We tested the performance of ALFA-K against synthetic data generated by our ABM. Simulations were performed on multiple randomly generated landscapes across a range of λ (see: methods, SX Fig.). In these test simulations, populations on the most complex landscapes (Fig.2A, $\lambda=0.2$) tended not to increase in fitness due to being stuck at local maxima in the fitness landscape. Populations evolving on less complex fitness landscapes tended to have gradual increases in fitness over time (Fig.2A, $\lambda=1.6$). Data from the first 200 days of each simulated population was used to train ALFA-K. We trained ALFA-K varying both the number of longitudinal samples and the hyperparameter N, then evaluated the results using a cross validation procedure (Fig.2B) which tests the ability of ALFA-K to infer the fitness of karyotypes withheld from the input data (see Methods). ALFA-K performance was not sensitive to the value of the hyperparameter N within the tested range. It was however sensitive to landscape complexity and the number of longitudinal samples in the input, with at least 4 samples needed to obtain satisfactory results. We next tested the ability of ALFA-K to predict population evolution for the time period from 200-300 days that was withheld from the training data. We used the angle metric (SX Fig.) to evaluate predictive performance, in which values below 90 degrees are taken as good predictions. Landscapes with good cross-validation scores ($R^2>0.3$) predicted future population evolution well (Fig. 2C). The results from the forward prediction tests agreed with the cross validation test in terms of sensitivity to landscape complexity, number of sampled timepoints, and the value of the hyperparameter N (Fig. 2D). Finally, we evaluated the robustness of ALFA-K to different values of missegregation rate (Fig. 2E). The procedure was robust to a wide range of missegregation rates up to a threshold value, which occurred when karyotype became too unstable to estimate the fitness of subpopulations across multiple longitudinal timepoints. Up to this threshold however, increasing missegregation rate benefits ALFA-K by allowing a larger region of the fitness landscape to be charted.

A collage of a graph

Description automatically generated

Figure 2 Validation of ALFA-K against synthetic data. A) Mean fitness of ABM cell populations evolving on artificial fitness landscapes of varying complexity (as determined by $\lambda$). Data from the first 200 days of each simulation was demarcated by the red vertical lines was used to train ALFA-K. B) Cross validation results of ALFA-K for varying numbers of sampled timepoints and values of the hyperparameter N. c) Evolutionary prediction results of ALFA-K agggregated across all landscapes at various times in the validation period. Results are grouped by performance on the cross validation test and summarized by the angle metric. D) Evolutionary prediction results of ALFA-K for varying numbers of sampled timepoints and values of the hyperparameter N. Prediction results are summarized by the angle metric. E)