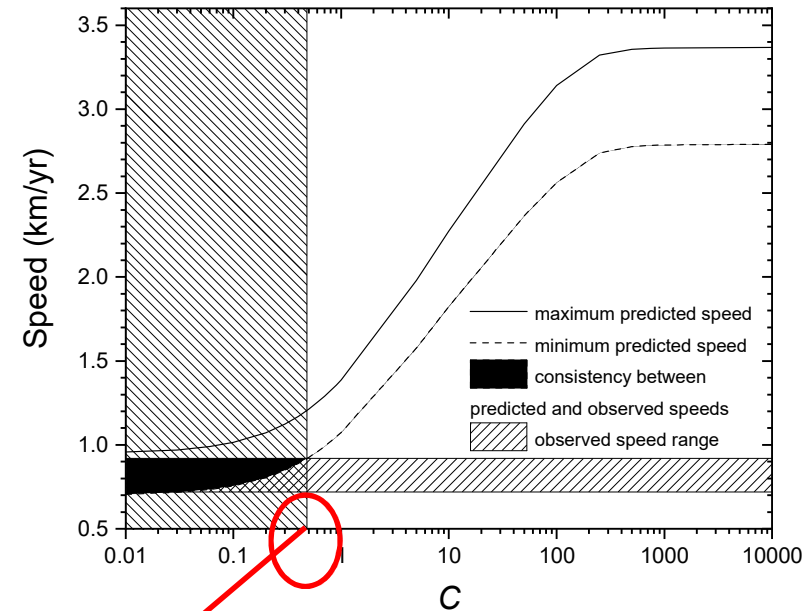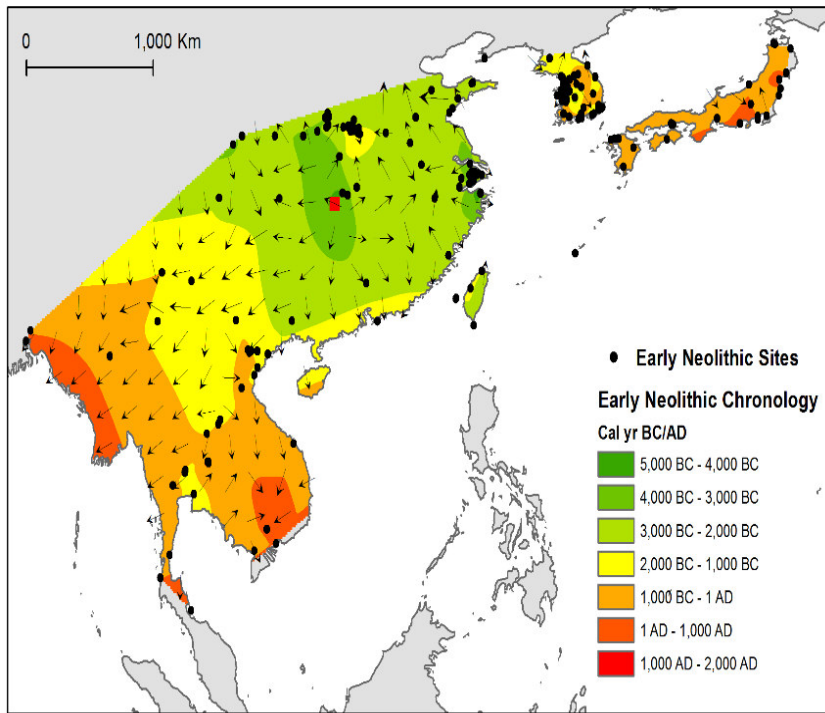
The spread of domesticated rice

The results depend on the dispersal kernel of rice cultivators:

1. Mehrai (1984), previous slide: $C<0.5$, cultural effect = 0%-24%.

2. Sing & Sing (2015): $C<0.3$, cultural effect =0%-13%.

3. Shukla (2015): $C<0.3$, cultural effect=0%-12%.

In all 3 cases: cultural effect<50%, so mainly demic.

Genetics

mtDNA haplogroup K: absent in hunter-gatherers

This pattern in early farmers suggests interbreeding with HGs
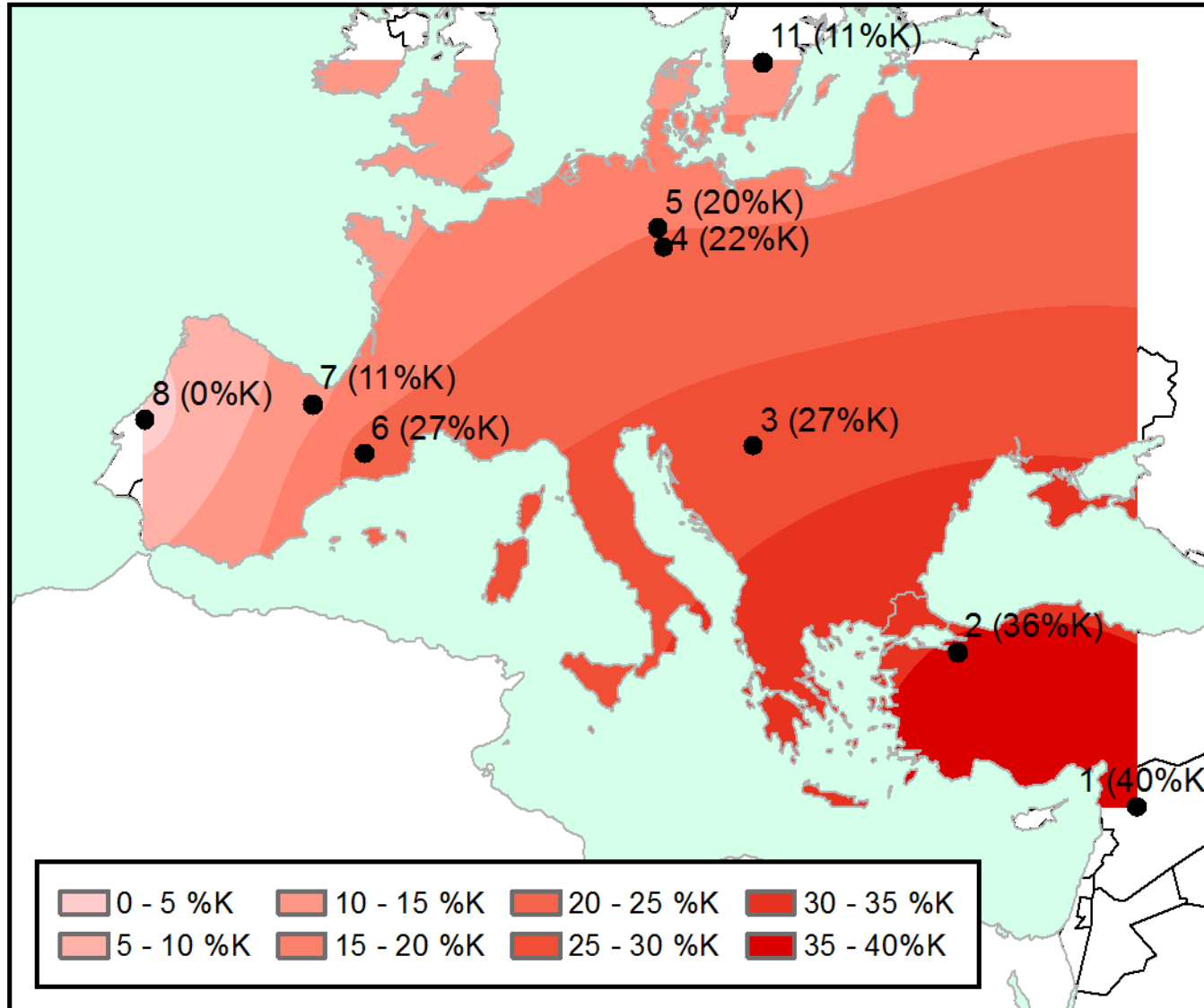
Isern, Fort & de Rioja, *Sci. Rep.* (2017)

Simulations

Grid of square cells. Initially farmers only at the cell containing the oldest site in Upper Mesopotamia (Abu Hureyra) with a %K such that we obtain the observed %K (47.4%) at the average location and date of the 15 early farmers in Upper Mesopotamia whose mtDNA is known.

