Supporting Information for

Network-based genetic monitoring of landscape fragmentation

Ohad Peled, Jaehee Kim and Gili Greenbaum

Gili Greenbaum.
E-mail: gil.G@mail.huji.ac.il

This PDF file includes:

Supporting Information Text

Deriving genetic measures for population networks. We begin by formulating a transformation from a migration network M with K nodes to a coalescent-time matrix T, which is a symmetric matrix with positive entries. Entry T_{ij} denotes the expected coalescence time of a pair of individuals sampled from populations i and j, and T_{ii} denotes the expected coalescence time of a pair of individuals sampled from the same population. For most analyses, we assume that all populations are of the same size N, and we scale time accordingly to coalescent units of $2N$ generations. Under these notations, the transformation from M to T can be computed by solving a set of linear equations (1):

$$\begin{cases} (1+M_{ii})T_{ii} - \sum_{k\neq i} M_{ik}T_{ik} = 1, \\ \frac{1}{2}(M_{ii}+M_{jj})T_{ij} - \sum_{k\neq i}\frac{M_{ik}}{2}T_{jk} - \sum_{k\neq i}\frac{M_{jk}}{2}T_{ik} = 1, \end{cases} \quad [S1]$$

with $i \neq j$ and $M_{ii} = -\sum_{k\neq i} M_{ik}$ for $i = 1, ..., K$. Eq. S1 is defined if all T_{ij} are finite, which is true if and only if the migration network is connected (i.e., consists of a single component). When M can be decomposed into several network components, Eq. S1 is defined within each component.

Using Eq. S1, we track the genetic diversity of each population and compute its expected heterozygosity (2, 3):

$$H_i = (\theta_i T_{ii})/K, \quad [S2]$$

where $\theta_i = 4N\mu$ is the scaled mutation rate per site per generation under the infinite-site model (3). Because we assume equal population size and a fixed mutation rate for all populations, we simplify the notation and study the 'unscaled expected heterozygosity' $H_i = T_{ii}/K$. In other words, heterozygosity is the coalescence time for two individuals within each population normalized by the total number of populations in the network. In our model, H_i can exceed 1, typically in populations that act as central gene flow hubs in the network. This is because the probability of two individuals in the central population to coalesce decreases, while the probability of the individuals to drift apart from one another increases. This pattern contrasts with the fully connected island model, where individuals are always within one migration event for each pair of populations. Thus, heterozygosity in our framework should be interpreted in relative terms.

Next, we formulate a transformation from the coalescent times matrix T to the F_{ST} matrix F, whose entry F_{ij} is a pairwise F_{ST} (4) value between populations i and j. For low mutation rates, F can be approximated from T using a set of non-linear equations (2, 5):

$$F_{ij} = \frac{T_T^{ij} - T_S^{ij}}{T_T^{ij}}, \quad [S3]$$

where $T_S^{ij} = (T_{ii} + T_{jj})/2$ is the expected within-population coalescence time, and $T_T^{ij} = (T_{ij} + T_S^{ij})/2$ is the expected coalescence time of two individuals sampled from these two populations. By definition, F is symmetric, with zero diagonal entries. Eq. S3 is derived under the assumption of small mutation rates, which was shown to be valid under many realistic scenarios (6). Following Eq. S1, Eq. S3 is also defined only for populations in the same component of M. For two populations i and j in different components of M, we set $F_{ij} = 1$ because there is no gene flow between the populations and they are therefore maximally differentiated.

Robustness analysis. To evaluate the robustness of our results to model assumptions, we repeated the analysis while varying key parameters. Specifically, we examined the effect of changing the (i) initial network topology, (ii) mean migration rate, (iii) variance of migration rates among edges, and (iv) variance of population sizes.

Initial network topology. To account for alternative patterns of gene flow in our initial network, we considered two additional models. (i) The Erdős–Rényi (ER) random network (7), in which, for K populations, each pair of populations is connected by an edge with probability p. To generate a well- but not fully-connected initial network, we set $p = 0.1$ (125 edges in total). To allow spatially explicit analysis that can be compared to the RGG, we embedded this model in Euclidean space, with nodes placed uniformly at random. (ii) The small-world Watts–Strogatz (WS) network is constructed by connecting each population to its g nearest neighbors in a grid topology, and then rewiring each edge with probability p to connect to a randomly chosen population, introducing long-range connections while preserving the total number of edges. The WS network can represent species with a life history of many short-distance dispersal events and an occasional long-distance dispersal event. We use a modified variant of this model to incorporate spatial characteristics into the network (8, 9) and apply the `grid_graph` function in the Python library NetworkX (10) to generate a two-dimensional network with $K = 49$ nodes (a 7×7 matrix), setting $g = 4$ (4 neighbors for each node), and a re-wiring probability of $p = 0.015$. This setting converges to the stepping stone model (11) for $p = 0$.

