



HAL
open science

Cultural transmission of reproductive success impacts genomic diversity, coalescent tree topologies and demographic inferences

Jérémy Guez, Guillaume Achaz, François Bienvenu, Jean Cury, Bruno Toupance, Évelyne Heyer, Flora Jay, Frédéric Austerlitz

► **To cite this version:**

Jérémy Guez, Guillaume Achaz, François Bienvenu, Jean Cury, Bruno Toupance, et al.. Cultural transmission of reproductive success impacts genomic diversity, coalescent tree topologies and demographic inferences. *Genetics*, 2023, 223 (4), pp.iyad007. 10.1093/genetics/iyad007 . hal-03875721

HAL Id: hal-03875721

<https://hal.science/hal-03875721v1>

Submitted on 17 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cultural transmission of reproductive success impacts genomic diversity, coalescent tree topologies and demographic inferences

Jérémy Guez^{1,2,*}, Guillaume Achaz^{1,3}, François Bienvenu⁴, Jean Cury⁵, Bruno Toupance¹, Évelyne Heyer¹, Flora Jay^{2,‡}, and Frédéric Austerlitz^{1,‡}

¹UMR 7206 Eco-Anthropologie, CNRS, MNHN, Université Paris Cité, 75116, Paris, France

²Université Paris-Saclay, CNRS, INRIA, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400, Orsay, France.

³SMILE group, Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS UMR 7241, INSERM U 1050, 75005, Paris, France

⁴Institute for Theoretical Studies, ETH Zürich, 8092, Zürich, Switzerland

⁵SEED, U1284, INSERM, Université Paris Cité, 75004, Paris, France

[‡]These authors contributed equally to this work.

*Corresponding author: jeremy.guez@outlook.fr

Abstract

Cultural transmission of reproductive success (CTRS) has been observed in many human populations as well as other animals. CTRS consists of a positive correlation of nongenetic origin between the progeny size of parents and children. This correlation can result from various factors, such as the social influence of parents on their children, the increase of children’s survival through allocare from uncles and aunts, or the transmission of resources. Here, we study the evolution of genomic diversity over time under CTRS. CTRS has a threefold impact on population genetics: (1) the effective population size decreases when CTRS starts, mimicking a population contraction, and increases back to its original value when CTRS stops; (2) coalescent tree topologies are distorted under CTRS, with higher imbalance and a higher number of polytomies; and (3) branch lengths are reduced nonhomogenously, with a higher impact on older branches. Under long-lasting CTRS, the effective population size stabilizes but the distortion of tree topology and the nonhomogenous branch length reduction remain, yielding U-shaped site frequency spectra (SFS) under a constant population size. We show that this yields a bias in SFS-based demographic inference. Considering that CTRS was detected in numerous human and animal populations worldwide, one should be cautious because inferring population past histories from genomic data can be biased by this cultural process.

population genetics; evolution; cultural process; demographic inference; genetic diversity; coalescent tree shape; imbalanced topology

1 Introduction

In recent years, numerous studies have investigated the interactions between human culture and genetics. In some cases, cultural changes yield genetic adaptations. This was the case, for example, for lactase persistence that likely evolved independently in different human populations in Eurasia and Africa, due to the emergence of pastoralism (Swallow, 2003; Bersaglieri *et al.*, 2004; Tishkoff *et al.*, 2007; Gerbault *et al.*, 2011; Segurel *et al.*, 2020). Nevertheless, cultural processes can affect human genetic evolution without involving natural selection (Heyer *et al.*, 2012): (i) polygamy (including polyandry and polygyny), (ii) descent rules (patrilineal, matrilineal, or cognatic), and (iii) cultural transmission of reproductive success (CTRS).

CTRS is a positive correlation in the number of children between parents and children resulting from nongenetic causes. In that case, individuals with many siblings tend to have more children

12 than average. This transmission can result from multiple nongenetic causes: the social influence of
13 parents on their children (Barber, 2001; de Valk, 2013; Kolk, 2014), the increase in child survival when
14 uncles and aunts are present (allocare) (Heyer *et al.*, 2012; Lawson and Mace, 2011; Murphy, 2013)
15 or the transmission of resources from parents to children. Such resources can be material resources
16 (Sorokowski *et al.*, 2013), social resources (e.g., transmission of rank or of polygyny; Heyer *et al.*,
17 2012), or cultural resources (such as hunting skills; Mulder *et al.*, 2009). Furthermore, transmission
18 of migration propensity across generations can have an effect similar to CTRS, with some lineages
19 growing less than others due to their larger tendency to leave the population (Gagnon and Heyer,
20 2001; Gagnon *et al.*, 2006).

21 CTRS yields a decrease in effective population size and genetic diversity, and may increase the
22 frequency of severe genetic disorders (Austerlitz and Heyer, 1998). The time to the most recent
23 common ancestor is reduced, yet in a nonhomogenous way as the tree branches closer to the root are
24 more strongly shortened (Sibert *et al.*, 2002). While these patterns can result from other evolutionary
25 processes (e.g. bottlenecks, expansions), a more specific effect of CTRS is its impact on the topology of
26 coalescent trees: CTRS yields imbalanced trees as it increases the proportion of lineages corresponding
27 to large families (Sibert *et al.*, 2002). This specific property has been used in particular for inferring
28 the transmission of reproductive success (TRS) on Y chromosome and mitochondrial DNA (Blum
29 *et al.*, 2006; Heyer *et al.*, 2015). Since natural selection also implies a TRS, it is difficult to assess
30 whether the imbalanced trees of nonrecombining uniparental markers result from natural selection or
31 CTRS. Therefore, it is important to study the impact of CTRS on the nuclear genome. Recombination
32 should indeed restrict the effects of natural selection to the genomic regions around selected loci (Li
33 and Wiehe, 2013). Conversely, CTRS will yield an imbalance signal across the whole genome because
34 in that case reproductive success is not linked to any locus in particular.

35 Studying the impact of CTRS on genomic diversity is particularly relevant, as it is a rather common
36 phenomenon. Several demographic studies have shown a parents-children correlation in the number of
37 children ranging generally between 0.1 and 0.25 (e.g., Murphy, 1999; Murphy and Wang, 2001; Gagnon
38 and Heyer, 2001; Pluzhnikov *et al.*, 2007). There has been an extensive debate about whether these
39 correlations result from cultural (Potter and Kantner, 1955; Duncan *et al.*, 1965) or genetic (Kohler
40 *et al.*, 1999; Rodgers *et al.*, 2001; Mills and Troup, 2015) transmission, the second case corresponding
41 to natural selection. The correlations may, in fact, often be caused by both genetic and cultural
42 transmission, along with interactions between genetics and the environment (Murphy, 2013), making
43 the disentangling of those processes particularly difficult, especially as they can vary across populations
44 and time. For instance, contemporary populations tend to have a stronger intergenerational correlation
45 than populations that predate the demographic transition (Murphy, 1999; Murphy and Wang, 2001).
46 Furthermore, this phenomenon is not limited to humans and has been described in various species such
47 as hyenas (Engh *et al.*, 2000), Japanese macaques (Kawai, 1958), whales (Whitehead, 1998), dolphins
48 (Frere *et al.*, 2010), and cheetahs (Kelly, 2001).

49 Another reason for studying the impact of CTRS on genomic diversity lies in its putative ability
50 to impact summary statistics commonly used to infer other processes. For instance, site frequency
51 spectra (SFS), which might be impacted by CTRS, are widely used for demographic inferences, either
52 alone (e.g. $\delta a \delta i$ (Gutenkunst *et al.*, 2009), Fastsimcoal (Excoffier *et al.*, 2013), Stairway Plot (Liu and
53 Fu, 2020), ABC-DL (Mondal *et al.*, 2019)) or jointly with other summary statistics (e.g., Sheehan and
54 Song, 2016; Boitard *et al.*, 2016; Jay *et al.*, 2019; Terhorst *et al.*, 2017). These inference tools could
55 thus be biased when applied to populations that have been affected by CTRS during part of their
56 history. Understanding the interactions between CTRS and demographic changes is therefore relevant
57 not only for inferring CTRS itself but also for improving demographic inferences, which is of broad
58 interest (Beichman *et al.*, 2018).

59 This article pursues three objectives. First, we aim to improve our understanding of the impact
60 of CTRS on nuclear genomes using simulations. Brandenburg *et al.* (2012) performed a simulation
61 study that investigated the impact of CTRS on small sequences, ignoring intragenic recombination.
62 Here, we study its impact on large recombining sequence data, adding numerous summary statistics
63 not previously explored in CTRS scenarios. The summary statistics we assess are mainly of two kinds:
64 (i) population genomic statistics, such as genetic diversity, Tajima's D and SFS, and (ii) various tree
65 topology indices, such as tree imbalance indices and number of polytomies. In addition, we investigate
66 the interaction of demographic changes and CTRS, as we expect human populations to undergo both
67 types of processes. In particular, we look into the effect of an expansion occurring before and during

68 CTRS, an interaction that has not yet been explored. Second, we investigate the impact of CTRS
69 duration and the persistence of ancient CTRS signals in the genome by measuring the evolution of
70 the summary statistics over time (before, during, and after CTRS). In particular, this allows us to
71 assess the impact of very short periods of CTRS on population genetics. Although long-lasting CTRS
72 is not theoretically excluded, available anthropological evidence only indicates the presence of CTRS
73 over short periods. For example, pedigrees from the Saguenay-Lac-Saint-Jean population show CTRS
74 for 12 generations (Austerlitz and Heyer, 1998). For CTRS induced by variance in fertility among
75 lineages within a population, the persistence of CTRS requires that individuals can trace back their
76 lineage affiliation for several generations (in central Asia, Chaix *et al.* (2004) estimated this number
77 of generations to be 7–10 depending on the population). Finally, we assess whether CTRS impacts
78 demographic inference. For various CTRS scenarios, we compare the true and estimated instantaneous
79 growth factor and timing of expansion.

80 2 Methods

81 2.1 Model

82 We implemented the CTRS model designed by Sibert *et al.* (2002) and Brandenburg *et al.* (2012) using
83 the forward-in-time simulation framework SLiM (Haller and Messer, 2019). Individuals are diploid and
84 monogamous, generations are nonoverlapping, and the population has a fixed number of individuals N
85 with a 1:1 sex-ratio. At each generation, couples are formed uniformly at random before reproduction
86 and never separated. One parental couple is randomly drawn from the population for each newborn
87 child. This process is repeated until N offspring are produced. The probability p_i for a given couple i
88 of being drawn for reproduction is given by:

$$p_i = \frac{\gamma_i(b) \times s_i^\alpha}{\sum_{j=1}^{N_c} \gamma_j(b) \times s_j^\alpha},$$

89 where s_i is the average sibship size of the two members of couple i , α is the parameter controlling
90 the intensity of CTRS and b is the parameter controlling the variance in reproductive success. We
91 denote N_c as the number of couples ($N_c = N/2$). The higher α is, the stronger the CTRS ($\alpha = 0$
92 means no CTRS, $\alpha = 2$ means a very strong CTRS). $\gamma_i(b)$ is a random gamma distributed variable
93 drawn independently for each couple i , with shape parameter b and mean 1. Here, we considered only
94 two cases: $b \rightarrow \infty$ (low variance in reproductive success, resulting in a Poisson-like distribution for
95 the progeny size in the absence of CTRS, as $\lim_{b \rightarrow \infty} \gamma(b) = 1$) or $b = 1$ (high variance, resulting in a
96 geometric-like distribution, as $\gamma(1)$ is an exponential of mean 1 distribution). Some results are shown
97 for both values of b , but we focused mainly on the $b = 1$ case, as Austerlitz and Heyer (1998) found
98 that the geometric-like model was more consistent with demographic data than the Poisson-like model
99 and better explained the occurrence of genetic diseases in Saguenay-Lac-Saint-Jean.

100 For the demographic parameters, we compared two scenarios of constant population sizes (200 and
101 5000 individuals) and explored a scenario of sudden demographic expansion by a fivefold factor (200
102 to 1000 individuals). This expansion occurred 300 generations before the present.

103 2.2 Simulations

104 Unless specified otherwise, the simulations correspond to 200 replicates per scenario, a population size
105 of 1000 individuals and a sample size of 30 individuals. Genomes were made of one chromosome of 10^7
106 bp in length, with a recombination rate and mutation rate of 10^{-8} per bp, which are commonly used
107 parameters in human population modeling. We used the geometric-like model ($b = 1$) since Austerlitz
108 and Heyer (1998) showed it was more realistic than the Poisson-like model ($b = \infty$) in the population
109 of Saguenay-Lac-Saint-Jean where CTRS is documented from pedigree datasets. Coalescent trees are
110 built in two steps: (1) forward-in-time simulations using our model implemented in SLiM (Haller and
111 Messer, 2019) starting before the beginning of CTRS, resulting in trees that did not fully coalesce
112 when the CTRS period is short, (2) a backward neutral coalescent process in order to complete the
113 trees from the first step (i.e., to reach the most recent common ancestors throughout the genome).
114 This step uses the *tskit* package functionality called *recapitation* (Kelleher *et al.*, 2016, 2019).

115 To assess the impact of CTRS on reproduction, we measured three demographic parameters : (1)
 116 the correlation between progeny sizes of all individuals and their parents' progeny sizes as a function
 117 of α , the strength of CTRS; (2) the variance of progeny size, and (3) the distribution of progeny sizes
 118 in the population for $\alpha = 0, 1$ and 2 .

119 To investigate the effect of CTRS across time, we measured the genomic summary statistics on
 120 batches of individuals sampled through time for the following scenario: 2000 generations of CTRS,
 121 followed by 2000 generations with no CTRS. Every 50 generations, individuals were sampled for
 122 analysis. Following any cultural change (starting or stopping CTRS), we sampled more frequently
 123 to capture rapid fluctuations of summary statistics (at generations 2, 5, 10, 15, and 20 postchange).