All other grid cells are initially empty of farmers and with HGs at their saturation density.

At each node in the grid and time step (1 generation=32 yr), we compute 3 processes:
(1) Dispersal (38% do not migrate, from ethnography). Migration threshold: migration only if the farmer density is > 0.06 farmers/km^2, from archaeology and ethnography.
(2) Cultural transmission: next slide.
(3) Reproduction: next slide.

24

Simulations

(2) Cultural transmission:

P_N =farmers who _have_ haplogroup K.

P_X = farmers who <u>do _not have_</u> haplogroup K.

P_{HG} = hunter-gatherers (all without haplogroup K).

$$\%K = \frac{P_N}{P_N + P_X}$$

Cultural transmission theory (Cavalli-Sforza & Feldman 1981; Fort 2011, 2012): Interbreeding: $couples\ HN = C\dfrac{P_{HG}P_N}{P_{HG}+P_N+P_X}$

$$couples\ HX = C\frac{P_{HG}P_X}{P_{HG}+P_N+P_X}$$

random mating for farmers$\rightarrow couples\ NX = \dfrac{P_N P_X}{P_N+P_X}$

(3) Reproduction: each couple of farmers has 2Ro children (Ro=2.45). Genetically mixed matings (HN and NX) have 50% children N and 50% children X.

Archaeologists have identified 2 routes
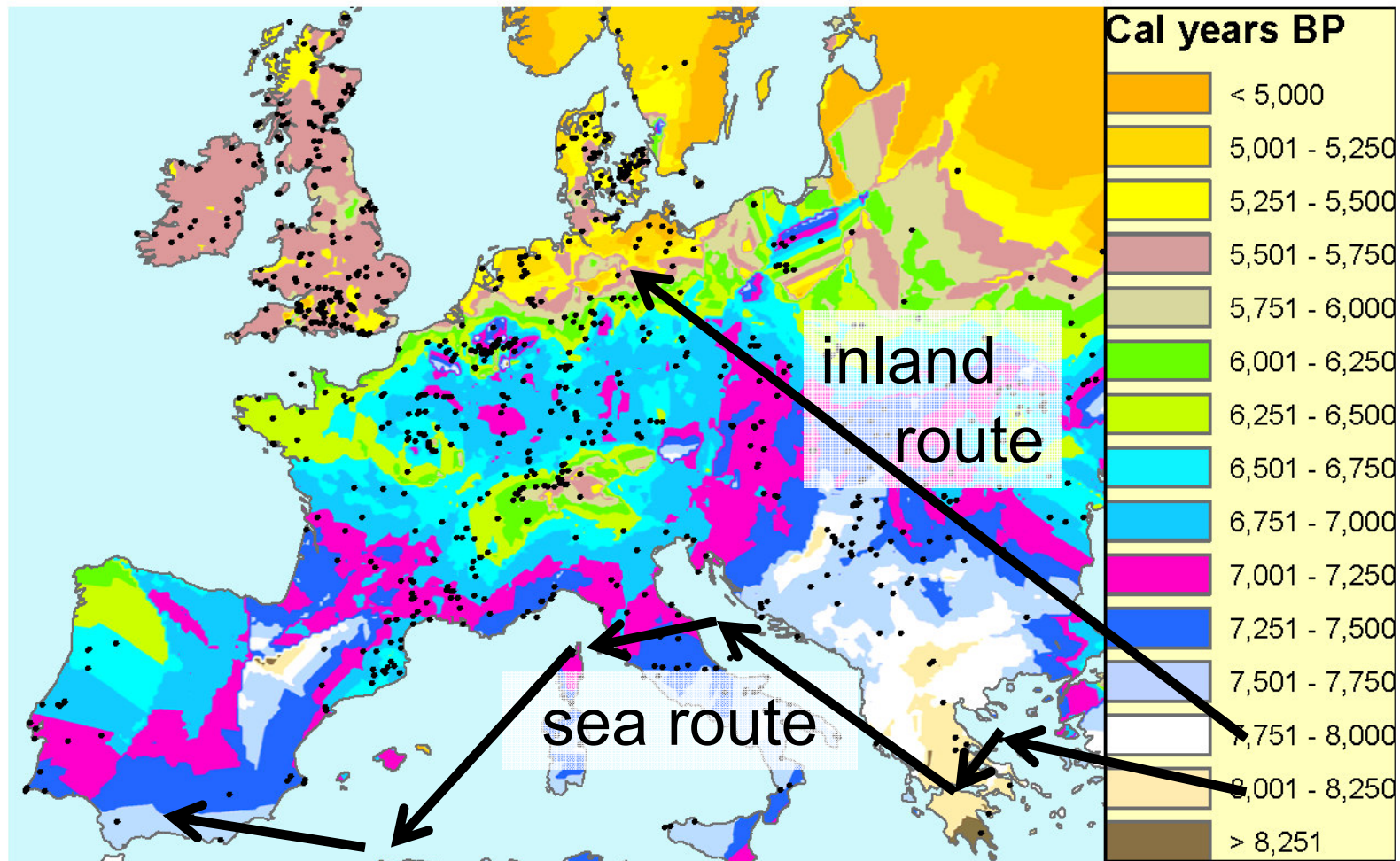
Now we have ancient genetic data for both routes