Mean migration rate. To examine the effect of the mean migration rate on our results, which was $M = 1$ in our main analyses, we considered models with alternative mean migration rates $M = 0.2$ and $M = 5$.

Variance of migration rates. To examine the effect of unequal migration rates on our results, we varied migration rates to different extents while preserving the same mean $M = 1$. For this purpose, we treat the network as a directed network, allowing migration rate $M_{ij} \neq M_{ji}$. After generating a network M using the RGG model, we assign a weight to each directed edge by drawing from a Normal distribution $M_{ij} \sim \mathcal{N}(1, \sigma^2)$, and we consider different values of σ (0.1, 0.3 and 0.5). We

clipped this Normal distribution at a lower bound of 0.1 and upper bound of 4 to ensure reasonable migration rates between connected nodes.To preserve the conservative migration constraint required for using the transformation from migration rates to coalescence times, we apply the following calibration procedure. We first enforce node-wise conservativeness (total emigration equals total immigration per population), to allow directional asymmetry ($M_{ij} \neq M_{ji}$) while preserving demographic balance at every population. Let E be the number of directed edges in the network. We index edge rates by a vector $\mathbf{x} \in \mathbb{R}^E$ and define the node–edge incidence matrix $A \in \mathbb{R}^{K \times E}$ by assigning $+1$ in row i and -1 in row j to the column corresponding to edge $i \to j$, with zeros elsewhere. A vector $\mathbf{m}$ is conservative if $A\mathbf{m} = \mathbf{0}$. We then project $\mathbf{x}$ orthogonally onto the null space of A,

$$P \;=\; I_E \;-\; A^{\top}\big(AA^{\top}\big)^{+}A, \qquad \mathbf{m}^{(1)} \;=\; P\,\mathbf{x},$$

where $(\cdot)^{+}$ denotes the Moore–Penrose pseudoinverse and I_E is the $E \times E$ identity matrix. If any entry of $\mathbf{m}^{(1)}$ falls outside our bounds ($M_{ij} = 0.1$ or $M_{ij} = 4$), we clip to the nearest bound and re-project once more,

$$\mathbf{m} \;=\; P\left[\mathbf{m}^{(1)}\right]_{\text{clipped}}.$$

The resulting conservative vector $\mathbf{m}$ is then placed in the off-diagonal entries of the migration matrix M. During fragmentation, deleting an undirected connection removes both directed edges $i \to j$ and $j \to i$; we then re-impose conservativeness on the remaining edges by applying the same projector with the updated incidence matrix A. This guarantees conservative migration ($A\mathbf{m} = \mathbf{0}$) at every step of the fragmentation while minimally perturbing remaining rates within the prescribed bounds. We first ran 100 simulation replicates to examine the relationship between the standard deviation of the migration rates (σ) and its corresponding H_e and pairwise F_{ST} (Fig. S7). To account for the weighted directed networks in the correlations analysis (Fig. 5), for betweenness centrality, shortest path distance, and random walk distance calculations, pairwise distance was first converted to edge lengths $\ell_{ij} = 1/M_{ij}$ (so higher migration implies shorter distance).

Variance of population sizes. To examine the effect of unequal population sizes on our results, we varied population sizes to different extents while preserving the same mean $N = 100$. For a given network, the population size for each node was drawn from a normal distribution $N_i \sim \mathcal{N}(100, \sigma^2)$, which was clipped at $N = 10$ and $N = 500$, and we considered different values of σ (10, 30 and 50). The relative population size for population i was defined as $c_i = \frac{N_i}{N}$, so that $mean(\mathbf{c}) = 1$. To preserve the conservative migration constraint required for using the transformation from migration rates to coalescence times, we also need to apply a calibration procedure, which requires treating the migration matrix as a weighted directed network, as when varying migration rates. In the case of unequal population sizes, the conservative migration constraint is $c_i \sum_{j \neq i} M_{ij} = \sum_{j \neq i} c_j\, M_{ji}$ (12). We therefore define a weighted directed network M^w by column scaling the undirected network (assuming all edge weights are $M_{ij} = 1$ for each pair of connected populations i and j), i.e., $M^w_{ij} = c_j M_{ij}$. This guarantees size-weighted conservativeness in every deme i, $c_i \sum_j M^w_{ij} = \sum_j c_j M^w_{ji}$. Because our fragmentation processes remove edges as undirected pairs $i \leftrightarrow j$, and we have$M^w_{ij} = c_j M_{ij}$, each fragmentation step preserves the conservativeness property at each step. At each step, coalescence times were then obtained by solving Eq. S4 with a deme-specific within-deme coalescence rate $1/c_i$ on the diagonal equations. We first ran 100 simulation replicates to examine the relationship between the standard deviation of the population sizes (σ) and its corresponding H_e and pairwise F_{ST} (Fig. S7). For the correlation analysis between the genetic measures and network metrics, we used the same adjustments as described above for varying migration rates.

