Supplementary Material

Supplementary Methods

Inference of the Yoruba demography with ∂a∂i:

Demographic function used of the *Linear* model:

```
def linear_growth(params, n1, pts):
    T = params
    xx = dadi.Numerics.default_grid(pts)
    phi = dadi.PhiManip.phi_1D(xx)
    nu = 1e-9 #fixed initial population size
    nu_func = lambda t: nu + ( ( 1.0 - nu ) * t ) / T
    phi = dadi.Integration.one_pop(phi, xx, T, nu=nu_func)
    sfs = dadi.Spectrum.from_phi(phi, ns, (xx,))
    return sfs}
```

Demographic function used of the *Exponential* model:

```
def exponential_growth(params, n1, pts):
    T = params
    xx = dadi.Numerics.default_grid(pts)
    phi = dadi.PhiManip.phi_1D(xx)
    nu = 1e-1 #fixed initial population size
    nu_func = lambda t: nu * ( ( 1.0 / nu ) ** ( t  / T ) )
    phi = dadi.Integration.one_pop(phi, xx, T, nu=nu_func)
    sfs = dadi.Spectrum.from_phi(phi, ns, (xx,))
    return sfs
```

Demographic function used of the *Sudden* model:

```
def sudden_growth(params, n1, pts):
    T = params
    xx = dadi.Numerics.default_grid(pts)
    phi = dadi.PhiManip.phi_1D(xx)
    Tb = 1.0 #fixed time before growth event
    nu = 1e-2 #fixed population size before growth event
```

```
nuF = 1.0 #fixed population size after growth event
phi = dadi.Integration.one_pop(phi, xx, Tb, nu=nu)
phi = dadi.Integration.one_pop(phi, xx, T, nu=nuF)
sfs = dadi.Spectrum.from_phi(phi, ns, (xx,))
return sfs
```

Interval for the parameter to optimize and initial value for optimization:

- *Linear*: $T \in [0, 10]$ and $T_0 = 3$
- *Exponential*: $T \in [0, 25]$ and $T_0 = 5$
- *Sudden*: $T \in [0, 10]$ and $T_0 = 1$

For the three models, the grid point settings for the extrapolation is `pts_l=[300,400,500]`. The function used for the optimization is `optimize_log` with `maxiter=3`. The script for each demographic model was run 100 times, and we kept the parameter value with the best maximum log composite likelihood. For the exponential, to retrieve the rate of the exponential growth as we had parametrized it in our model, we compute

$$\tau = -\frac{T_{opt}}{\ln(nu)}$$

where T_{opt} is the optimized parameter value and $nu = 0.1$ (see demographic function above).

Fixation trajectories for the models with implicit demography

For the *Conditioned* model, we use the Wright-Fisher diffusion conditioned upon fixation (Lambert, 2008) to simulate trajectories of fixation:

$$dX_t = (1 - X_t)dt + \sqrt{X_t(1 - X_t)}\, dB_t$$

where X_t is the random variable accounting for the frequency of the allele at time t and B_t is Brownian motion. We simulate the trajectories starting at $X_0 = 0$ with $dt = 0.0001$ and we stop the trajectories when X_t reaches 1. To account for the specificity of the *Conditioned* model, we keep only trajectories that reach fixation in a time smaller than the optimized parameter value $\widehat{\tau}$. Similarly, for the Birth-Death model, we use the critical Feller diffusion (Lambert, 2008):

$$dX_t = \sqrt{2X_t}dB_t$$

and we run trajectories until time reaches the optimized parameter value $\widehat{\tau}$. We keep trajectories for which $X_{\widehat{\tau}} \in (U_n, U_{n+1})$, where $U_k = \sum_{i=1}^{k} V_i$ and the V_i's are independent exponential random variables with mean $1/n$. This procedure amounts to conditioning upon sampling n individuals at time $\widehat{\tau}$. Indeed, for mathematical reasons, the standard way of sampling in a branching population is not to fix the sample size, but to sample each individual in the population independently with the same probability p. Assuming that individuals are linearly ordered, the number W of individuals between two consecutively sampled individuals then follows a geometric law of parameter p. In the model used in the paper and

in Delaporte *et al.* (2016), we further condition on the sample size n with the relation $p = n/N$. So if we measure W in units of N individuals, we are left with $V = \frac{p}{n}W$. Now as $p \to 0$ (sparse sampling), V converges to an exponential random variable of parameter n. Thus, the individuals sampled in the population are separated by exponential random variables of parameter n, and can thus be represented by the points $(U_i)_{i\geq 1}$. Therefore, sampling n individuals is equivalent to keeping trajectories for which $X_{\widehat{\tau}} \in (U_n, U_{n+1})$.