Initially there are farmers only at the cell with the **oldest PPNB site in Upper Mesopotamia (Abu Hureyra**, <9,038 cal BC) at a date (8,718 cal BC) such that the simulations agree with the data along the inland route (red).
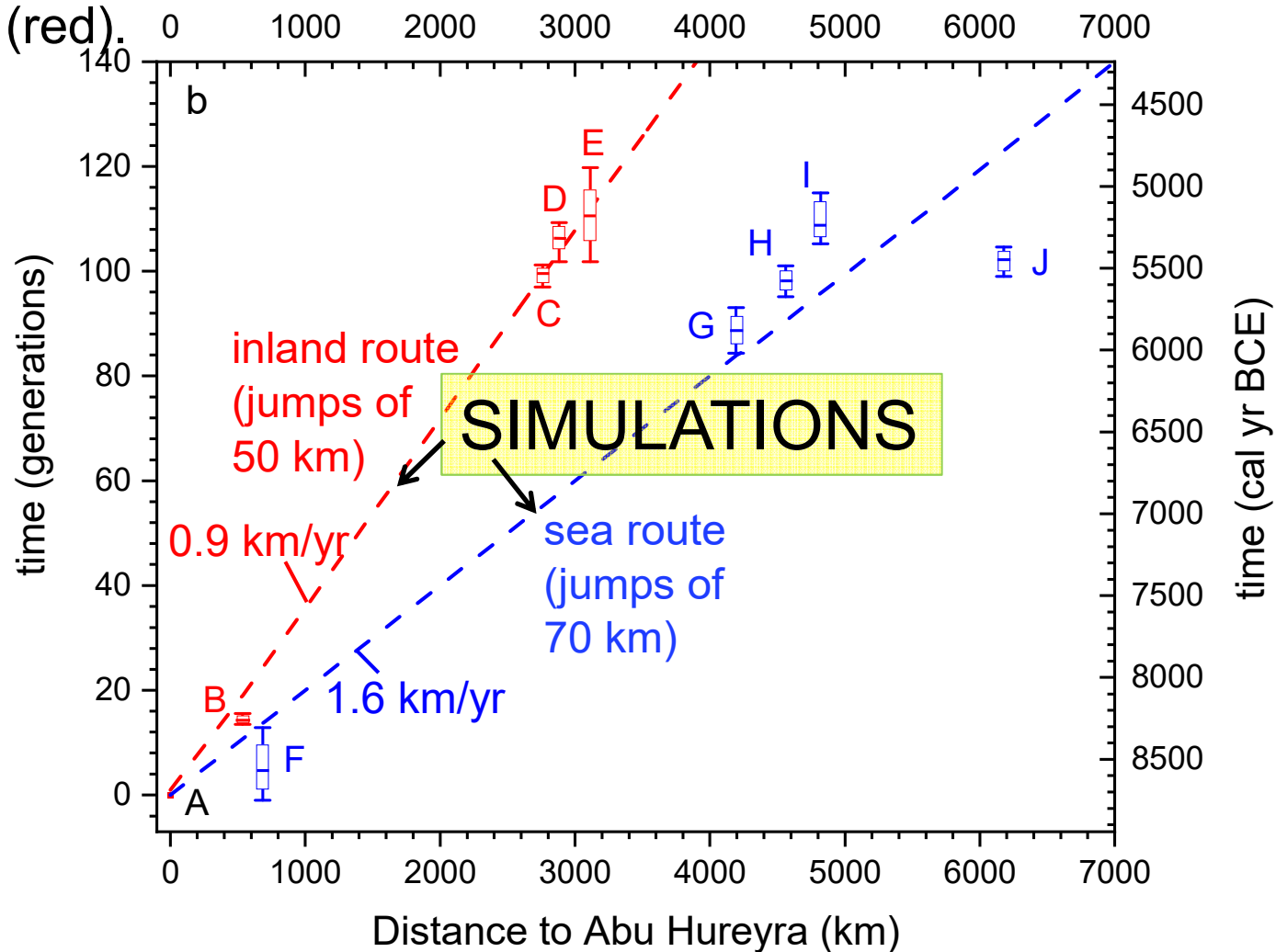
Inland route: simulations with jumps of 50 km per generation (value from ethnography)

Sea route: best fit for simulations with jumps of 70 km

27
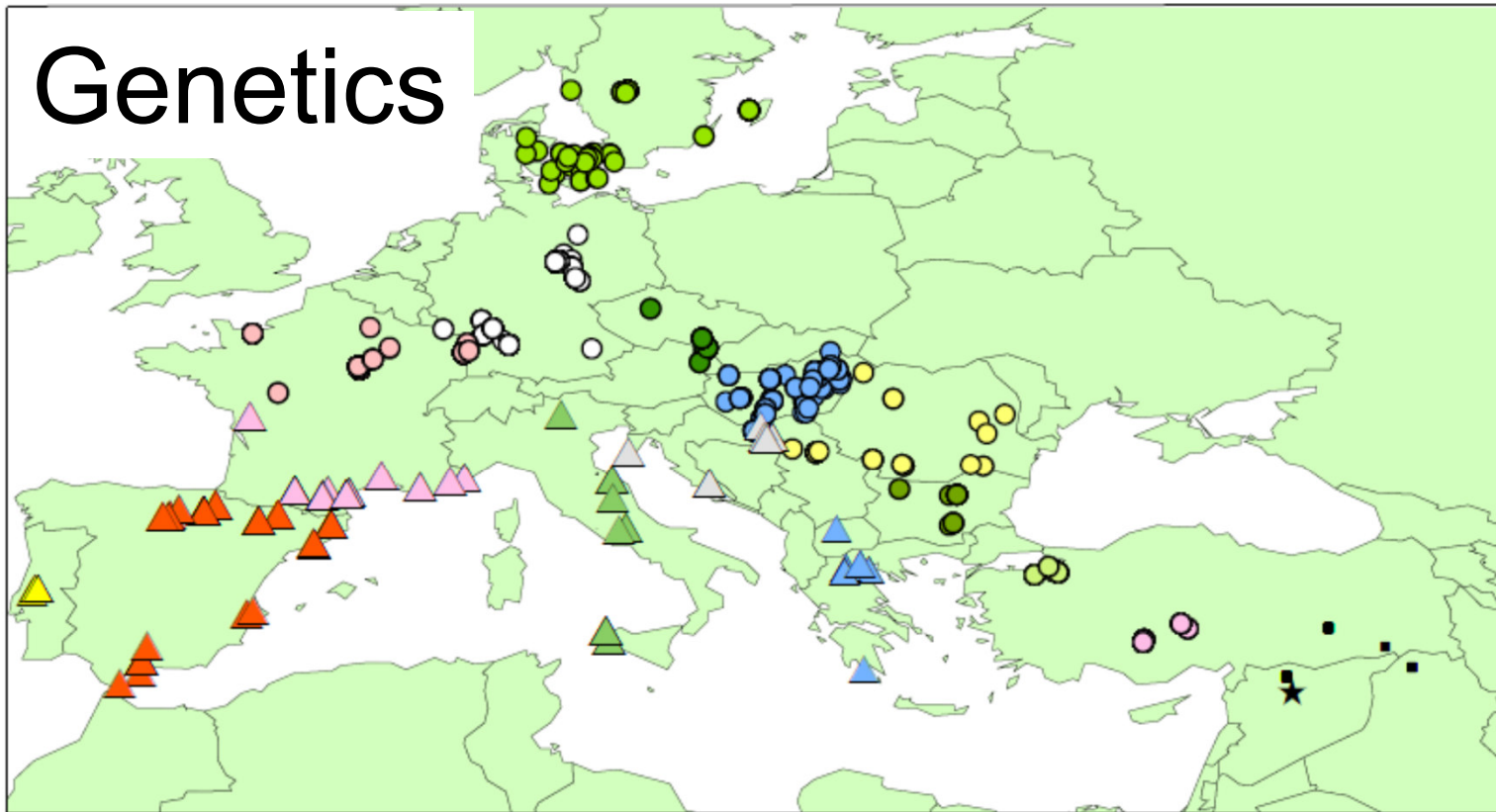
Genetics

Sites with genetic data. Circles and triangles stand for inland and Mediterranean routes

★	Abu Hureyra (oldest PPNB site in region 1, no genetic data)	○	9 Scandinavia
▪	1 Northern Mesopotamia (pressumed region of origin)	○	10 Northern France
○	2 Central Anatolia	▲	11 Greece and N. Macedonia
○	3 Western Anatolia	▲	12 Croatia
●	4 Bulgaria (except Malak Preslavets)	▲	13 Italy
○	5 Romania and Serbia (except Iron Gates)	▲	14 Southern France
●	6 Hungary	▲	15 Spain
●	7 Austria and Czech R.	▲	16 Portugal
○	8 Germany		

28

(a) Oldest Neolithic genetic regional data

inland route (jumps of 50 km)

$C = 0$
$C = 0.02$
$C = 0.06$
$C = 0.07$
$C = 0.09$
$C = 0.08$

SIMULATIONS

3000 km

Haplogroup K (%)

Great-circle distance to Abu Hureyra (km)