124 2.3 Summary statistics

125 To assess the effects of CTRS on the genome, we explored the following diversity summary statistics
 126 as a function of time using the *tskit* package (Kelleher *et al.*, 2016, 2019): (1) the number of trees per
 127 chromosome, which is the number of recombination breakpoints plus 1, (2) the number of pairwise
 128 differences among the sampled chromosomes, (3) the average number of pairwise differences per tree,
 129 and (4) the number of SNPs in the chromosomes, (5) the average number of SNPs per tree, (6)
 130 Tajima's D , (7) the unfolded site frequency spectrum (SFS). For the SFS, we computed a transformed
 131 version (Lapierre *et al.*, 2017) that consists of multiplying singletons by 1, doubletons by 2, and n -tons
 132 by n . We then divided all bins by θ , which is estimated by taking the average of all bins so that the
 133 expected transformed SFS for the neutral case is a flat line with a value of 1.

134 We computed the theoretical effective size N_{exp} according to the equation $N_{\text{exp}} = 4N/(2 + s^2)$,
 135 where s^2 is the variance in progeny size (Wright, 1938; Ewens, 2016). This formula computes the
 136 effective size as a function of the census population size N and the variance in progeny size only. We
 137 compared N_{exp} to the observed effective size N_{obs} which was computed as follows: $N_{\text{obs}} = \theta/(4\mu L)$,
 138 with the average number of pairwise differences, $\hat{\theta}_\pi$, as an estimator of θ , L the genome length and μ
 139 the mutation rate per base pair.

140 We also computed various topology indices, to assess the effect of CTRS on the topology of coa-
 141 lescent trees, with the help of the *tskit* package (Kelleher *et al.*, 2016, 2019). Balance and imbalance
 142 indices: (1) I_b , the Brandenburg imbalance index (Brandenburg *et al.*, 2012; Blum *et al.*, 2006); (2) I_s^* ,
 143 a normalized Sackin imbalance index (Sackin, 1972; Shao and Sokal, 1990); (3) I_{ce}^* and I_{ca}^* , two mod-
 144 ified versions of the Colless imbalance index (Colless, 1982), ; (4) the B_1 balance index (Shao and
 145 Sokal, 1990); (5) the B_2 balance index (Shao and Sokal, 1990; Bienvenu *et al.*, 2021). Other topology
 146 indices: (1) the number of polytomies (nodes that have more than two direct children); (2) the number
 147 of interior nodes (all nodes excluding leaves and root). To compare different indices, we also used their
 148 standardized versions using their mean and standard deviation at generations preceding CTRS.

149 I_b , I_s^* , I_{ca}^* and I_{ce}^* measure the imbalance of trees, meaning that those indices take higher values
 150 for more imbalanced trees. I_b was computed using the script provided by Brandenburg *et al.* (2012).
 151 For one tree, I_b is the average of $I_{b,\text{node}}$ computed for each node in the tree according to the formula:

$$I_{b,\text{node}} = \frac{B - m_{s,l}}{D - m_{s,l}}, \text{ with } m_{s,l} = 2B_{s,l,\text{coal}} - D,$$

152 where s is the number of direct subnodes under the considered node and l the number of leaves
 153 descending from it. For each direct subnode under the considered node, leaves are counted and the
 154 maximum value is denoted B . D is the maximum value that B can possibly take (i.e., in the most
 155 imbalanced configuration) and is equal to $l - s + 1$. Thus, $\frac{B}{D}$ is the level of imbalance at this specific
 156 node. The correction factor $m_{s,l}$ enforces the expectation of I_b to be 0.5 for a standard population
 157 without CTRS. This parameter is evaluated based on simulations: $B_{s,l,\text{coal}}$ is the average B value of
 158 1000 simulated random Kingman's (1982) incomplete coalescent trees with l leaves that were stopped
 159 when s parent nodes remained.

160 The Sackin imbalance index I_s is computed by counting for each leaf the number of nodes to reach
 161 the root and summing up all values. The Colless imbalance index I_c is computed by counting for each
 162 node (except for the root in our case) the difference in the number of leaves between its two children
 163 and summing up all values. However, this can be done only for binary trees. To handle polytomies, we
 164 designed two modified versions of the Colless imbalance index, I_{ce} and I_{ca} . For I_{ce} , the two children
 165 chosen for calculating the difference are those with the highest and lowest number of leaves (e , as for

166 extreme number of leaves). I_{ca} is computed by taking the average of differences for all pairs of children
167 among all children of a given node (a , as for average). Since the Sackin and Colless indices minimum
168 and maximum values depend on the number of nodes (Shao and Sokal, 1990) which varies across trees
169 when permitting polytomies, we computed a corrected version of the Sackin (I_s^*) and Colless (I_{ce}^* and
170 I_{ca}^*) indices which divides the index of each tree by the number of its interior nodes.

171 B_1 and B_2 are balance indices; we thus expect their value to be lower when trees are imbalanced.
172 The B_1 balance index is computed by counting for each node the maximum path length to its leaves
173 and taking the inverse of this value before summing up all of the values (one value per interior node).
174 The B_2 balance index is based on p_k the probabilities to reach the leaf k assuming a random walk
175 starting from the root and choosing a random direction at each node. B_2 is equal to the Shannon
176 entropy of the p_k ; a uniform distribution (an entropy of 1) corresponds to a balanced tree (Shao and
177 Sokal, 1990; Bienvenu *et al.*, 2021).

178 Because of recombination, one chromosome corresponds to a sequence of coalescent trees. Summary
179 statistics can be computed for each of the trees, with close trees having similar values. To consider
180 the various histories represented by each of those trees, we explored not only the average summary
181 statistics but also the shape of their distributions across the genome. The summary statistics were
182 computed separately on each tree along the genome using the *tskit* package.

183 We also assessed the effect of sample size (number of individuals sampled) and of number of genomic
184 regions on the power of detecting CTRS, using a Wilcoxon test with the significance threshold set to
185 0.01. For this assessment, we simulated 3,000 independent genomic regions of 1 Mb for two populations
186 of 1000 individuals: one that went through a CTRS process of strength $\alpha = 1$ during 20 generations
187 before present, and one with $\alpha = 0$ (no CTRS). We then sampled 5, 10, 30, 60, 90, and 120 diploid
188 individuals from each of the two sets of 3,000 simulated regions and the four summary statistics (I_b ,
189 number of polytomies, B_1 , and Tajima’s D) on all of them (2 scenarios \times 3,000 regions \times 6 sample sizes
190 \times 4 summary statistics computations). For each sample size, we sampled 3, 4, 5, . . . , 100 regions from
191 the two sets of 3,000 simulated regions, before using a Wilcoxon test to compare the four summary
192 statistics values between the two populations ($\alpha = 0$ and $\alpha = 1$). For each combination of sample
193 size and number of sampled replicates (6 \times 98 combinations), the sampling among replicates and the
194 Wilcoxon test were repeated 1000 times, with the proportion of P -values lower than or equal to 0.01
195 equaling the power of the test.

196 2.4 Assessing demography inference bias

197 To assess the bias in SFS-based demography inference, we used the software *$\delta a \delta i$* (Gutenkunst *et al.*,
198 2009) with a one-event model. Two scenarios were studied: (1) a sudden fivefold expansion in popu-
199 lation size that occurred 280 generations before a short period of CTRS (20 generations); and (2) a
200 sudden fivefold expansion in population size that occurred during CTRS, after the first 1200 gener-
201 ations of a 1500-generations period of CTRS (Figure 1). We chose a fivefold sudden expansion as a
202 simple illustration of a demographic event, which has the advantage of mimicking the past Neolithic
203 expansion in human population history. From 30 diploid individuals sampled 300 generations after the
204 demographic event, we inferred two parameters: the growth factor (expected value of 5) of the popula-
205 tion and the number of generations since the event (expected value of 300 generations). The strength
206 of CTRS was set to $\alpha = 1$. We compared the quality of inference in both scenarios to equivalent
207 demographic scenarios without CTRS ($\alpha = 0$).

208 We inferred the parameters of 200 replicates for each of the four scenarios (scenarios 1 and 2 with
209 $\alpha = 0$ or 1). Because the *$\delta a \delta i$* optimization algorithm depends on the initialization of the model
210 parameters, we repeated the inference three times for each replicate with different initialization values.
211 We set the boundaries for the inferred growth factor at [0.01; 100] and for the inferred growth time at
212 [0; 5] (time is expressed in $2N$ generations in *$\delta a \delta i$* , where N is the population size before the event).
213 When the results were too close to the boundaries (> 99 or $< 1/99$ for the growth factor, > 4.9 or
214 < 0.1 for the time since the event), the results were discarded. For each replicate, the remaining results
215 among the three trials were kept, and their median was considered as the inferred parameter for this
216 replicate. To convert time into generations, we multiplied the inferred time value of each replicate r
217 by $2\hat{N}_r$; where \hat{N}_r denotes the ancestral population size estimated for replicate r , using a $\hat{\theta}_r$ estimate
218 computed by *$\delta a \delta i$* .

219 We removed outliers among replicates (i.e., values that were higher than $Q3 + 1.5 \times IQR$ and lower
220 than $Q1 - 1.5 \times IQR$, with $Q3$ being the third quartile, $Q1$ being the first quartile and IQR being the

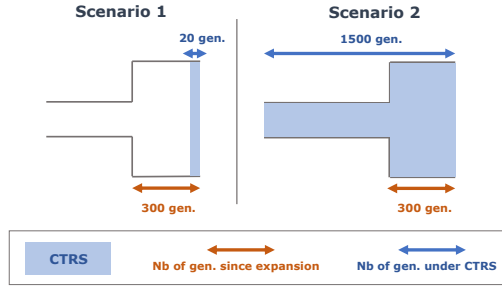


Figure 1: The two studied scenarios for SFS computation and $\delta a \delta i$ inference. In both scenarios, the expansion event occurs 300 generations before SFS computation and $\delta a \delta i$ inference. Scenario 1: 20 generations of CTRS before the present. Scenario 2: 1500 generations of CTRS before present.

interquartile range). We then computed the mean squared relative error (MSRE) and relative bias.

3 Results and discussion

3.1 Impact of CTRS on reproductive patterns

To assess the impact of CTRS on reproductive patterns, we simulated various strengths of CTRS (defined by α) for two models of variance in reproductive success (low variance with $b = \infty$ and high variance with $b = 1$). We computed the Pearson correlation coefficient between parents and children $\text{Cor}_{P,C}$ and the variance and distribution of progeny size. As expected, $\text{Cor}_{P,C}$ increases with α . However, this effect is weaker for smaller population sizes. This is due to an increased effect of stochastic processes in small populations, counteracting the impact of parents on children's progeny size (Figure 2a). The slope of the relationship between $\text{Cor}_{P,C}$ and α is also lower for the $b = 1$ model than for the $b = \infty$ model (Figure 2a). Indeed, the higher variance in progeny size in the $b = 1$ model decreases the correlations, compared with the $b = \infty$ model.

Higher values of α yield more extreme progeny sizes (Figure 2b-C, purple compared with orange and green) and a higher variance (Supp. Fig. S1). This variance reaches a plateau after a few generations (Supp. Fig. S1). At this plateau, the exact progeny size distribution differs depending on the model: compared with the $b = \infty$ model, the $b = 1$ model yields a higher proportion of couples with no offspring and a lower proportion of couples with medium-sized families (1–3 children) (Figure 2b versus 2c).

3.2 Impact of CTRS on the genome

3.2.1 Effective population size

We then assessed the impact of CTRS on population genomic parameters. When CTRS begins, genomic diversity, measured either as the number of SNPs (Supp. Fig. S2a) or as the number of pairwise differences (Fig. 3a), declines and eventually reaches a plateau, showing a decrease in effective population size of 40% for the $b = \infty$ model and of 75% for the for the $b = 1$ model (for $\alpha = 1$, at the plateau), demonstrating a stronger effect of CTRS under the second model (Fig. 3b).

Because of this decrease in effective population size, the number of coalescent trees across the genome is lower due to fewer recombination events, and the TMRCA is smaller (Supp. Fig. S2b-C). For all these parameters, the plateau is lower for $\alpha = 2$, since it yields lower effective population sizes than $\alpha = 1$. Moreover, the higher α is, the faster the plateau is reached. This happens because genetic drift, which is stronger when α is high, swiftly erases past diversity. As soon as CTRS stops, diversity starts to increase slowly (Figure 3a), taking more time to recover than it took to decrease. Indeed, as the effective population size becomes larger, drift becomes weaker and the impact of past events lasts longer (i.e., diversity is close to equilibrium after $10N_e$ generations).