For this part of unequal population sizes, we solve the modified Eq. S1 with a deme-specific within-deme coalescence rate $\frac{1}{c_i}$ on the diagonal equations (1, 12)

$$\begin{cases} (\dfrac{1}{c_i} + M_{ii})T_{ii} - \displaystyle\sum_{k \neq i} M_{ik}T_{ik} = 1, \\ \dfrac{1}{2}(M_{ii} + M_{jj})T_{ij} - \displaystyle\sum_{k \neq i} \frac{M_{ik}}{2}T_{jk} - \sum_{k \neq i} \frac{M_{jk}}{2}T_{ik} = 1. \end{cases} \qquad \text{[S4]}$$

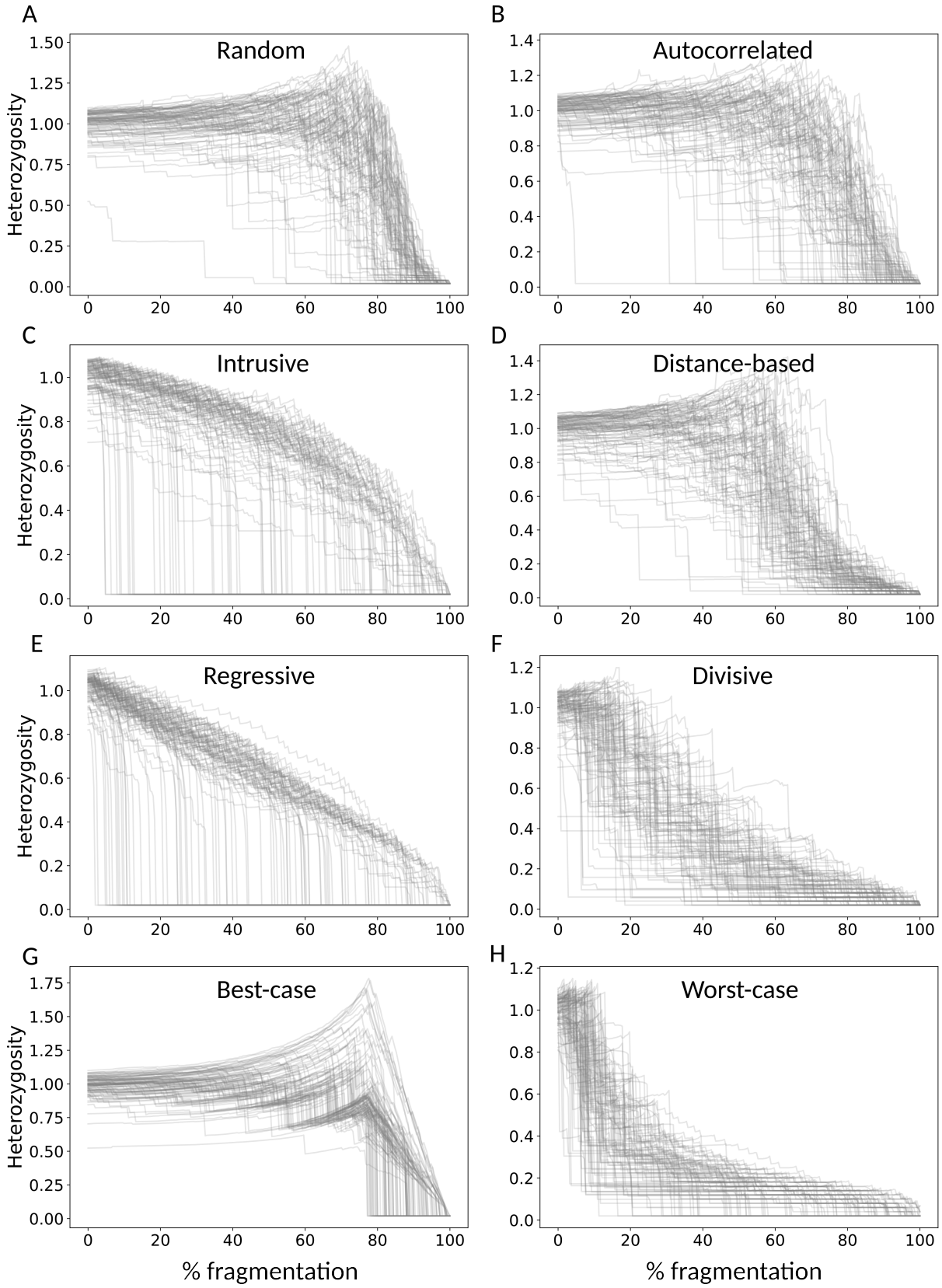

Fig. S1. Trajectories of H_e throughout fragmentation. (A-H) 100 randomly sampled populations across replicates for each fragmentation scenario, corresponding to the analysis in Fig. 2.