Best fits: $C = 0.07$-0.08

29
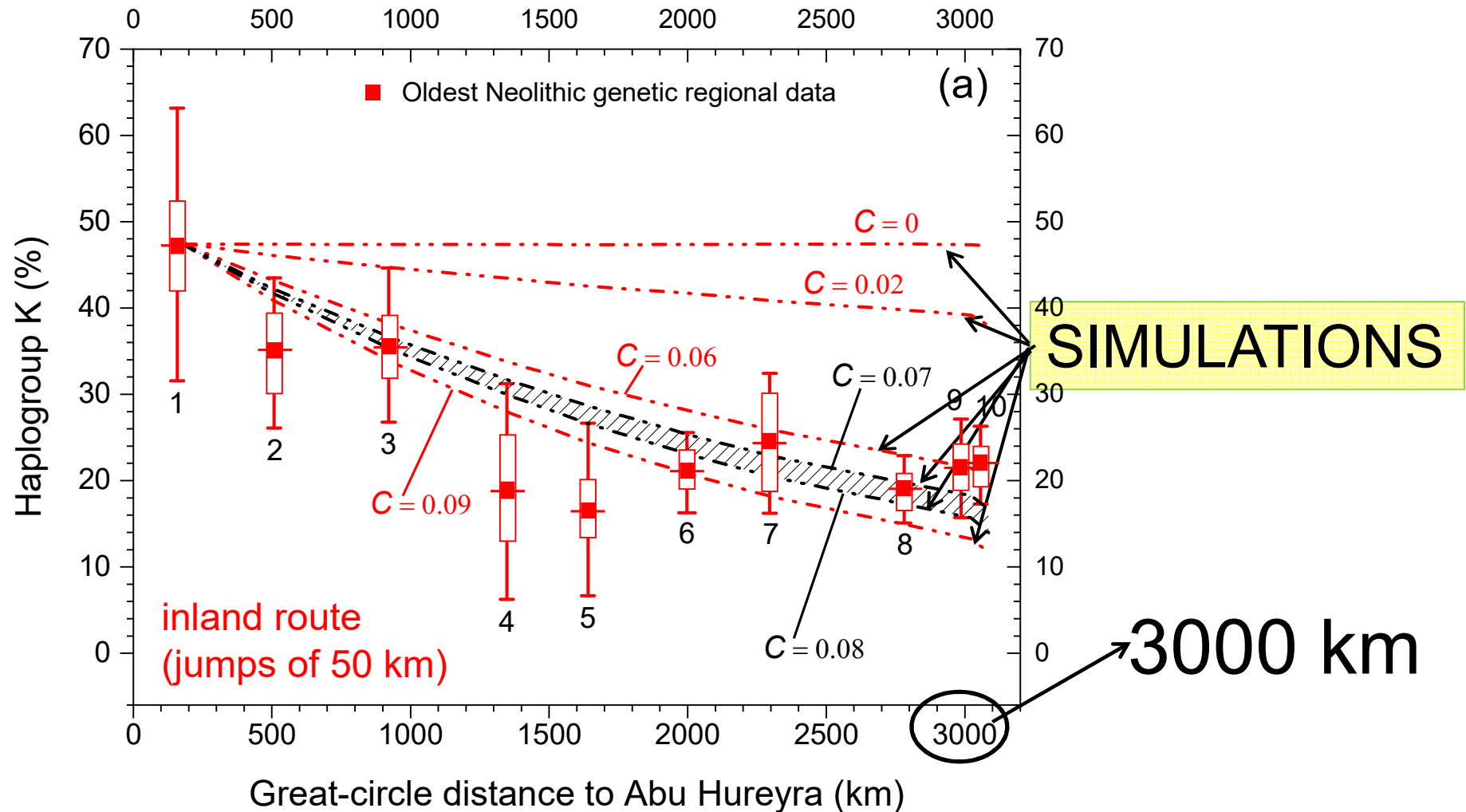
Mediterranean genetic cline

(b) Oldest Neolithic genetic regional data

Haplogroup K (%)

sea route (jumps of 70 km)

Distance along the coast to Abu Hureyra (km)

SIMULATIONS

6000 km
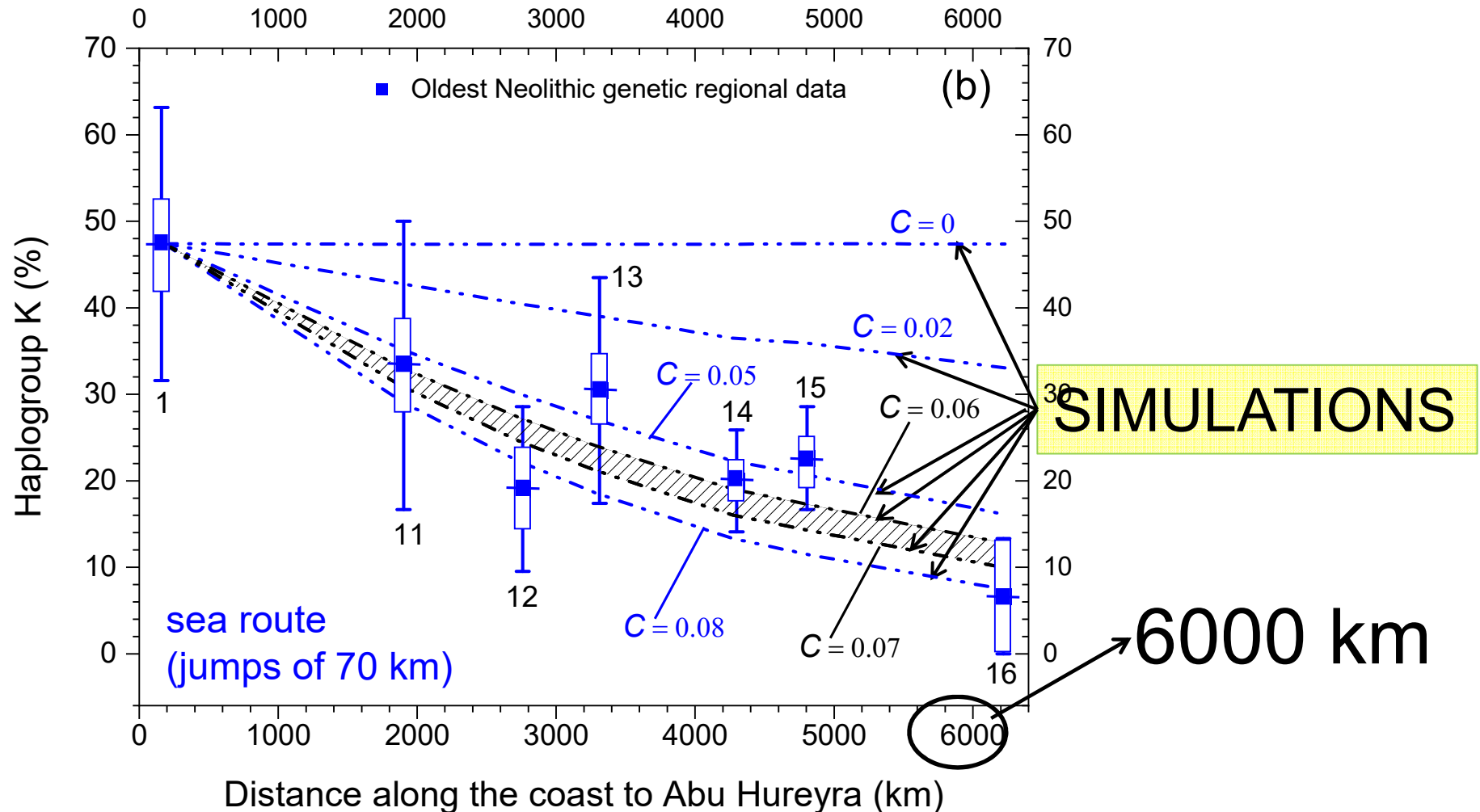
**Best fit: C = 0.06-0.07.
Esentially the same as for the inland route!**

30

The dispersal behavior depends on geography:
-early farmers moved longer distances per generation along the sea route.
In turn this led to:
-a faster spread rate along the sea route,
-a lower slope of the genetic cline along the sea route (due to less interbreeding events per unit distance).