This decrease in effective population size results both from the increase in the variance of progeny size due to CTRS and the transmission of progeny size itself, which amplifies allele fixations by helping alleles carried by large lineages to spread faster in the population. To assess the respective impact of these two factors on effective population size, we compared N_{exp} (the expected effective population

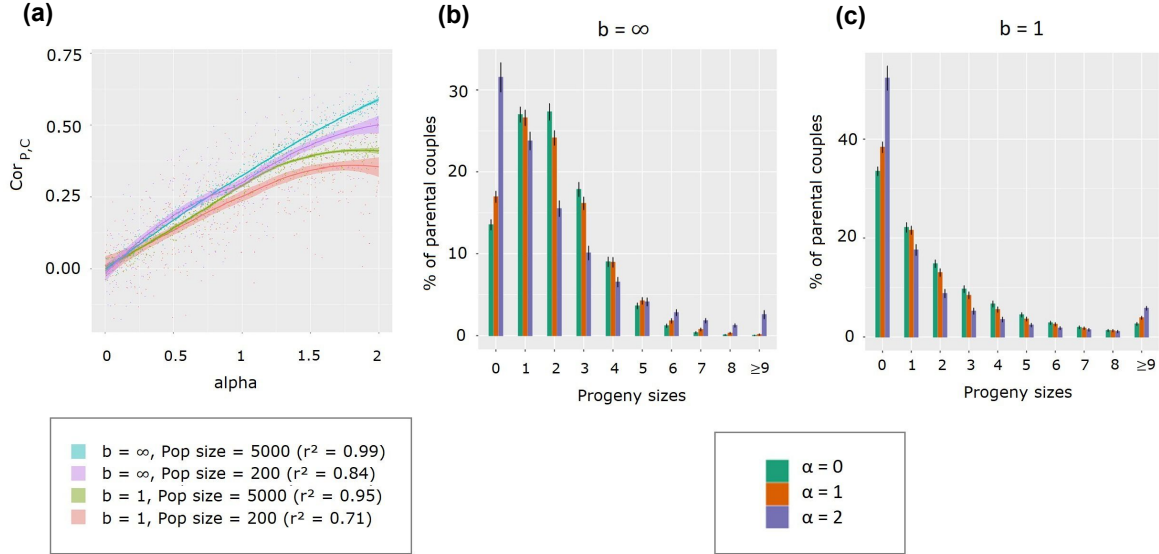


Figure 2: Impact of CTRS on two population reproduction variables. (a) Correlation between parents and children progeny size as a function of α , for four scenarios. In brackets: correlation between $Cor_{P,C}$ and α for each scenario. Lines are drawn using locally weighted regression with the 95% confidence interval using the function `loess` of the R package `ggplot2`. (b) Distribution of progeny sizes for $\alpha = 0$ (green), 1 (orange) and 2 (purple), population size = 1000. The $b = \infty$ model is used (low variance of reproductive success). (c) Distribution of progeny sizes for $\alpha = 0$ (green), 1 (orange) and 2 (purple), population size = 1000. The $b = 1$ model is used (low variance of reproductive success).

size when taking into account the variance in progeny size only), to N_{obs} which is impacted by both components (Fig. 3b). We show that while a substantial decrease in effective population size is caused by the increased variance in progeny size, most of this decrease is due to the transmission component (around 70% of the decrease in the $b = \infty$ model and 65% of the decrease in the $b = 1$ model, for $\alpha = 1$).

3.2.2 Tajima's D

Tajima's D follows a more complex pattern than does genetic diversity. This pattern can be decomposed into four steps (Figure 4a): (1) as soon as CTRS begins, it increases rapidly towards a peak in positive values then (2) it decreases toward a plateau in negative values, (3) when CTRS stops, it rapidly decreases again toward more negative values, and (4) it slowly recovers to pre-CTRS levels. The first peak (1) results from a sudden decrease in effective population size when CTRS starts, as explained above, yielding a demographic contraction-like signal with positive values of D . Once this contraction signal is erased (i.e., the effective population size is still lower but there is no "memory" of the ancient effective population size due to an MRCA born after the change), D reaches a negative plateau at equilibrium; (2): the population is composed of many related individuals coming from large family lineages and few individuals from small family lineages, the latter yielding an excess of rare alleles. The nonhomogenous reduction of coalescent times, stronger for the branches closer to the root (Sibert *et al.*, 2002), also contributes to this excess of rare alleles. When CTRS stops, the decrease toward more negative values (3) is due to the increase in effective population size (expansion-like event). This negative peak is followed by a slow recovery (4) until the expansion signal is completely erased. These steps are not followed at the same pace along the genome: some coalescent trees will enter the equilibrium stage, while others retain a strong signal of the effective population size contraction, transiently yielding a bimodal distribution of D across the genome (Supp. Fig. S3b and C for $\alpha = 2$, Figure S3d for $\alpha = 1$).

Thus, understanding the effect of CTRS on Tajima's D requires accounting for three processes: changes in effective population size, an increased variance in relatedness among individuals as compared

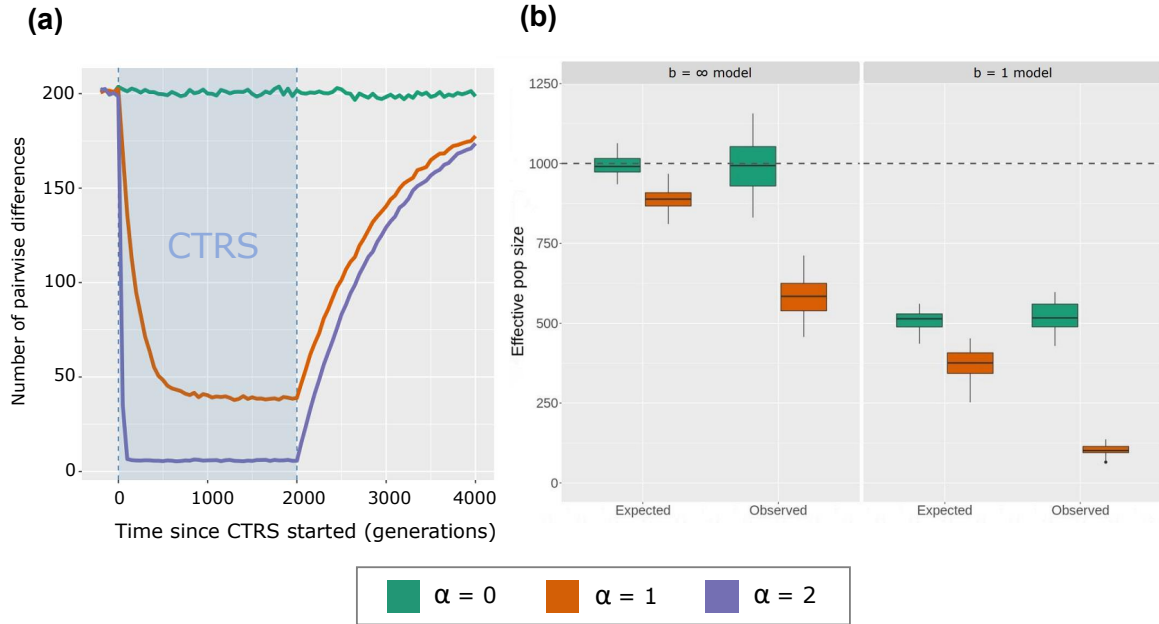


Figure 3: Factors of effective population size decrease under CTRS. (a) Average number of pairwise differences across time for three levels of CTRS: $\alpha = 0$, $\alpha = 1$ and $\alpha = 2$. In all cases, the $b = 1$ model of variance in progeny size is used. The blue rectangle corresponds to the period when populations are under CTRS. Generations are counted from the beginning of CTRS. (b) Expected effective population size given the observed offspring variance (N_{exp}) and observed effective population size measured using the number of pairwise differences at the plateau in Panel a as an estimator of θ (N_{obs}), for $\alpha = 0$ and $\alpha = 1$ and both models of variance in progeny size ($b = \infty$ and $b = 1$). The dotted line represents the census N value, which is 1000 individuals.

283 with a neutral population and a non homogeneous reduction in branch lengths. Timing is then an
 284 important factor: the relationship between α and Tajima’s D changes over time after the beginning
 285 of CTRS, and the impact of CTRS on genetic diversity and D persists long after CTRS has stopped.

286 The interaction between demographic events and CTRS is also important, since both can happen
 287 in the same period of human history. When a fivefold expansion occurs during the equilibrium stage,
 288 Tajima’s D decreases as expected, but the extent of this decrease depends on α : the stronger α is, the
 289 weaker the decrease will be, showing the nonadditivity of the two processes regarding D (Figure 4b,
 290 generation 1200). The recovery from the effect of this fivefold expansion also depends on α : when $\alpha =$
 291 1, Tajima’s D recovers faster than with no CTRS ($\alpha = 0$) (Figure 4b, generations 1,200 – 1,500). This
 292 is due to the smaller population effective size when $\alpha = 1$, which quickly erases past signals. Thus, we
 293 expect populations under CTRS to lose the genetic signals of past demographic events faster.

294 3.2.3 Coalescent tree topology

295 It is likely that neither diversity indices nor Tajima’s D would be sufficient alone to infer CTRS in
 296 population genetics data, since demographic events also impact these statistics. In contrast, the shape
 297 of coalescent trees has been shown to display a CTRS-specific signal, with trees being more imbalanced
 298 only when CTRS is present, irrespective of the variation in total population size. Brandenburg *et al.*’s
 299 (2012) imbalance index I_b (Figure 5a) grows rapidly when CTRS starts and decreases as soon as it
 300 stops, recovering in a few dozens of generations, unlike Tajima’s D (Figure 4a), which did not fully
 301 recover after $2N = 2000$ generations. The number of polytomies follows a pattern similar across time
 302 as I_b (Supp. Fig. S4). However, this increased number of polytomies can stem from the contraction in
 303 effective size yielded by CTRS (4-fold decrease when $\alpha = 1$ and $b = 1$), as coalescent rates are higher
 304 for smaller population sizes, increasing the probabilities of polytomies. To assess this hypothesis, we
 305 compared the number of polytomies after 500 generations of CTRS ($\alpha = 1$ and $b = 1$) to the number
 306 of polytomies after a 4-fold contraction 500 generations before the present, without CTRS. The results

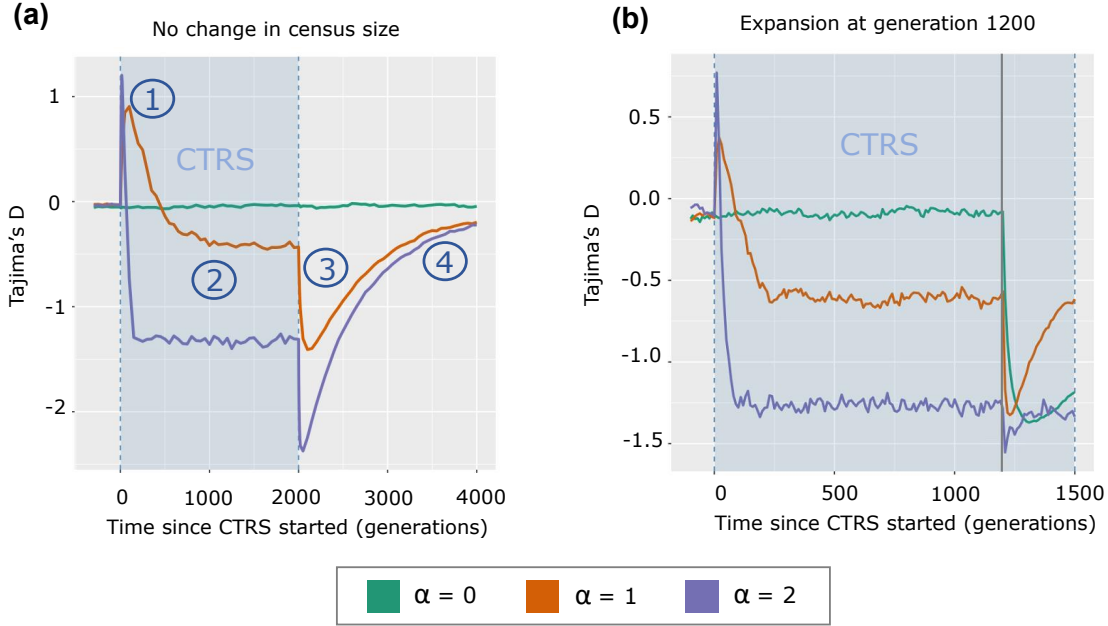


Figure 4: Tajima’s D through time under various CTRS and demographic conditions. (a-b) The blue rectangle corresponds to the period when populations are under CTRS. Generations are counted from the beginning of CTRS. In all cases, the $b = 1$ model of variance in progeny size is used. (a) Tajima’s D across generations for three values of α (0, 1, and 2), with a constant population size of 1000 individuals. (b) Tajima’s D across generations for three values of α (0, 1, and 2). A fivefold expansion event occurs at generation 1200 (200 individuals to 1000 individuals — gray vertical line).

307 show that the 4-fold contraction indeed yields a higher number of polytomies than the neutral case,
 308 but a lower number of polytomies compared with the scenario of CTRS (Supp. Fig. S5a). Thus, the
 309 increased number of polytomies under CTRS is caused not only by the contraction of the effective
 310 size, but also by the transmission property of CTRS. The same comparison for I_b shows that none
 311 of the imbalance under CTRS is due to the contraction of effective size, as the mean imbalance after
 312 contraction is equal to the mean imbalance of the neutral case, with a higher variance due to the
 313 smaller population size (Supp. Fig. S5b).

314 The distribution of I_b across the genome was bell-shaped and unimodal for all tested strengths of
 315 CTRS ($\alpha = 0, 1,$ and 2), with a shift toward high values when α increased (Supp. Fig. S6). This is
 316 because CTRS is not conveyed by any locus in particular, unlike natural selection, for which we could
 317 expect in some cases a multimodal distribution due to imbalanced trees in the region under selection
 318 and balanced trees elsewhere in the genome. Unlike the distribution of Tajima’s D (Supp. Fig. S3),
 319 the distribution of I_b does not evolve during the process of CTRS, as shown when comparing the
 320 distributions after 20 and 500 generations of CTRS (Supp. Fig. S6). In fact, I_b is only impacted by the
 321 imbalance property of coalescent trees and thus only displays its effects, which are constant through
 322 time after the first few generations, contrary to Tajima’s D , which is affected by imbalance and by
 323 changes in effective size as well, with the latter’s effects depending strongly on time.

3.2.4 Short-lasting CTRS

324 We have thus far simulated cases of long-lasting CTRS, in order to investigate the values of the different
 325 statistics at the equilibrium state under CTRS (Figure 4). However, as the CTRS duration could be
 326 much shorter in reality, we also investigated cases where CTRS lasted for only a few generations. This
 327 situation was simulated for both low ($b = \infty$) and high variance in progeny-size ($b = 1$). We show that
 328 two or three generations of CTRS are sufficient to have an impact on genetic statistics (Supp. Fig. S7).
 329 Tajima’s D displays an effect under medium ($\alpha = 1$) and high levels of CTRS ($\alpha = 2$), for both models
 330 of variance in progeny-size ($b = \infty$ and $b = 1$). Conversely, I_b seems affected under medium levels
 331 of CTRS only in the case of high variance in progeny size. Note that these realistic short periods of
 332

333 CTRS lead to an increase in Tajima’s D toward positive values due to the effective size contraction, as
 334 explained above. Finally, we show that after such a short period of CTRS, a few generations without
 335 CTRS are not sufficient to erase the effects on the genome (Supp. Fig. S7).

