

## File S1: Analysis of Indian populations

Following MOORJANI *et al.* (2013), we ran the  $F_4$  ratio tool in the ADMIXTOOLS package (PATTERSON *et al.* 2012) on Georgian, Basque, Yoruba, Onge and the focal Indian population to estimate ANI ancestry proportions in these populations (Fig. 5). We fit a latitudinal cline to these ancestry proportions (Eq. 1) returning a cline center at  $24^\circ 4'N$  and  $\sigma\sqrt{\tau} = 25.4$ . Because the gradient of ancestry could run along any geographic axis, we also tried to fit ancestry proportion clines to various transects using linear combinations of latitude and longitude. Since these did not produce substantially better fits than latitude alone, we chose to use latitude as our geographic axis (results not shown). Through this analysis, we aimed to closely follow the procedure outlined in (MOORJANI *et al.* 2013) to generate LD curves and improve model-fitting.

We then generated co-ancestry decay curves in ALDER for each of these samples, using weightings from Basque and Onge parental populations as proxies for the ANI and ASI populations (see MOORJANI *et al.* (2013)). We consider three possible contact zone scenarios under our geographic model: One in which all population samples form a contact zone and, based on the findings of earlier studies, one that comprises only the Indo-European and one that comprises only the Dravidian populations. We initially attempted to fit the  $\tau$ ,  $\sigma$  and  $F$  parameters in Eq. 12 simultaneously, but faced some difficulty as there appears to be limited information about  $F$ . This results in wide range of values fitting the data equally well, but give rise to very different surfaces for  $\sigma$  and  $\tau$ . We attributed this to a deficit of information in the curves, leading to non-identifiability, due to relative low levels of differentiation and relatively rapid decay of ancestry-LD. The difficulty in estimating the intercept of admixture-LD curves had been noted before (LOH *et al.* 2013), and can reflect the fact that very close pairs of markers are discarded to remove the effects of LD in the ancestral populations. This results in the fitted curve being relatively unconstrained near  $r = 0$ . To remedy this, we estimated  $\mathcal{F}$  using an approach similar to that taken by MOORJANI *et al.* (2013). Using MIXMAPPER (LIPSON *et al.* 2013), we estimated the value of  $\mathcal{F}$  as  $2F_2(ANI; ASI)^2$  using the Onge and Basque populations as present day proxies. We then fit values of  $\sigma$  and  $\tau$  under the range of  $F_2$  values computed by MIXMAPPER ((0.015, 0.042)). The profile likelihood surface was generated over 20 values of  $\mathcal{F}$ . We also use the value estimated above as the cline center for all three fits.

We first fit our LD curves to all populations under a model in which all Indo-European and Dravidian populations are the outcome of a single admixture contact zone. The best fit was approximately 220 generations since contact with  $\sigma = (0.9 \text{ degrees} \approx 100 \text{ km})/\text{generation}$  (Fig. 5). Fits to the subset of populations classified as Indo-European yielded a contact zone age of approximately 200 generations, and  $\sigma = (1.3 \text{ degrees} \approx 144 \text{ km})/\text{generation}$  (Fig. S5). Finally, we fit the subset of Dravidian populations (Fig. S5), which found a best fit of 460 generations with  $\sigma = (1.2 \text{ degrees} \approx 133 \text{ km})/\text{generation}$  on a relatively flat surface. This is likely because there is very little information in the decay of LD in this subset given there are so few Dravidian populations, and that the LD curves are relatively flat. The profile likelihood surface was generated over 40 evenly distributed values of  $\mathcal{F}$  spanning the values inferred above using MIXMAPPER. For all three groups of populations we used our earlier estimate for cline center.