In sharp contrast to this:
The number of farmers that acculturated a HG (**about 3.6%* of farmers, or C = 0.07*)** was the same along both routes. **It did not depend on geography but only on the transition in the subsistence economy and its associated way of life.**

*fraction of farmers $= \dfrac{P_F(x,y,t+1) - P_F(x,y,t)}{P_F(x,y,t)} = C \dfrac{P_{HG}(x,y,t)}{P_{HG}(x,y,t) + P_F(x,y,t)} = \dfrac{100\,C}{1 + \dfrac{p_{F\,min}}{p_{HG\,max}}}$

Uncertainties

·previous slides: about **3.6%** of farmers acculturated a HG ($\boldsymbol{C \approx 0.07}$).

·Taking into account the uncertainties in the parameter values ($p_{F\,max}$, $p_{HG\,max}$, $p_{F\,min}$, R_0) and in the initial frequencies of haplogroup K: **1% - 8%** of farmers acculturated a HG ($\boldsymbol{0.03 < C < 0.14}$).

We go back to a figure obtained in a previous slide from Archaeology*:

– – effect for the minimum predicted speed
—— effect for the maximum predicted speed

Archaeology: 0<C<2.5

Genetics: 0.03<C<0.14
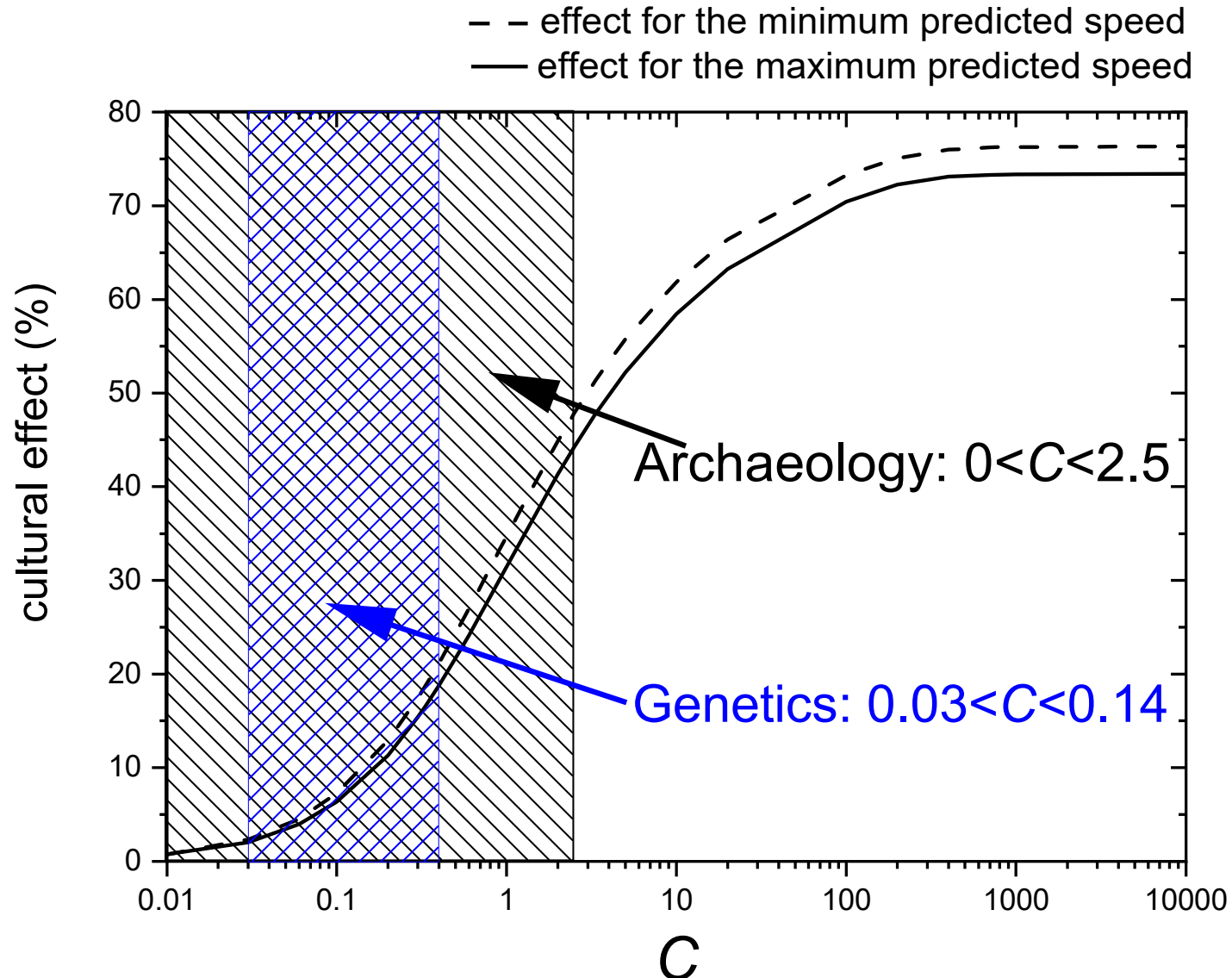
*Although the model used to obtain this figure it is not exactly the same (because it applies several dispersal distances and no migration threshold), we can use it to compare approximately Archaeology to Genetics.