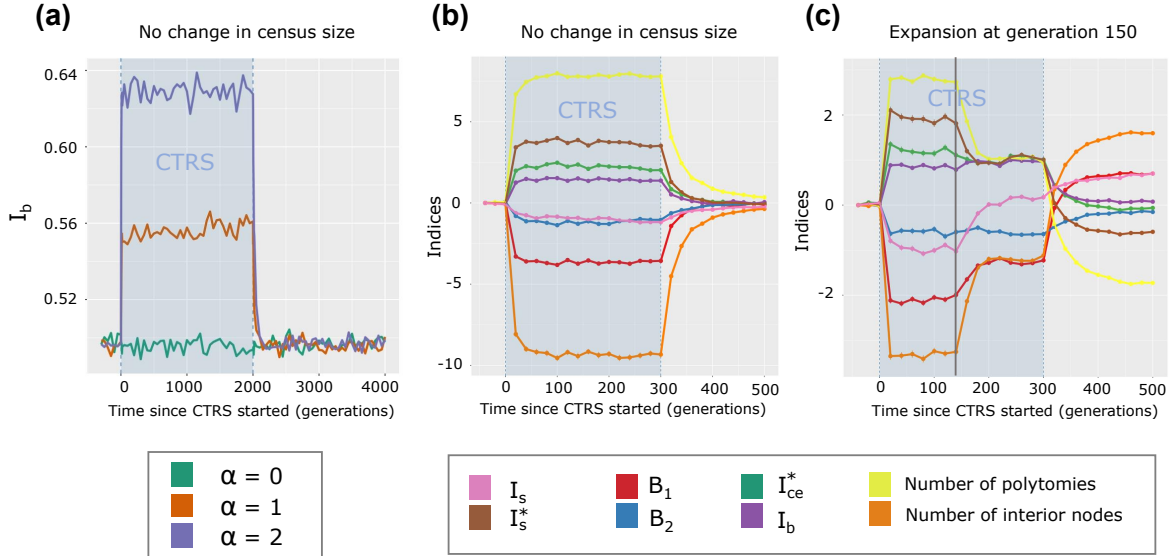


Figure 5: Imbalance indices over time. (a-c) The blue rectangle corresponds to the period when populations are under CTRS. Generations are counted from the beginning of CTRS. In all cases, the $b = 1$ model of variance in progeny size is used. (a) I_b across generations for three values of α (0, 1, and 2). (b) Various indices across generations for $\alpha = 1$. For each point, bars show the standard error of the mean. (c) Various indices across generations for $\alpha = 1$. An expansion event occurs at generation 150 (vertical gray line). For each point, bars show the standard error of the mean.

3.2.5 CTRS detection

337 Some indices seem to be more effective for CTRS detection than others (Figure 5b). When $\alpha = 1$,
 338 of all tree (im)balance indices, B_1 and I_s^* are the most affected, with a shift of 3 to 4 SD, while this
 339 shift is only between 1 and 2 SD for other (im)balance indices such as I_b , I_s , B_2 . I_{ca}^* and I_{ce}^* , the two
 340 Colless indices handling polytomies, display a similar pattern with a shift of 2 SD (Supp. Fig. S8).
 341 However, I_{ce}^* seems slightly more affected by CTRS, due probably to its algorithm focusing on children
 342 with an extreme number of leaves (see Methods). The number of interior nodes and the number of
 343 polytomies are affected by CTRS more than all other measured indices, with a shift of 8 to 9 SD (Figure
 344 5b). Interestingly, each of these indices seems to contain specific information about tree topology, as
 345 the correlations between their absolute values range between 0.99 and -0.17, although they all are
 346 correlated to α (Supp. Fig. S9). Thus, a method combining various indices (e.g., using approximate
 347 Bayesian computation) might be able to detect CTRS from population genomic data more accurately
 348 than a method using a single index. Furthermore, not all indices are robust to demographic events, as
 349 shown in Figure 5c: only I_b and B_2 seem unchanged when an expansion occurs during CTRS (vertical
 350 gray line at generation 150), with a small change for I_{ce}^* and wider changes for other indices. The
 351 remaining indices are all affected by the demographic event, although they still show tree imbalance
 352 of samples collected after the event (except for I_s , which reaches 0 soon after the event).

353 As with many evolutionary processes, the ability to detect CTRS also depends on the number
 354 of sampled individuals and loci. We assessed the effect of these two parameters on our ability to
 355 discriminate two scenarios using a Wilcoxon rank test: one of 20 generations of CTRS (strength
 356 $\alpha = 1$) before present and one without CTRS ($\alpha = 0$). We show that for all four studied summary
 357 statistics (i.e., I_b , B_2 , Number of polytomies and Tajima’s D), power increases with both the number

358 of sampled individuals and the number of sampled loci (Supp. Fig. S10). The number of polytomies
359 and Tajima’s D are the most effective indices, with the first index reaching a power above 0.95 (at
360 Type I error = 0.01) for 60 genomic regions of 1 Mb and 10 sampled individuals, and the second
361 reaching this power for 100 genomic regions of 1 Mb and 10 sampled individuals. However, as shown
362 previously, both indices are also impacted by changes in census population size and cannot thus be
363 used alone for CTRS inference. Conversely, I_b and B_2 are independent from changes in population
364 size, but display a much lower power of detection compared with the two previous indices. I_b needs
365 30 individuals and 100 genomic regions of 1 Mb in order to reach a power of 0.95, while B_2 needs 90
366 individuals and 100 genomic regions of 1 Mb to reach this power of detection. For CTRS detection,
367 the number of individuals seems to have a stronger impact on power of detection than the number
368 of genomic regions, with a power above 0.9 reached with I_b for 100 individuals and 10 independent
369 regions of 1 Mb, compared with a power of 0.15 with 10 individuals and 100 independent regions of
370 1 Mb, possibly due to the need to have a minimum number of sampled individuals in order to assess
371 topological properties of the population coalescent trees. As stated above, we expect a combination
372 of multiple indices using methods such as ABC to be even more effective for CTRS estimation from
373 genomic data, compared with single indices. Additionally, using the distribution of indices along the
374 genome might provide more information about past CTRS compared with the use of mere averages.

375 In conclusion, the evolution of Tajima’s D and imbalance measures over time highlights the com-
376 plexity and the timing of CTRS impacts on population genetics. When CTRS starts or stops, sudden
377 changes in effective population size occur. During the process, CTRS affects coalescent tree topology
378 (imbalance and number of polytomies) and branch lengths with a nonhomogenous reduction (young
379 branches less impacted than old branches). Imbalance is due to the transmission process, which yields
380 asymmetrical genealogies. The higher number of polytomies stems from the higher coalescence rate.
381 The nonhomogenous branch length reduction is similar to what occurs during an expansion. Although
382 the effective population size remains stable during CTRS, a pseudoexpansion occurs, due to the expan-
383 sion of large family lineages, which is compensated by the extinction of small family lineages (Sibert
384 *et al.*, 2002). All of these mechanisms affect the genomic signal commonly used for population genetic
385 inferences, and the next section will illustrate, based on simulations of an instantaneous expansion,
386 how demographic inference is impacted both before and after CTRS equilibrium.

387 3.3 Impact of CTRS on demographic inference

388 In this section, we investigate the impact of CTRS on demographic inference before and after CTRS
389 equilibrium. In the first case, the genomic signal of expansion is affected by the distortion in tree
390 topology (i.e., imbalance and higher number of polytomies) and by the recent change in effective
391 population size, while in the second case only changes in tree topology remain. We explored the
392 “Before CTRS equilibrium” scenario by inferring demography 20 generations after the beginning of
393 CTRS, and the “At equilibrium” scenario by inferring demography 1500 generations after the beginning
394 of CTRS. The 5-fold expansion event to be inferred occurs in both scenarios 300 generations before
395 the inference (more details in Methods).

396 Before CTRS equilibrium, we measured a strong bias in the demography inferred by δadi . When
397 $\alpha = 1$, the inferred growth factor has a median of 3 instead of 5 (relative bias = -0.37, MSRE =
398 0.18, compared with 0 and 0.04, respectively, for $\alpha = 0$) (Figure 6c). δadi inferences are based solely
399 on the SFS. After 20 generations of CTRS and without any change in census population size, SFS
400 shows a marked deficit of rare alleles due to the contraction of effective population size caused by
401 the initiation of CTRS, and an excess of common alleles due to this contraction combined with the
402 presence of many related individuals coming from large family lineages (Figure 6a). Conversely, in a
403 scenario of 20 generations of CTRS following an event of expansion, the SFS for $\alpha = 1$ is expectedly
404 a mix between the expansion-only pattern ($\alpha = 0$) and the CTRS pattern for $\alpha = 1$ (Figure 6b). In
405 this case, the SFS displays a smaller excess of rare alleles compared with the expansion-only pattern.
406 Since the excess of rare alleles is the main signal of expansions, a smaller expansion is inferred. The
407 contraction of the effective population size due to the initiation of CTRS reduces the excess of rare
408 alleles caused by the expansion event, yielding an inference of a smaller growth factor. Time since the
409 demographic event is also inferred less accurately after a period of 20 generations of CTRS (for $\alpha = 0$:
410 relative bias = -0.17, MSRE = 0.06; for $\alpha = 1$: relative bias = 0.22, MSRE = 0.21).

411 At CTRS equilibrium, for $\alpha = 1$, a median growth factor of 3.8 is inferred instead of 5 (relative
412 bias = -0.18, MSRE = 0.16, compared with -0.01 and 0.04, respectively, for $\alpha = 0$) (Figure 6g). The

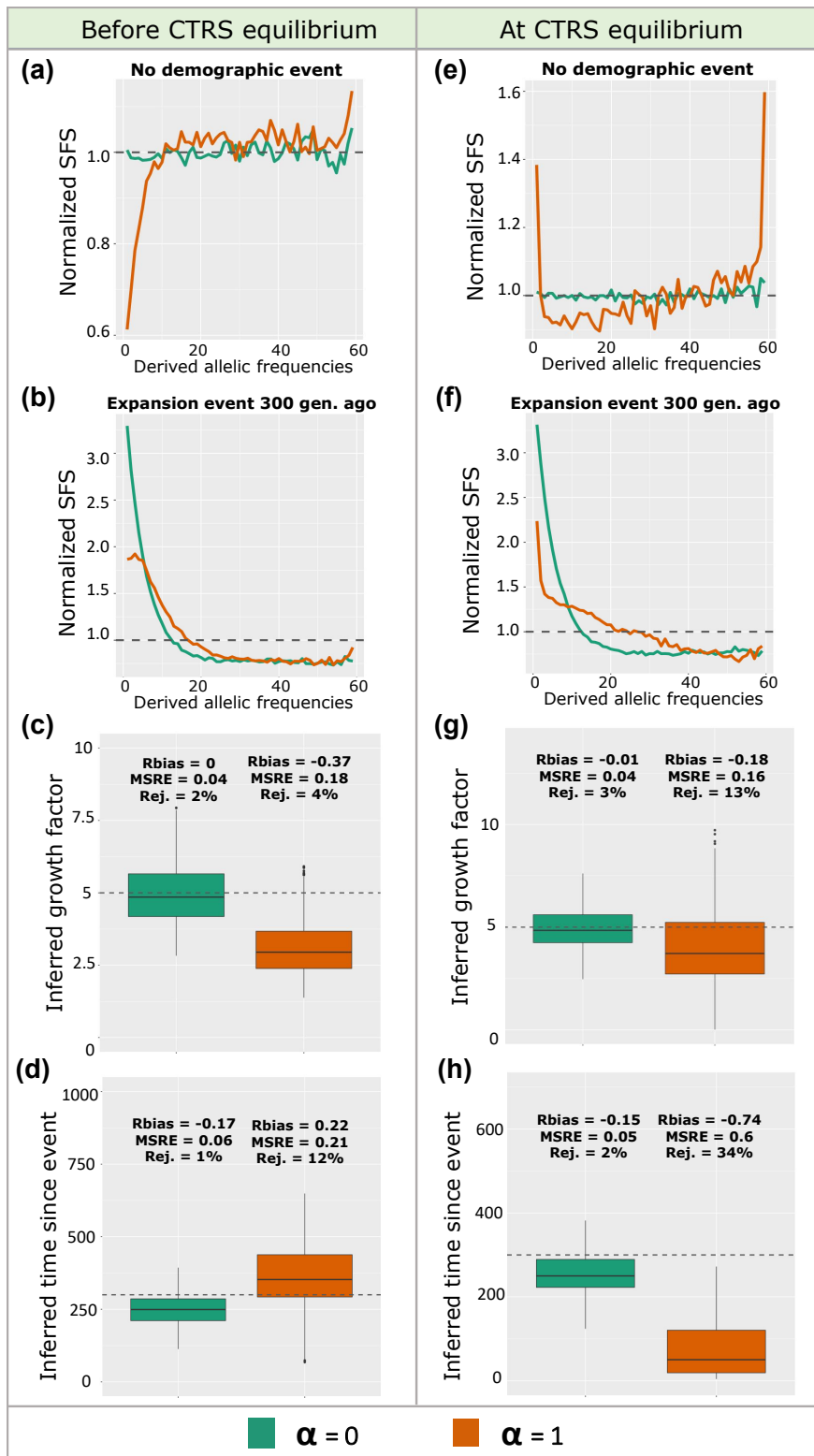


Figure 6: SFS and δadi inference of expansion parameters at two stages of CTRS. (a) and (e) SFS for $\alpha = 0$ and 1 with no demographic event. (b) and (f) SFS for $\alpha = 0$ and 1 after a 5-fold expansion 300 generations ago. (c) and (g) inferred growth factor for $\alpha = 0$ and 1, after a 5-fold expansion 300 generations ago. (d) and (h) inferred number of generations since expansion for $\alpha = 0$ and 1, after a 5-fold expansion 300 generations ago. (a-d) Scenario “Before CTRS equilibrium” (20 generations of CTRS before present). (e-f) Scenario “At CTRS equilibrium” (1500 generations of CTRS before present). MSRE, relative bias and percentage of rejected replicates displayed above each boxplot. In all cases, the $b = 1$ model of variance in progeny size is used.