For both models, we average over 5 000 trajectories.

References

Delaporte, C., G. Achaz, and A. Lambert, 2016 Mutational pattern of a sample from a critical branching population. Journal of mathematical biology**:** 1–38.

Lambert, A., 2008 Population Dynamics and Random Genealogies. Stochastic Models *24*(sup1)**:** 45–163.

Supplementary Figures

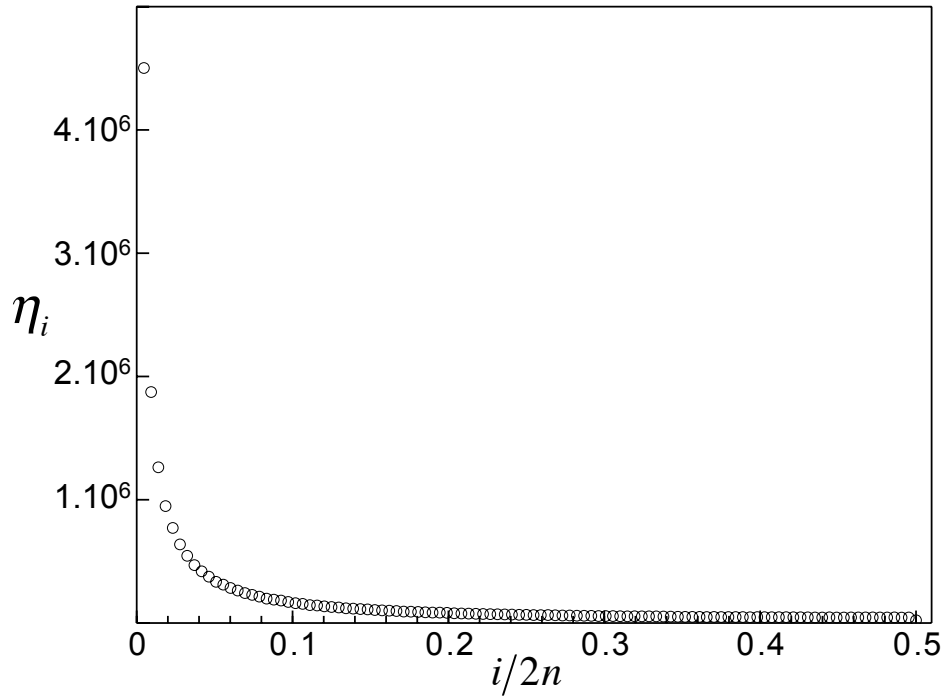

Figure S1: Yoruba Site Frequency Spectrum. The SFS is folded. The total number of sites in the SFS is $S = 20\,417\,698$.

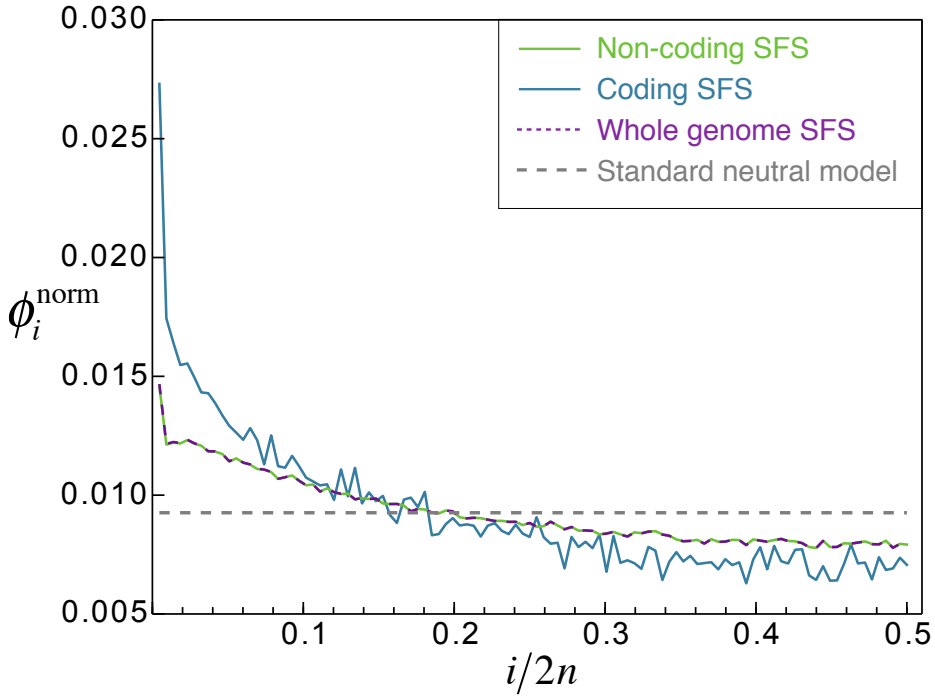

Figure S2: Coding and non-coding Yoruba SFS. In blue, SFS for coding parts of the genome. In green, SFS for the non-coding parts of the genome. The dashed purple line is the whole-genome SFS. The grey dashed line is the expected SFS under the standard neutral model without demography. The SFS are folded, transformed and normalized (see Methods).