33

Linguistics

Can we apply the same approach as for Archaeology?
We need speeds. Some authors have estimated them:
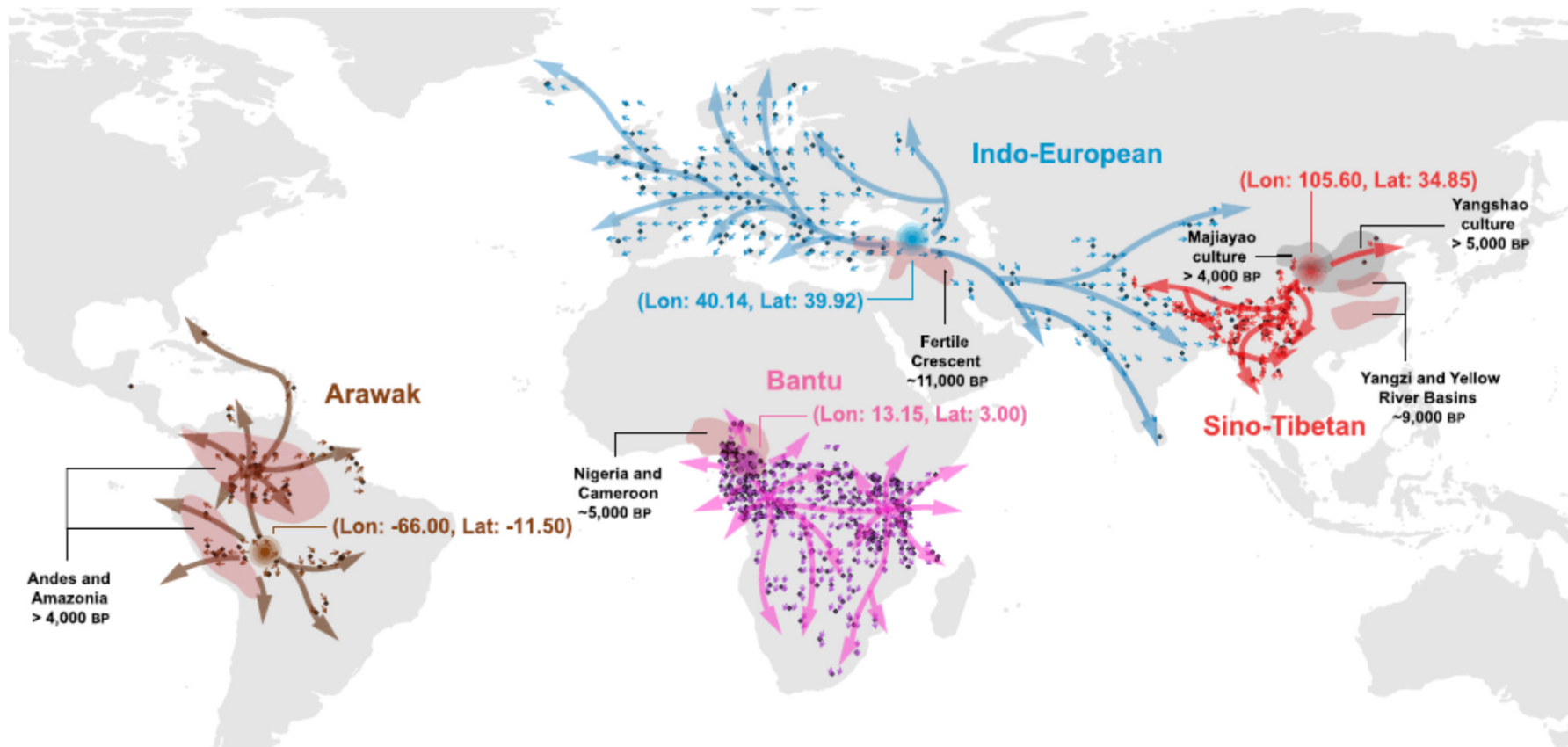
Yang, Sun, Jin, & Zhang, *Nature Comm.* (2024)

Linguistics

Indo-European languages

Data from Menghan and Stan

Equation	y = a + b*x
Plot	yr BP
Weight	No Weighting
Intercept	2336.50596 ± 252.7769
Slope	-0.50054 ± 0.09413
Residual Sum of Square	4.21716E8
Pearson's r	-0.34967
R-Square (COD)	0.12227
Adj. R-Square	0.11795
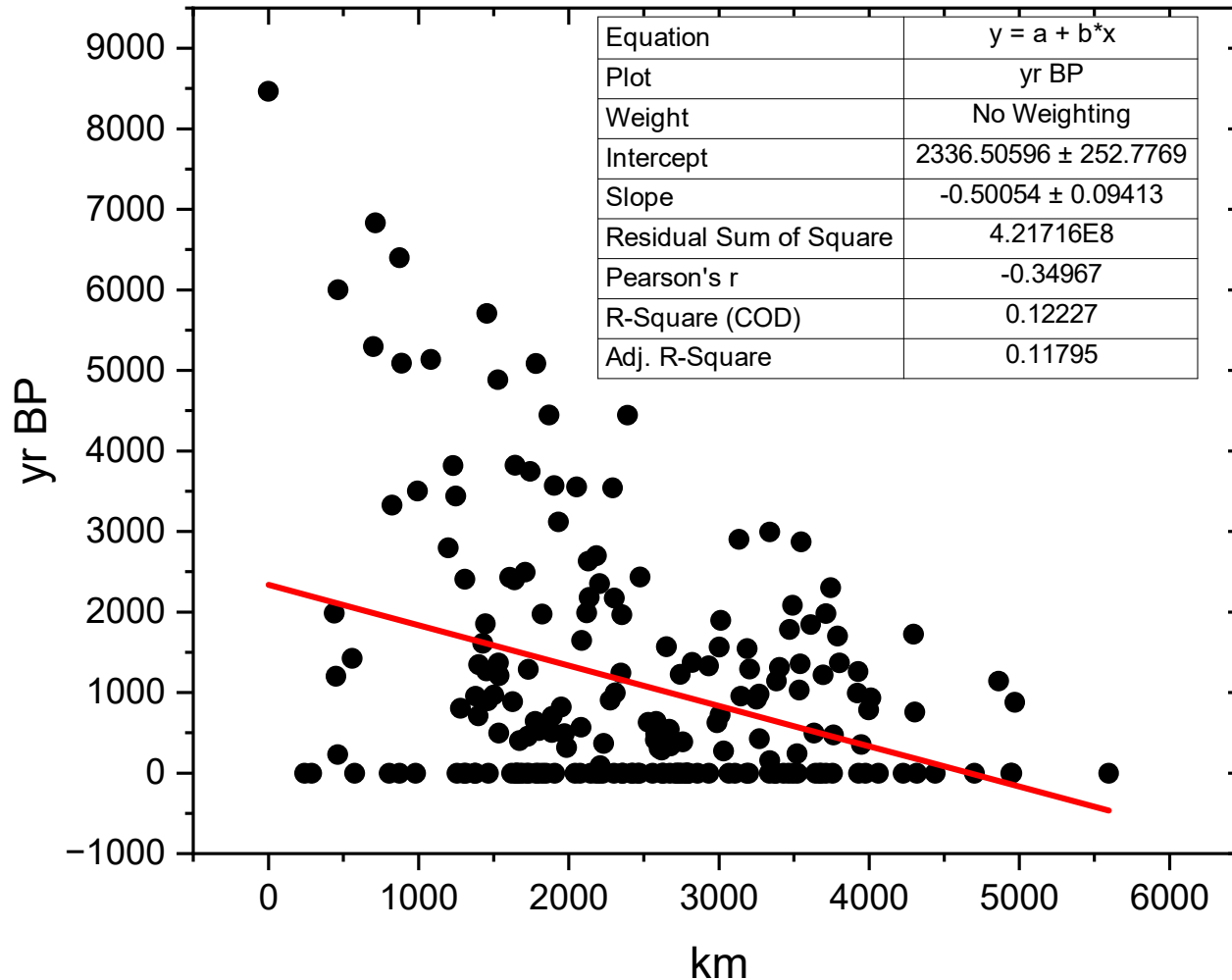
Very poor correlation (r=0.35). Not surprising: the same happens in Archaeology.

We should use the oldest date for a given distance.