413 SFS at CTRS equilibrium with no demographic event is U-shaped (Figure 6e). Tree imbalance and
414 the higher number yield the excess of rare and common alleles, while nonhomogenous reduction of
415 branch lengths contributes to the excess of rare alleles. When a demographic expansion occurs at
416 CTRS equilibrium, the SFS displays a tilted U-shape, with less excess of rare alleles in comparison to
417 the expansion-only scenario (Figure 6f). This is due to the smaller effective population size during the
418 generations where CTRS occurs, which induces an accelerated loss of part of the rare alleles created
419 by the fivefold expansion event. Since rare alleles are the main traces of this past expansion event,
420 a smaller expansion is inferred. The inferred time since the demographic event when the population
421 experienced 1500 generations of CTRS was strongly biased, with a median inference of 50 generations
422 since the demographic event instead of 300 ($\alpha = 0$: relative bias = -0.15, MSRE = 0.05; $\alpha = 1$: relative
423 bias = -0.74, MSRE = 0.6) (Figure 6h).

424 We thus showed that after a period of CTRS, whether short (20 generations) or long (1500 gen-
425 erations), past growth factors of expansion events are underestimated with an SFS-based inference
426 method, due to a lack of rare alleles compared with the neutral case scenario. The time since the
427 expansion event can be largely underestimated if it happened after a long period of CTRS and slightly
428 overestimated after a short period of CTRS.

429 4 Conclusions

430 Many studies evaluating CTRS strength in human populations rely on the computation of correlations
431 between parents and children progeny size from pedigree datasets (Murphy, 1999). However, we show
432 here that this measure cannot by itself account for the magnitude of CTRS effects on population
433 genetics. Indeed, under the high variance in progeny size model ($b = 1$), correlations are lower than
434 under the low variance model ($b = \infty$), while the impacts on population genetics are increased. Thus,
435 a more precise evaluation of CTRS from pedigree data would require considering the distributions
436 of parents and child progeny sizes in addition to the correlation values. Furthermore, the higher
437 correlations under the low variance model ($b = \infty$) could explain the higher correlations observed in
438 populations that exhibited a demographic transition (Murphy, 1999; Jennings *et al.*, 2012; Jennings
439 and Leslie, 2013). Indeed, a main characteristic of this transition is a decrease in progeny size variance.
440 Finally, we observe that CTRS has a stronger impact on effective size than the variance introduced in
441 the model. This result is supported by measurements in the Saguenay–Lac-Saint-Jean population for
442 similar levels of progeny size correlation (Heyer *et al.*, 2012).

443 CTRS impacts genomic diversity in two ways: (i) when CTRS begins or ends, populations undergo
444 a decrease (resp. increase) in effective size that impacts several population genetic statistics such as
445 Tajima’s D and SFS. This lower effective size stems from the increased variance in progeny size under
446 CTRS and from the transmission component itself. We could show that the latter accounts for most
447 part of the decrease in effective population size under CTRS. (ii) During the CTRS process and shortly
448 after the process stops, coalescent tree topologies (i.e., tree shape properties that are not related to
449 branch length) are distorted, which also impacts Tajima’s D and SFS. When CTRS lasts long enough,
450 the effect of the change in effective size disappears while tree topology distortion persists, inducing
451 lower genetic diversity and a U-shaped SFS. These two processes start together but have different
452 dynamics, yielding a complex effect on population genetics over time.

453 We showed that the distortion in coalescent tree topology affects two topological properties: (1)
454 trees are more imbalanced, which can be shown with balance and imbalance indices, and (2) the number
455 of polytomies increases. In theory, both of these effects could happen independently, as binary trees can
456 be imbalanced and polytomies do not necessarily induce imbalance. However, under CTRS, we show
457 that trees undergo a complex change in their topology, with an interplay between these two properties
458 of imbalance and polytomies. These two effects increase the proportions of rare and common alleles,
459 while a nonhomogenous reduction in branch lengths (Sibert *et al.*, 2002) increases only the proportion
460 of rare alleles, yielding a U-shaped SFS. Further studies could evaluate the relative impacts and possible
461 interactions between these processes.

462 The impact of CTRS on SFS explains why the SFS-based demographic inference performed by
463 $\delta a d i$ was biased for populations undergoing CTRS. After a few generations of CTRS, the growth
464 factors of past expansion events are underestimated. This result implies that past expansions, such as
465 the Neolithic ones, might be underestimated in populations experiencing CTRS, at least when inferred
466 based on SFS. After many generations under CTRS, the timing of expansion is strongly underestimated

467 as well. Furthermore, due to the decrease in effective population size induced by CTRS, past expansion
468 signals were lost more rapidly, when compared with scenarios without CTRS. Similarly, the signal of
469 other past events, such as bottlenecks, selection or migration, is expected to be erased more rapidly
470 in the presence of CTRS. We established that CTRS impacts an SFS-based inference method and
471 expect other approaches to be affected given that CTRS distorts coalescent trees, which are directly
472 or indirectly at the core of any inference method. CTRS is thus one more process among others that
473 can affect demographic inference (e.g., purifying and background selection (Johri *et al.*, 2021; Pouyet
474 *et al.*, 2018), biased gene conversion (Pouyet *et al.*, 2018), population structure (Mazet *et al.*, 2016),
475 selection, gene conversion, and biased sampling in microbial populations (Lapierre *et al.*, 2016)).

476 To disentangle the effects of demographic events from CTRS, imbalance indices that are unaffected
477 by variations in the census population size can be used. We showed that the power of detection of
478 CTRS from genomic data is less impacted by the number of independent regions than by the number
479 of sequenced individuals that should be high enough, a condition easily achieved with modern datasets.
480 However, these indices are computed from coalescent trees which first need to be reconstructed from
481 genomic data (e.g., using tools such as ARGweaver (Rasmussen *et al.*, 2014), tsinfer (Kelleher *et al.*,
482 2019), or relate (Speidel *et al.*, 2019)). This tree reconstruction step might not be able to infer a
483 perfectly accurate topology, yielding potential biases in the estimated (im)balance indices. Moreover,
484 in addition to the expected imprecision of the reconstruction of neutral trees, the behavior of these
485 tools under CTRS remains to be checked. Another possibility would be to build and train deep learning
486 networks directly on raw genomic data without reconstructing coalescent trees, as in Sanchez *et al.*
487 (2021), which would prevent the introduction of biases due to tree reconstruction, but might require
488 a larger amount of simulated data for training. To generate this large dataset, it would be useful
489 to develop a backward coalescent model of CTRS, as forward-in-time simulations are particularly
490 time-consuming.

491 Finally, we should address the question of the similarity between CTRS and natural selection:
492 in both cases, some individuals have more offspring than others and transmit this higher fertility
493 to their descendants. However, in the case of CTRS, fertility is culturally transmitted, whereas for
494 selection, it is genetically transmitted. The question is to what extent these processes affect the
495 genome differently. Without recombination, one might expect qualitatively similar effects of the two
496 processes on the genome: lower diversity and similar patterns for Tajima's D over time. Moreover,
497 tree topology is also expected to be distorted with an increase in imbalance (Fay and Wu, 2000; Li,
498 2011; Li and Wiehe, 2013) and number of polytomies (Durrett and Schweinsberg, 2005; Neher and
499 Hallatschek, 2013) under selection. The resemblance of the two processes is confirmed by a similar
500 U-shaped signature in SFS: selection also yields an excess of rare (Braverman *et al.*, 1995) and common
501 alleles (Fay and Wu, 2000).

502 However, a fairly clear difference exists between the CTRS model (based on the α parameter) used
503 here and the commonly used model of positive selection (based on the selection coefficient s , (Wright,
504 1932)). Under this model of selection, the beneficial allele can go to fixation, and selection stops at
505 that point. However, in the case of CTRS, the model is constructed in such a way that the TRS may
506 continue indefinitely. The CTRS model would more closely resemble a positive selection model with a
507 high mutation rate, preventing fixation. This difference between the two models makes sense relative
508 to reality: cultural transmission can be expected to be quite inaccurate in real life compared with
509 genetic transmission. This argument of "high mutation rate" in cultural transmission has been used to
510 resolve the so-called Fisher's paradox (Pettay *et al.*, 2005): how can correlations between parents' and
511 children's progeny size remain positive over time given the expected erosion of variance in the fertility
512 phenotype? The answer would be that these correlations stem from a CTRS and not a genetic TRS.
513 Thus, the unfaithful cultural transmission of fertility would explain why variance is maintained, with
514 the "high mutation rate" preventing the "fixation" of high-fertility cultural traits (Heyer *et al.*, 2012).
515 This difference in fixation between the two models might yield distinctive dynamics in population
516 genetics statistics. To further compare CTRS and selection models, an analytical reconciliation that
517 would link α to the selection coefficient would be pertinent.

518 A second difference between CTRS and selection appears when recombination is considered. In this
519 case, the selection signal is restricted over time to the locus under selection, as recombination events
520 accumulate, with a remaining local effect on nearby loci due to hitchhiking (Smith and Haigh, 1974).
521 The length of the region impacted by hitchhiking depends on the recombination rate, as well as on the
522 time under which selection has been acting. When fixation occurs, this time is equivalent to the time to

523 fixation, which is inversely proportional to the selection coefficient s (Kim and Stephan, 2002; Stephan,
524 2019). In human populations, even selection events that started rather recently have been shown to
525 give rise to a signal restricted to only a few megabases. For example, in the case of the selection
526 for lactase persistence in Africa (event dated to ~ 7000 years ago), the selection signal decreases very
527 rapidly over the 3 Mb sequenced (Tishkoff *et al.*, 2007). An even more recent selection event, such
528 as the one on the 3p12.1 chromosomal region in Mongolians, associated with energy metabolism and
529 reproductive traits, dated to approximately 50 generations ago (~ 1500 years), is almost undetectable
530 outside the 4 Mb region around the locus under selection (Nakayama *et al.*, 2017). Conversely, in the
531 case of CTRS, the effects are uniform over the whole genome since the transmission of fertility is not
532 conveyed by genetics: we showed in this paper the shift of the whole distribution of tree imbalances in
533 the genome toward higher values. We expect the distribution of indices across the genome to be quite
534 different in the case of selection, which would help distinguish between the two processes.

535 We can go farther and compare polygenic selection to CTRS, because of their propensity to affect
536 simultaneously distant loci in the genome. In particular, background selection, which has this ability
537 to affect large parts of the genome (Pouyet *et al.*, 2018), could strongly resemble CTRS in its effects.
538 Because of their potential similarity, distinguishing highly polygenic selection from CTRS might be
539 troublesome. However, it seems unlikely that even highly polygenic selection would have an effect
540 identical to CTRS for several reasons. First, the neutral parts of the genome are under the effect of
541 CTRS but not under that of polygenic selection (e.g., Pouyet *et al.* (2018) identified a set of SNPs that
542 are mostly unaffected by background selection). Second, in a polygenic selection, selective pressure
543 may have different parameters depending on the gene: the temporality may differ (selective pressure
544 does not start at the same time on each gene) as well as intensity (different selection coefficients for
545 each gene), yielding different coalescent trees across the genome (each gene tree telling its own history).
546 In fact, theoretical analyses showed different temporal dynamics in polygenic adaptation, with large
547 effect alleles contributing first, followed by small/intermediate-effect alleles (Hayward and Sella, 2022;
548 Barghi *et al.*, 2020). This process has been shown to be responsible for maize domestication, with a
549 central transcription factor (*teosinte branched 1*) driving adaptation (Studer *et al.*, 2011), although
550 most of the network controlled by this gene displays a selection signal as well (Wang *et al.*, 1999;
551 Studer *et al.*, 2017; Barghi *et al.*, 2020). Conversely, CTRS will tend to create trees that look similar
552 across the genome, since they are all affected uniformly by the same cultural history (a single α
553 parameter for the whole genome). Third, populations exchanging migrants will tend to have the same
554 alleles selected by multigenetic selection, whereas nongenetic TRS will select for different alleles in
555 each population (alleles randomly carried by large family lineages). Fourth, under polygenic selection,
556 genes can undergo a complex effect, combining not only the effects of their selection pressure, but also
557 the effects of nearby genes due to hitchhiking (Barton, 1995). This competing effect would not happen
558 under CTRS only, adding another difference between the effects of CTRS and of highly polygenic
559 selection. Ultimately, these three listed differences might help distinguish the two processes in real
560 data.

561 Furthermore, one may ask what happens when CTRS and selection are combined, which might be
562 the case in a number of populations. Competition between selection and CTRS might arise in the case
563 of a culturally fertile lineage carrying a disadvantageous allele. In fact, Austerlitz and Heyer (1998)
564 have shown that CTRS can increase the propensity of a population to maintain genetic diseases. This
565 increase in genetics disease can also stem from the reduction in diversity created by CTRS, under
566 which conditions, selection is less effective. Studying coalescent tree shapes under the combined effects
567 of selection and CTRS is also interesting: will trees be even more imbalanced compared with CTRS
568 alone, or is imbalance already saturated by CTRS? It is also possible that the sum of the two processes
569 will result in more balanced trees due to the aforementioned competition between them. The study of
570 the combination of these two processes is crucial to be able to distinguish them in real populations,
571 where both are likely to happen, in order to find their respective impact on genetic diversity and tree
572 topologies.