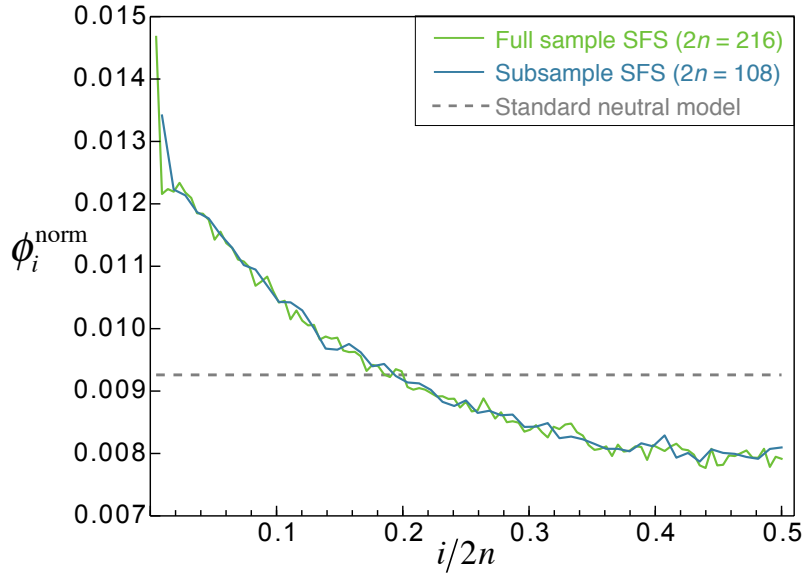

Figure S3: Subsample SFS of the Yoruba population. The green line is the SFS of the whole sample ($2n = 216$). The blue line is the SFS of a subsample containing half of the Yoruba individuals ($2n = 108$). The grey dashed line is the expected SFS under the standard neutral model without demography (with $2n = 216$). The SFS are folded, transformed and normalized (see Methods). For comparison, the subsample SFS was divided by 2 after normalization because it contains half as many values as the two other SFS.

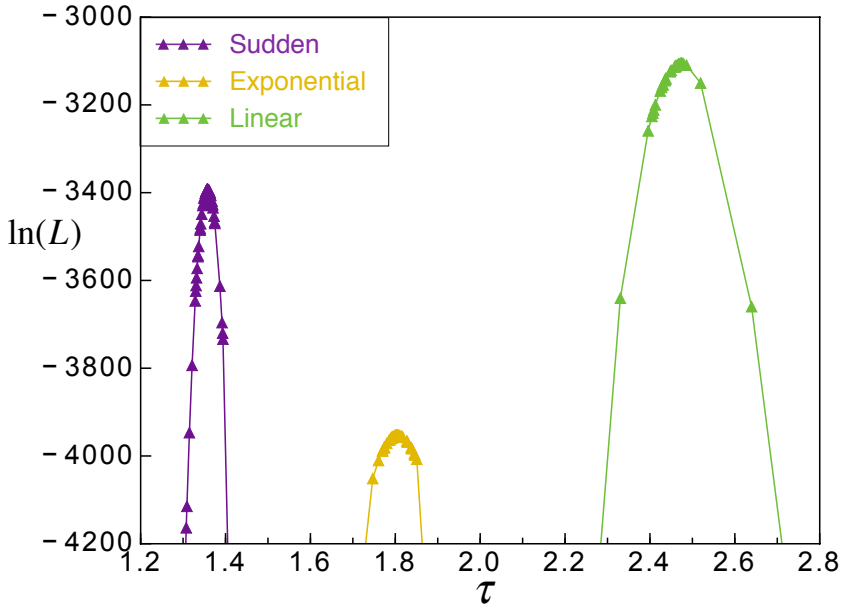

Figure S4: Maximum log composite likelihood obtained by the ∂a∂i method for the *Sudden*, *Exponential* and *Linear* models. We ran the method 100 times for each model. For each run, we report the maximum log composite likelihood with the corresponding τ value. The figure is zoomed on the best likelihood values (higher than -4 200). The number of points present in the plot (with log composite likelihood higher than -4 200) is 75 for the *Sudden* model, 93 for the *Linear* model and 95 for the *Exponential* model.