Very encouraging that **all data are in a triangle!**

Using the oldest date for each 400-km interval:

Equation	y = a + b*x
Plot	age
Weight	No Weighting
Intercept	7819.11676 ± 357.50138
Slope	-1.52423 ± 0.1065
Residual Sum of Squares	6016380.19899
Pearson's r	-0.9697
R-Square (COD)	0.94032
Adj. R-Square	0.93573
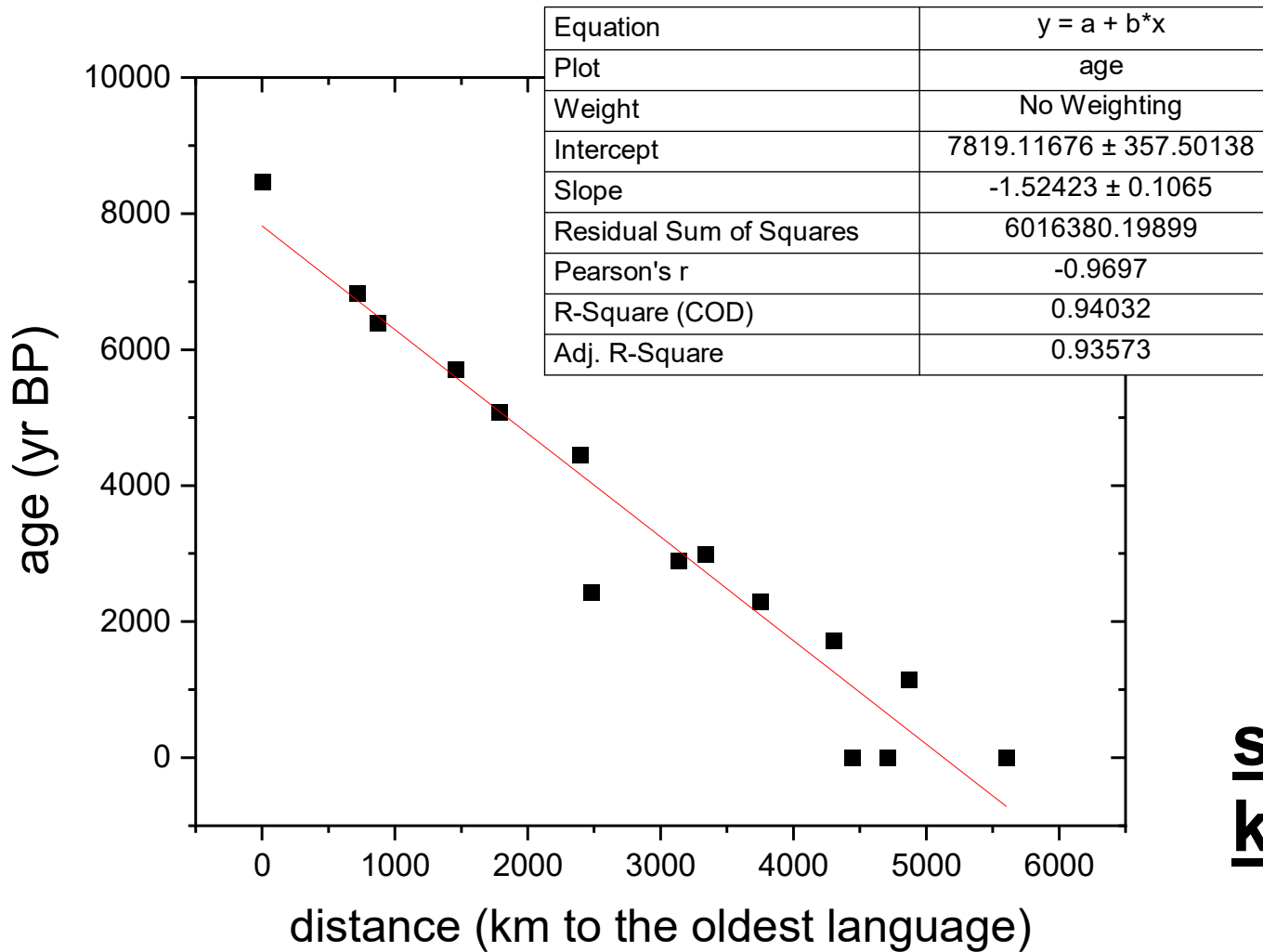
Very good correlation (r=0.97).

speed = 0.4-0.9 km/yr (95% CL)

age (yr BP)

distance (km to the oldest language)

We go back to a figure obtained in a previous slide from Archaeology, but change the spread rate for the Neolithic in Europe (0.9-1.3 km/yr) by the spread rate for Indo-European languages (**0.4-0.9 km/yr**):

◇ simulations

(a)

maximum predicted speed

s^*

minimum predicted speed

■ consistency between predicted and observed speeds

observed range of spread rate

Indo-European languages (0.4-0.9 km/yr)

▨ allowed range for C ($C<0.06$)

Neolithic spread rate (km/yr)

C

demic speeds ($C = 0$)
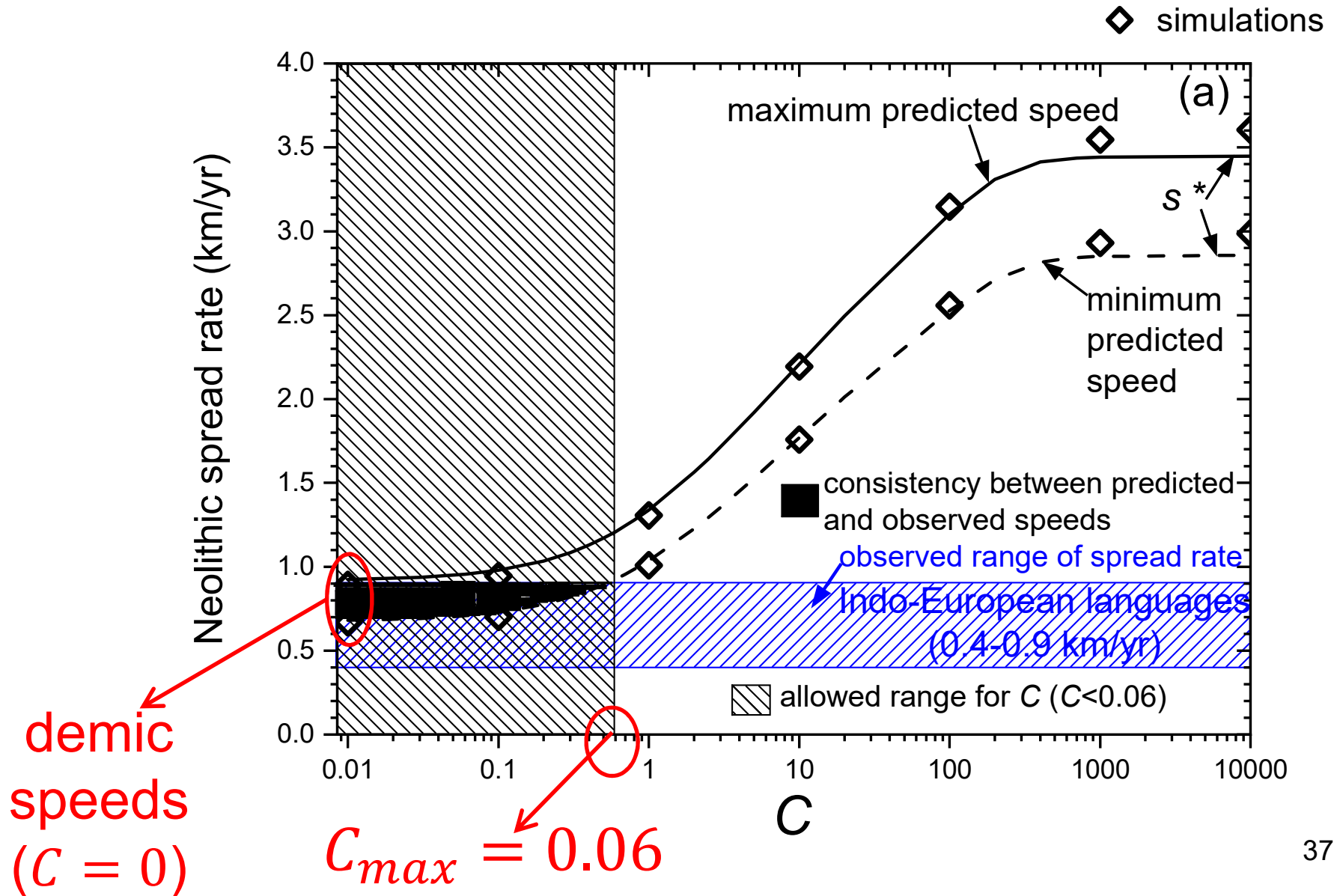
$C_{max} = 0.06$

Effect of cultural diffusion in the spread of Indo-European languages

Effect (%) = (speed – demic speed) /speed · 100

- – effect for the minimum predicted speed
- — effect for the maximum predicted speed