573 Finally, the analysis of CTRS provided here might be valid for any TRS that is not genetic. For
574 example, ecological inheritance (Odling-Smee, 1988; Danchin *et al.*, 2011), where an individual passes
575 on its environment to its offspring, could yield a similar process provided that: (1) the population is
576 settled in diverse environments, (2) the fitness varies with the environment, and (3) there is a vertical
577 transmission of the environment (Bonduriansky and Day, 2018). These conditions might be achieved
578 in plants whose seeds disperse little (Danchin *et al.*, 2011). Therefore, although the literature has

579 focused on *cultural TRS* until now (Blum *et al.*, 2006; Heyer *et al.*, 2012, 2015), one could generalize
580 this evolutionary process and call it *nongenetic TRS*.

581 **5 Data availability**

582 The SLiM code used to generate the simulated data and the Python code for summary statistics
583 computing and *δaδi* inference can be found at <https://github.com/jeremyguez/CTRS>.

584 **6 Acknowledgments**

585 We thank the editor and three anonymous reviewers for their helpful comments and suggestions. We
586 thank Matteo Fumagalli, Olivier François, Aurélien Tellier, Fanny Pouyet, Jean-Tristan Brandenburg,
587 Théophile Sanchez, Romain Laurent, Ferdinand Petit and Arnaud Quelin for the insightful interactions.

588 **7 Funding**

589 JG was supported by a French National Center for Scientific Research (CNRS) fellowship: 80Prime
590 (TransIA). FB was supported by Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zürich
591 Foundation. JC was supported by the Human Frontier Science Project (number RGY0075/2019). We
592 also thank ANR-20-CE45-0010-01 RoDAPoG.

593 **8 Conflicts of interest**

594 The authors declare that there is no conflict of interest.

595 **References**

- 596 Austerlitz F, Heyer E. 1998. Social transmission of reproductive behavior increases frequency of inher-
597 ited disorders in a young-expanding population. *Proceedings of the National Academy of Sciences*.
598 95:15140–15144.
- 599 Barber JS. 2001. The intergenerational transmission of age at first birth among married and unmarried
600 men and women. *Social Science Research*. 30:219–247.
- 601 Barghi N, Hermisson J, Schlötterer C. 2020. Polygenic adaptation: a unifying framework to understand
602 positive selection. *Nature Reviews Genetics*. 21:769–781. Number: 12 Publisher: Nature Publishing
603 Group.
- 604 Barton NH. 1995. Linkage and the limits to natural selection. *Genetics*. 140:821–841.
- 605 Beichman AC, Huerta-Sanchez E, Lohmueller KE. 2018. Using genomic data to infer historic population
606 dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics*. 49:433–
607 456.
- 608 Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE,
609 Hirschhorn JN. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *The*
610 *American Journal of Human Genetics*. 74:1111–1120.
- 611 Bienvenu F, Cardona G, Scornavacca C. 2021. Revisiting Shao and Sokal’s B_2 index of phylogenetic
612 balance. *Journal of Mathematical Biology*. 83:1–43.
- 613 Blum MGB, Heyer E, François O, Austerlitz F. 2006. Matrilineal fertility inheritance detected in
614 hunter-gatherer populations using the imbalance of gene genealogies. *PLOS Genetics*. 2:e122.
- 615 Boitard S, Rodríguez W, Jay F, Mona S, Austerlitz F. 2016. Inferring population size history from
616 large samples of genome-wide molecular data - an approximate Bayesian computation approach.
617 *PLOS Genetics*. 12:1–36.

- 618 Bonduriansky R, Day T. 2018. *Extended Heredity: A New Understanding of Inheritance and Evolution*.
619 Princeton University Press. Princeton (New Jersey). Google-Books-ID: 1j63DwAAQBAJ.
- 620 Brandenburg JT, Austerlitz F, Toupance B. 2012. Impact of fertility transmission and other sociode-
621 mographic factors on reproductive success and coalescent trees. *Genetics Research*. 94:121–131.
- 622 Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The hitchhiking effect on the
623 site frequency spectrum of DNA polymorphisms. *Genetics*. 140:783–796.
- 624 Chaix R, Austerlitz F, Khégay T, Jacquesson S, Hammer MF, Heyer E, Quintana-Murci L. 2004. The
625 genetic or mythical ancestry of descent groups: Lessons from the y chromosome. *The American*
626 *Journal of Human Genetics*. 75:1113–1116.
- 627 Colless DH. 1982. Review of “Phylogenetics: the theory and practice of phylogenetic systematics”.
628 *Systematic Zoology*. 31:100–104.
- 629 Danchin E, Charmantier A, Champagne FA, Mesoudi A, Pujol B, Blanchet S. 2011. Beyond DNA:
630 integrating inclusive inheritance into an extended theory of evolution. *Nature Reviews Genetics*.
631 12:475–486. Number: 7 Publisher: Nature Publishing Group.
- 632 de Valk HA. 2013. Intergenerational discrepancies in fertility preferences among immigrant and Dutch
633 families. *The History of the Family*. 18:209–225.
- 634 Duncan OD, Freedman R, Coble JM, Slesinger DP. 1965. Marital fertility and size of family of orien-
635 tation. *Demography*. 2:508–515.
- 636 Durrett R, Schweinsberg J. 2005. A coalescent model for the effect of advantageous mutations on the
637 genealogy of a population. *Stochastic Processes and their Applications*. 115:1628–1657.
- 638 Engh AL, Esch K, Smale L, Holekamp KE. 2000. Mechanisms of maternal rank ‘inheritance’ in the
639 spotted hyaena, *Crocuta crocuta*. *Animal Behaviour*. 60:323–332.
- 640 Ewens WJ. 2016. Effective population size, In: Kliman RM, editor, *Encyclopedia of evolutionary*
641 *biology*, Academic Press. Waltham (MA). pp. 494–497.
- 642 Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference
643 from genomic and SNP data. *PLOS Genetics*. 9:e1003905.
- 644 Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics*. 155:1405–1413.
- 645 Frere CH, Krutzen M, Mann J, Connor RC, Bejder L, Sherwin WB. 2010. Social and genetic interac-
646 tions drive fitness variation in a free-living dolphin population. *Proceedings of the National Academy*
647 *of Sciences*. 107:19949–19954.
- 648 Gagnon A, Heyer E. 2001. Intergenerational correlation of effective family size in early Québec
649 (Canada): Correlation of effective family size. *American Journal of Human Biology*. 13:645–659.
- 650 Gagnon A, Toupance B, Tremblay M, Beise J, Heyer E. 2006. Transmission of migration propen-
651 sity increases genetic divergence between populations. *American Journal of Physical Anthropology*.
652 129:630–636.
- 653 Gerbault P, Liebert A, Itan Y, Powell A, Currat M, Burger J, Swallow DM, Thomas MG. 2011.
654 Evolution of lactase persistence: an example of human niche construction. *Philosophical Transactions*
655 *of the Royal Society B: Biological Sciences*. 366:863–877.
- 656 Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demo-
657 graphic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*.
658 5:e1000695.
- 659 Haller BC, Messer PW. 2019. SLiM 3: Forward genetic simulations beyond the Wright–Fisher model.
660 *Molecular Biology and Evolution*. 36:632–637.
- 661 Hayward LK, Sella G. 2022. Polygenic adaptation after a sudden change in environment. *eLife*.
662 11:e66697.

- 663 Heyer E, Brandenburg JT, Leonardi M, Toupance B, Balaesque P, Hegay T, Aldashev A, Austerlitz
664 F. 2015. Patrilineal populations show more male transmission of reproductive success than cognatic
665 populations in Central Asia, which reduces their genetic diversity: Cultural transmission of fitness.
666 *American Journal of Physical Anthropology*. 157:537–543.
- 667 Heyer E, Chaix R, Pavard S, Austerlitz F. 2012. Sex-specific demographic behaviours that shape human
668 genomic variation: sex-specific behaviours and genomic variation. *Molecular Ecology*. 21:597–612.
- 669 Jay F, Boitard S, Austerlitz F. 2019. An ABC method for whole-genome sequence data: Inferring
670 paleolithic and neolithic human expansions. *Molecular Biology and Evolution*. 36:1565–1579.
- 671 Jennings J, Leslie P. 2013. Differences in intergenerational fertility associations by sex and race in Saba,
672 Dutch Caribbean, 1876-2004. *The history of the family : an international quarterly*. 18:135–153.
- 673 Jennings J, Sullivan A, Hacker J. 2012. Intergenerational Transmission of Reproductive Behavior during
674 the Demographic Transition. *The Journal of interdisciplinary history*. 42:543–69.
- 675 Johri P, Riall K, Becher H, Excoffier L, Charlesworth B, Jensen JD. 2021. The impact of purifying
676 and background selection on the inference of population history: Problems and prospects. *Molecular
677 Biology and Evolution*. 38:2986–3003.
- 678 Kawai M. 1958. On the rank system in a natural group of Japanese monkey (I): the basic and dependent
679 rank. *Primates*. 1:111–130.
- 680 Kelleher J, Etheridge AM, McVean G. 2016. Efficient Coalescent Simulation and Genealogical Analysis
681 for Large Sample Sizes. *PLOS Computational Biology*. 12:e1004842. Publisher: Public Library of
682 Science.
- 683 Kelleher J, Wong Y, Wohns AW, Fadil C, Albers PK, McVean G. 2019. Inferring whole-genome histories
684 in large population datasets. *Nature Genetics*. 51:1330–1338.
- 685 Kelly MJ. 2001. Lineage loss in serengeti cheetahs: Consequences of high reproductive variance and
686 heritability of fitness on effective population size. *Conservation Biology*. 15:11.
- 687 Kim Y, Stephan W. 2002. Detecting a Local Signature of Genetic Hitchhiking Along a Recombining
688 Chromosome. *Genetics*. 160:765–777.
- 689 Kingman JFC. 1982. The coalescent. *Stochastic Processes and their Applications*. 13:235–248.
- 690 Kohler HP, Rodgers JL, Christensen K. 1999. Is fertility behavior in our genes? Findings from a
691 Danish twin study. *Population and Development Review*. 25:253–288.
- 692 Kolk M. 2014. Multigenerational transmission of family size in contemporary Sweden. *Population
693 Studies*. 68:111–129.
- 694 Lapierre M, Blin C, Lambert A, Achaz G, Rocha EPC. 2016. The impact of selection, gene conversion,
695 and biased sampling on the assessment of microbial demography. *Molecular Biology and Evolution*.
696 33:1711–1725.
- 697 Lapierre M, Lambert A, Achaz G. 2017. Accuracy of demographic inferences from the Site Frequency
698 Spectrum: The case of the Yoruba population. *Genetics*. 206:439–449.
- 699 Lawson DW, Mace R. 2011. Parental investment and the optimization of human family size. *Philo-
700 sophical Transactions of the Royal Society B: Biological Sciences*. 366:333–343.
- 701 Li H. 2011. A New Test for Detecting Recent Positive Selection that is Free from the Confounding
702 Impacts of Demography. *Molecular Biology and Evolution*. 28:365–375.
- 703 Li H, Wiehe T. 2013. Coalescent Tree Imbalance and a Simple Test for Selective Sweeps Based on
704 Microsatellite Variation. *PLOS Computational Biology*. 9:e1003060. Publisher: Public Library of
705 Science.
- 706 Liu X, Fu YX. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency
707 spectra. *Genome Biology*. 21:280.

- 708 Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. 2016. On the importance of being struc-
709 tured: instantaneous coalescence rates and human evolution—lessons for ancestral population size
710 inference? *Heredity*. 116:362–371.
- 711 Mills MC, Troup FC. 2015. The biodemography of fertility: A review and future research frontiers.
712 *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*. 67:397–424.
- 713 Mondal M, Bertranpetit J, Lao O. 2019. Approximate Bayesian computation with deep learning sup-
714 ports a third archaic introgression in Asia and Oceania. *Nature Communications*. 10:246.
- 715 Mulder MB, Bowles S, Hertz T, Bell A, Beise J, Clark G, Fazzio I, Gurven M, Hill K, Hooper PL *et al.*
716 2009. Intergenerational wealth transmission and the dynamics of inequality in small-scale societies.
717 *Science*. 326:682–688.
- 718 Murphy M. 1999. Is the relationship between fertility of parents and children really weak? *Biodemog-*
719 *raphy and Social Biology*. 46:122–145.
- 720 Murphy M. 2013. The intergenerational transmission of reproductive behaviour: comparative perspec-
721 tives. *The History of the Family*. 18:107–115.
- 722 Murphy M, Wang D. 2001. Family-level continuities in childbearing in low-fertility societies. *European*
723 *Journal of Population / Revue Européenne de Démographie*. 17:75–96.
- 724 Nakayama K, Ohashi J, Watanabe K, Munkhtulga L, Iwamoto S. 2017. Evidence for Very Recent
725 Positive Selection in Mongolians. *Molecular Biology and Evolution*. 34:1936–1946.
- 726 Neher RA, Hallatschek O. 2013. Genealogies of rapidly adapting populations. *Proceedings of the Na-*
727 *tional Academy of Sciences*. 110:437–442.
- 728 Odling-Smee FJ. 1988. Niche-constructing phenotypes, In: , The MIT Press. Cambridge, MA, US. pp.
729 73–132.
- 730 Pettay JE, Kruuk LEB, Jokela J, Lummaa V. 2005. Heritability and genetic constraints of life-history
731 trait evolution in preindustrial humans. *Proceedings of the National Academy of Sciences*. 102:2838–
732 2843.
- 733 Pluzhnikov A, Nolan DK, Tan Z, McPeck MS, Ober C. 2007. Correlation of Intergenerational Family
734 Sizes Suggests a Genetic Component of Reproductive Fitness. *The American Journal of Human*
735 *Genetics*. 81:165–169. Publisher: Elsevier.
- 736 Potter RG, Kantner JF. 1955. Social and psychological factors affecting fertility. XXVIII. The influence
737 of siblings and friends on fertility. *The Milbank Memorial Fund Quarterly*. 33:246.
- 738 Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. 2018. Background selection and biased gene conversion
739 affect more than 95% of the human genome and bias demographic inferences. *eLife*. 7:e36317.
- 740 Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. 2014. Genome-wide inference of ancestral recombi-
741 nation graphs. *PLOS Genetics*. 10:e1004342.
- 742 Rodgers JL, Kohler HP, Kyvik KO, Christensen K. 2001. Behavior genetic modeling of human fertility:
743 findings from a contemporary Danish twin study. *Demography*. 38:14.
- 744 Sackin MJ. 1972. “Good” and “bad” phenograms. *Systematic Biology*. 21:225–226.
- 745 Sanchez T, Cury J, Charpiat G, Jay F. 2021. Deep learning for population size history inference:
746 Design, comparison and combination with approximate Bayesian computation. *Molecular Ecology*
747 *Resources*. 21:2645–2660.
- 748 Segurel L, Guarino-Vignon P, Marchi N, Lafosse S, Laurent R, Bon C, Fabre A, Hegay T, Heyer E.
749 2020. Why and when was lactase persistence selected for? Insights from Central Asian herders and
750 ancient DNA. *PLoS biology*. 18:e3000742.
- 751 Shao KT, Sokal RR. 1990. Tree balance. *Systematic Zoology*. 39:266–276.

- 752 Sheehan S, Song YS. 2016. Deep learning for population genetic inference. *PLOS Computational*
753 *Biology*. 12:1–28.
- 754 Sibert A, Austerlitz F, Heyer E. 2002. Wright–Fisher revisited: The case of fertility correlation. *The-*
755 *oretical Population Biology*. 62:181–197.
- 756 Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genetical Research*. 23:23–35.
- 757 Sorokowski P, Sorokowska A, Danel DP. 2013. Why pigs are important in Papua? Wealth, height and
758 reproductive success among the Yali tribe of West Papua. *Economics & Human Biology*. 11:382–390.
- 759 Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation for
760 thousands of samples. *Nature Genetics*. 51:1321–1329.
- 761 Stephan W. 2019. Selective Sweeps. *Genetics*. 211:5–13.
- 762 Studer A, Zhao Q, Ross-Ibarra J, Doebley J. 2011. Identification of a functional transposon insertion
763 in the maize domestication gene *tb1*. *Nature Genetics*. 43:1160–1163.
- 764 Studer AJ, Wang H, Doebley JF. 2017. Selection During Maize Domestication Targeted a Gene Net-
765 work Controlling Plant and Inflorescence Architecture. *Genetics*. 207:755–765.
- 766 Swallow DM. 2003. Genetics of lactase persistence and lactose intolerance. *Annual Review of Genetics*.
767 37:197–219.
- 768 Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from
769 hundreds of unphased whole-genomes. *Nature genetics*. 49:303–309.
- 770 Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM,
771 Hirbo JB, Osman M *et al.* 2007. Convergent adaptation of human lactase persistence in Africa and
772 Europe. *Nature Genetics*. 39:31–40. Number: 1 Publisher: Nature Publishing Group.
- 773 Wang RL, Stec A, Hey J, Lukens L, Doebley J. 1999. The limits of selection during maize domestication.
774 *Nature*. 398:236–239.
- 775 Whitehead H. 1998. Cultural selection and genetic diversity in matrilineal whales. *Science*. 282:1708–
776 1711.
- 777 Wright S. 1932. The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution. In: .
778 The Hague, The Netherlands. Pergamon Press.
- 779 Wright S. 1938. Size of population and breeding structure in relation to evolution. *Science*. 87:430–431.

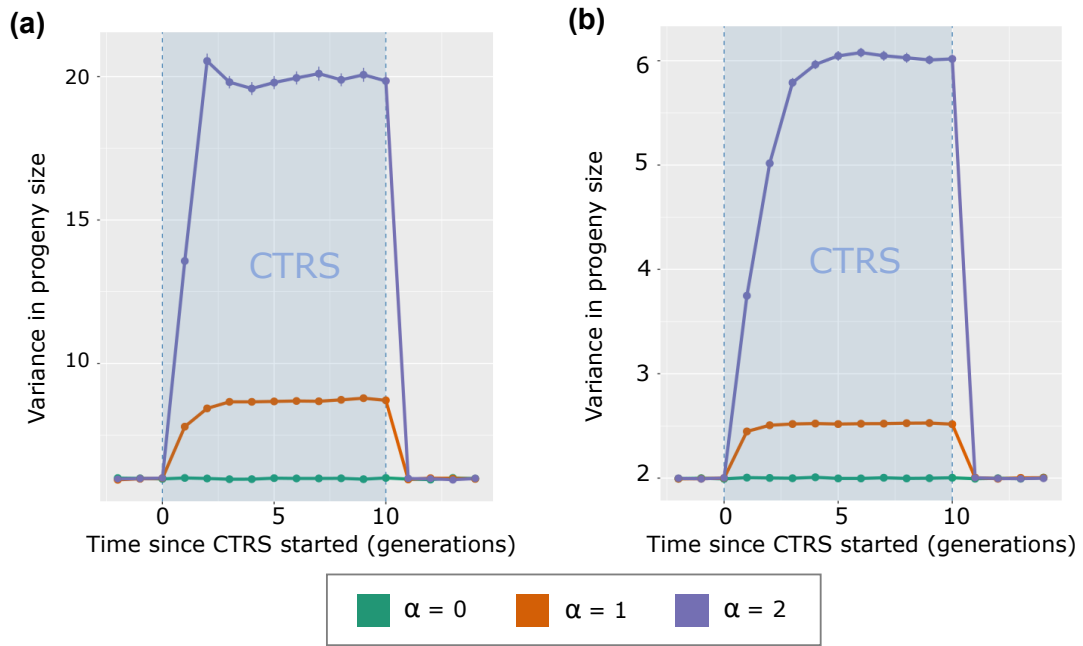


Figure S1: Variance of progeny size as a function of time for $\alpha = 0, 1,$ and 2 . The blue rectangle corresponds to the period when populations are under CTRS. Generations are counted from the beginning of CTRS. (a) $b = \infty$ model (low variance of progeny size). (b) $b = 1$ model (high variance of progeny size).

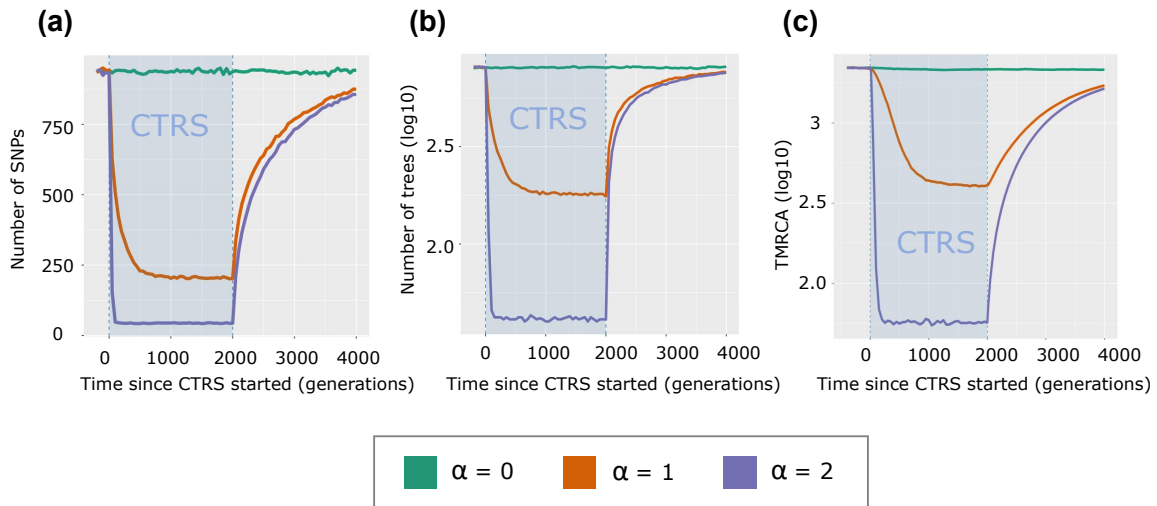


Figure S2: Number of SNPs (a), number of trees (log 10 scale) (b), and TMRCA (log 10 scale) (c) across generations. In all cases, the $b = 1$ model of variance in progeny size is used.

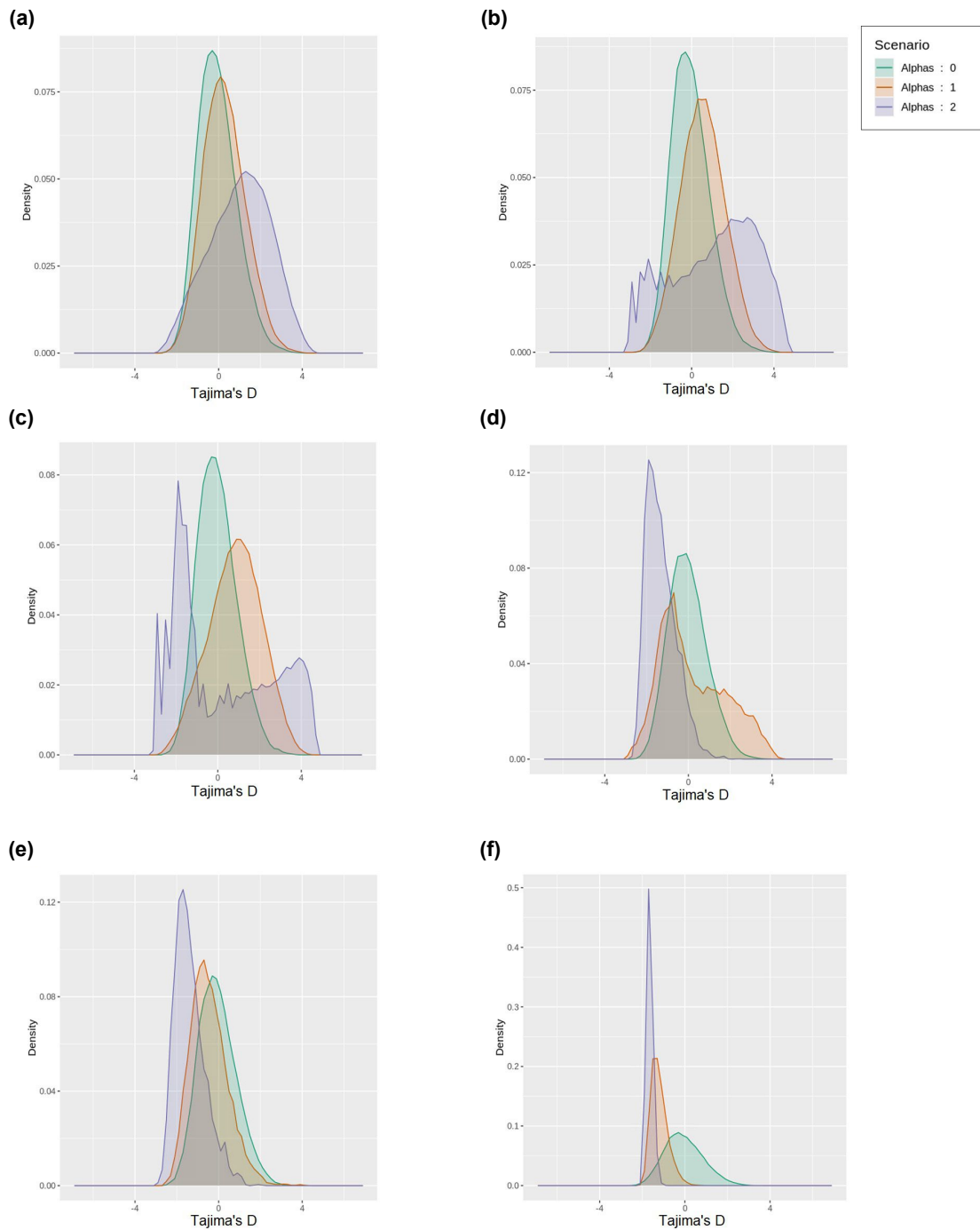


Figure S3: Distribution of Tajima's D across the genome. (a-e) 10, 20, 50, 500, 1500 generations since the starting of CTRS. (f) 500 generations without CTRS, after a period of 2000 generations of CTRS. The $b = 1$ model of variance in progeny size is used.

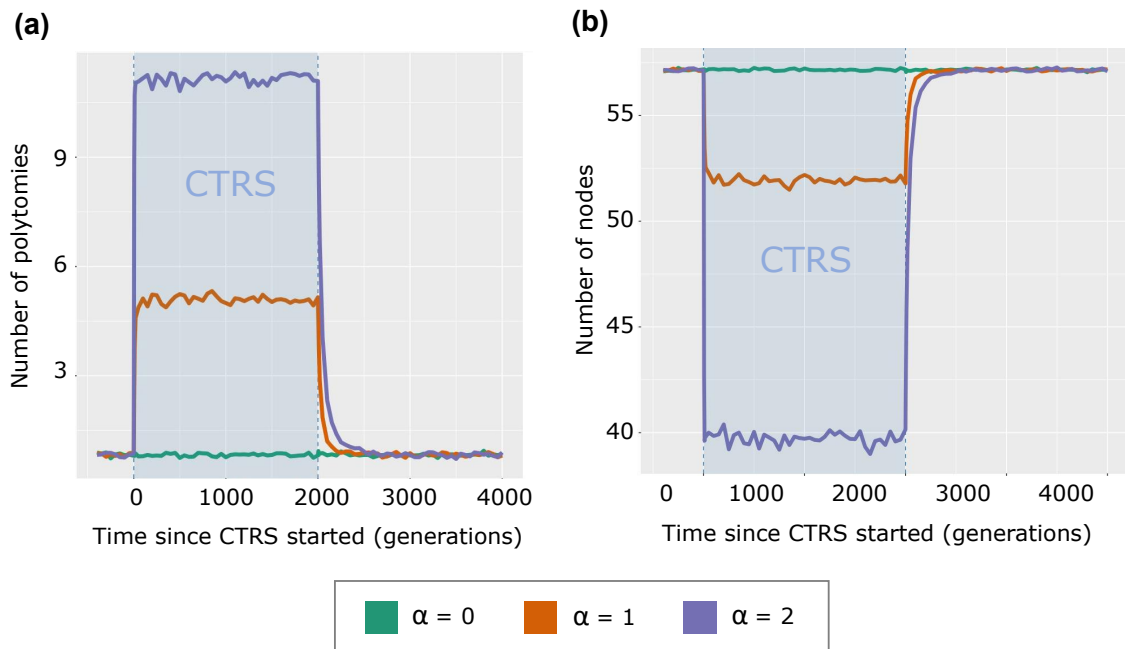


Figure S4: Number of polytomies (a) and number of nodes (b) throughout generations. The $b = 1$ model of variance in progeny size is used.