ALLOWED RANGE

(b)

cultural effect <5%

cultural effect (%)

C
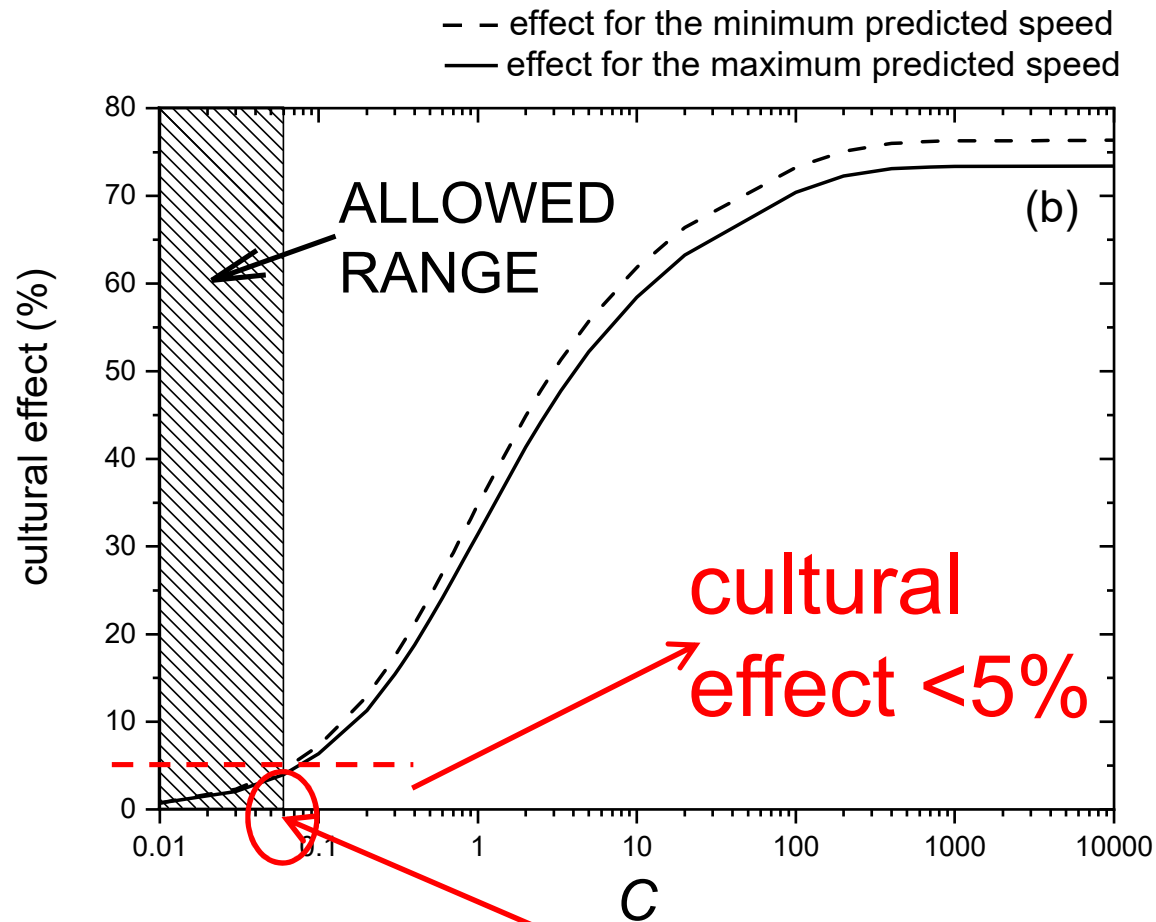
previuos slide → $C_{max} = 0.06$

Results for Indo-European languages:

- $C_{max} = 0.06 \rightarrow$ less tan 6%* of Indo-European (IE) speakers converted a non-IE speaker*.

- Cultural diffusion was responsible for <5% of the spread rate $\rightarrow$ demic diffusion was responsible for >95%.

- It was an overwhelmingly demic spread.

*In the two previous figures, we have used the same model as for the Neolithic in Europe (no dispersal threshold).

Next case studies to be analyzed

- Sino-Tibetan languages
- Bantu languages

+ compare to Indo-European

+ language expansions not driven by population growth? These models could be also applied.

Questions?

Appendix: Derivation of the cultural transmission Eqs.:

$$\begin{cases} \text{farmers } (F): \quad P_F' = P_F + f \dfrac{P_F P_H}{P_F + \gamma P_H} \\[2em] \text{hunter} - \text{gatherers } (H): \quad P_H' = P_H - f \dfrac{P_F P_H}{P_F + \gamma P_H} \end{cases}$$

Cavalli-Sforza & Feldman (*book* 1979):

$n =$ number of teachers (other tan parents) that a H individual contacts during his/her lifetime. $n \sim P_H + P_F$ yields the Lotka-Volterra Eqs., but n is independent of P_H and P_F (Dunbar 1993).

·Assume: of these teachers, a portion $u = \dfrac{P_F}{P_F + P_H}$ is of type F (in the next slide, we drop this assumption).

·Then: number of teachers of type F = $n\,u$.

·$g =$ probability that a H individual becomes F due to contact with a single F-teacher.

·Then: probab. that a H indiv. becomes F after n contacts = $1 -$ probab. that he/she does not = $1 - (1 - g)^{nu} \approx ngu = fu$ if $g \ll 1$.
$f = ng$. Thus: number of H indivs. becoming F per gen.= $fu\,P_H$. 42

Model in the previous slide

Number of H individuals becoming F per generation =

$$= f u\, P_H = f\, \frac{P_F\, P_H}{P_F + P_H}.$$

Therefore:

$$\begin{cases} \text{farmers } (F): \quad P_F' = P_F + f\, \dfrac{P_F P_H}{P_F + P_H} \\[2em] \text{hunter} - \text{gatherers } (H): \quad P_H' = P_H - f\, \dfrac{P_F P_H}{P_F + P_H} \end{cases}$$

Serious limitation:

If $P_H \ll P_F$ then $\quad P_H' = P_H(1 - f) \rightarrow f < 1$

If $P_F \ll P_H$ then $P_F' = P_F(1 + f) \rightarrow$ Since $f < 1$, each F indiv. can at most convert a single H in their lifetime.

More general model

In the previous model, we have assumed that each H individual has a portion of F teachers $= u = \dfrac{P_F}{P_F + P_H}$.

Thus, an H individual is equally likely to contact with an F or H individual.

Here we assume that, for learning purposes, an H individual contacts only with a fraction α of his/her F neighbors and a fraction β of his/her H neighbors. Then the number of F teachers that an H individual contacts is $n\dfrac{\alpha P_F}{\alpha P_F + \beta P_H} = n\dfrac{P_F}{P_F + \gamma P_H}$, where $\gamma = \dfrac{\beta}{\alpha}$. Replacing u by this in the former derivation, we obtain:

$$\begin{cases} P'_F = P_F + f\dfrac{P_F P_H}{P_F + \gamma P_H} \\[4mm] P'_H = P_H - f\dfrac{P_F P_H}{P_F + \gamma P_H} \end{cases}$$

Fort, *PNAS* (2012)

44