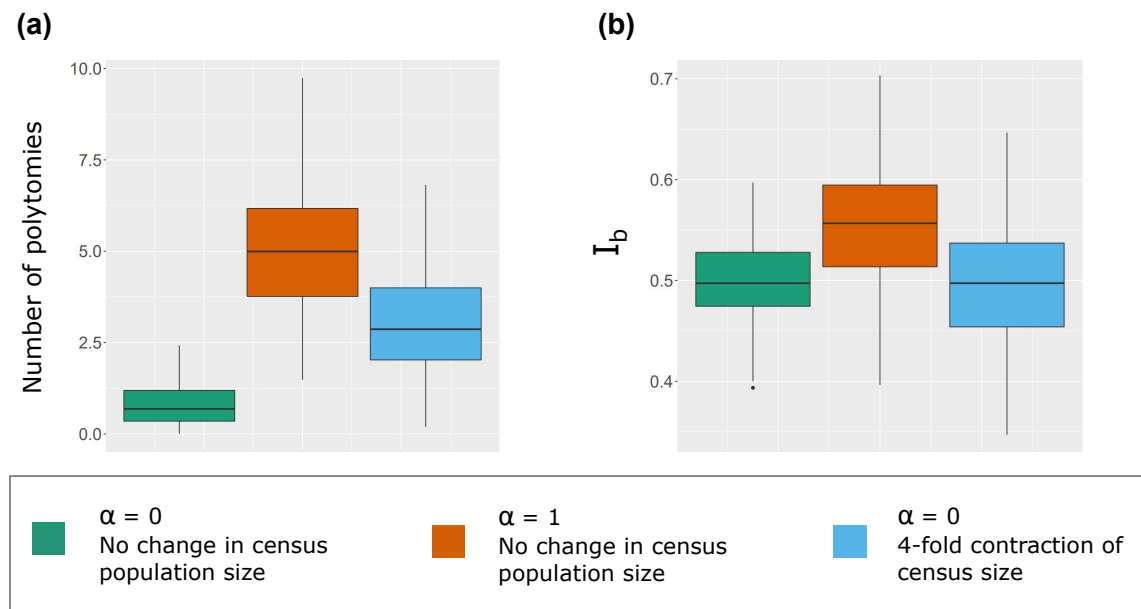


Figure S5: I_b and average number of polytomies for three scenarios. The CTRS in the $\alpha = 1$ scenario lasted for 500 generations before present. The 4-fold contraction happened 500 generations before present. In all cases, the $b = 1$ model of variance in progeny size is used.

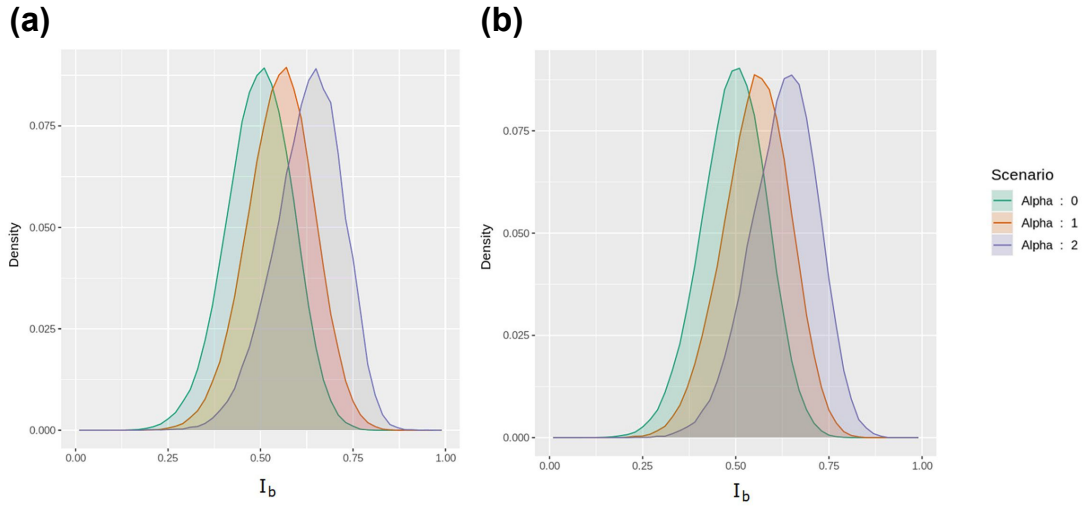


Figure S6: I_b distributions across the genome for $\alpha = 0, 1,$ and $2,$ after 20 (a) and 500 (b) generations of CTRS. The $b = 1$ model of variance in progeny size is used.

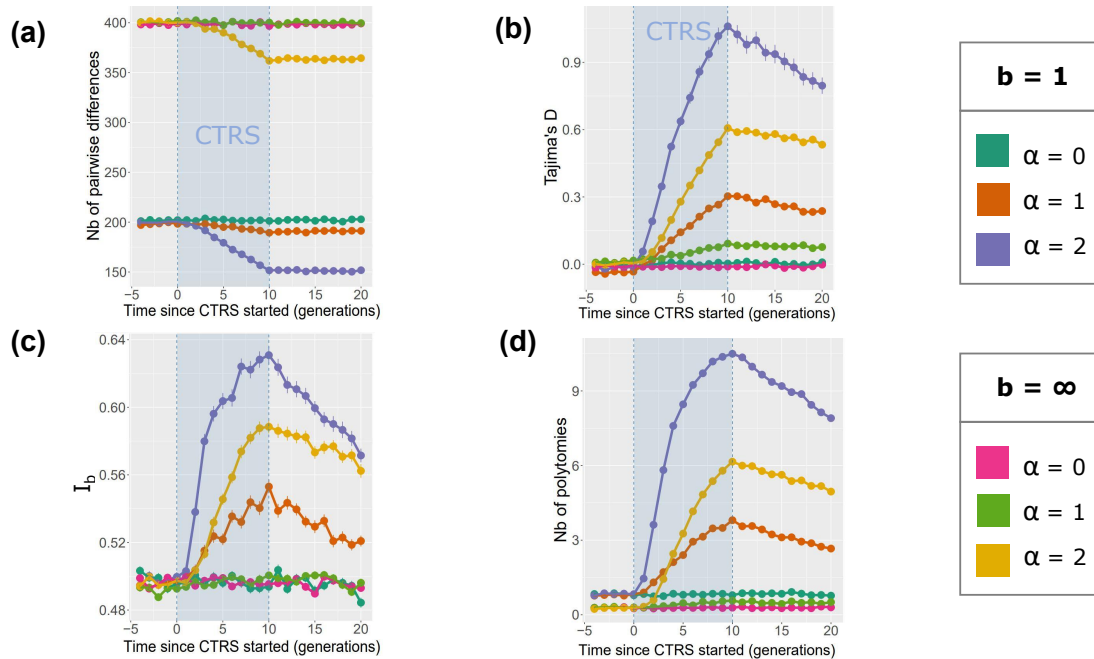


Figure S7: Number of pairwise differences, Tajima's $D,$ number of polytomies and I_b under 10 generations of CTRS followed by 10 generations without CTRS. Both $b = 1$ and $b = \infty$ models of variance in progeny size are used.

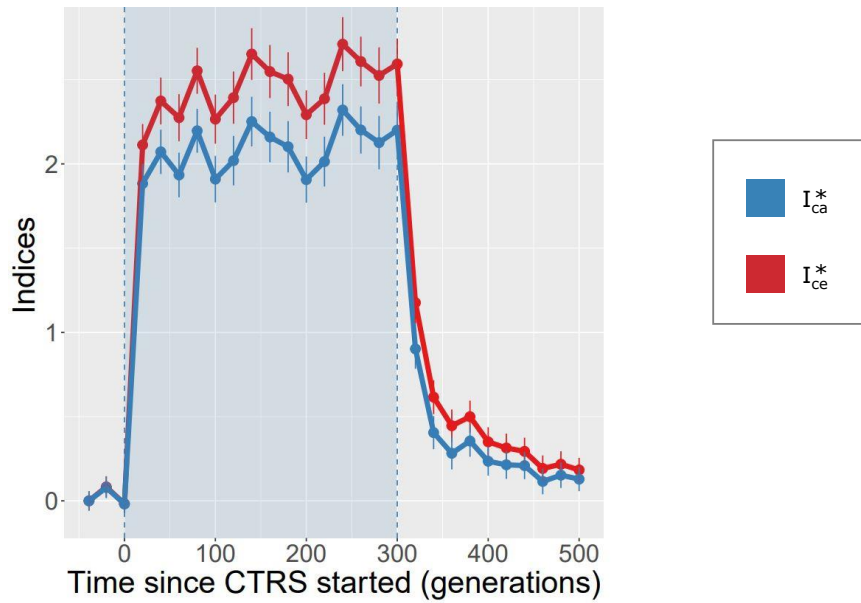


Figure S8: Two Colless index modifications to handle polytomies: I_{ca}^* and I_{ce}^* . See Methods for details on algorithms.

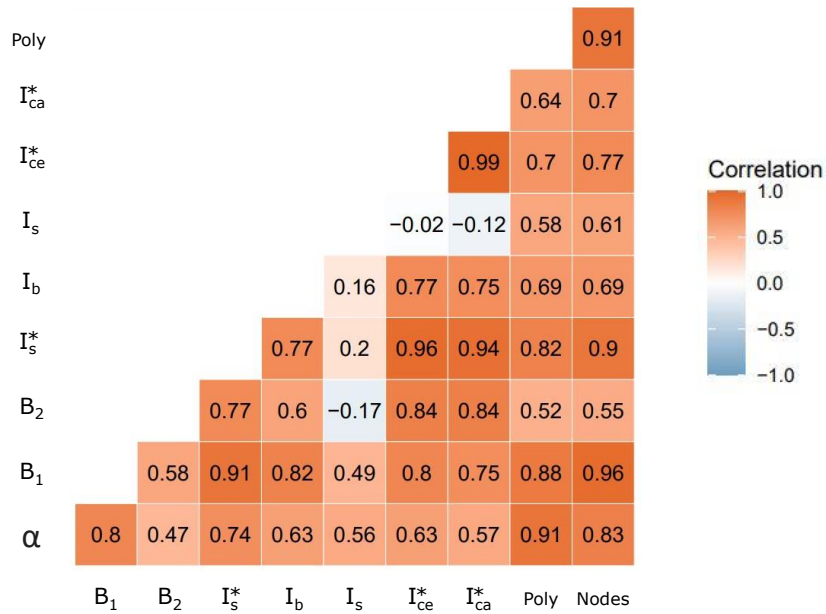


Figure S9: correlations between indices after 50 generations of CTRS.

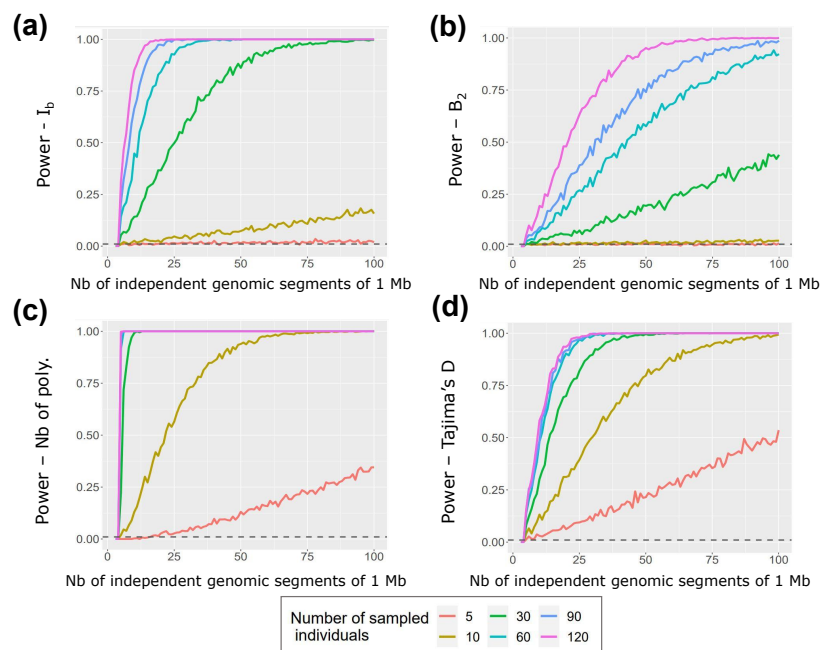


Figure S10: Power of distinguishing $\alpha = 1$ from $\alpha = 0$ scenarios (using a Wilcoxon test with the significance threshold set to 0.01), for 4 indices: I_b (a), B_2 (b), number of polytomies (c), Tajima's D (d).