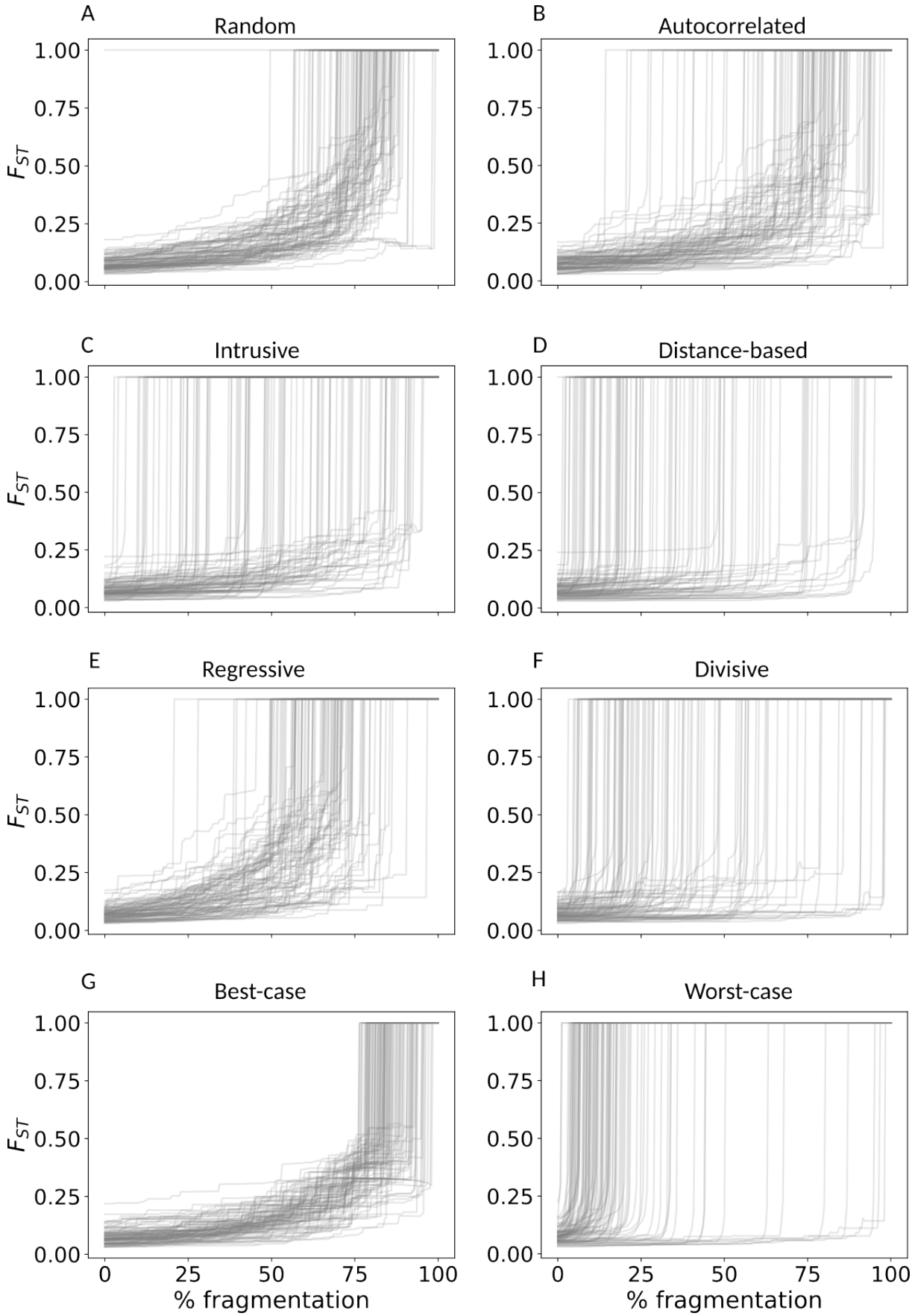

Fig. S2. Trajectories of F_{ST} throughout fragmentation. (A-H) 100 randomly sampled populations across replicates for each fragmentation scenario, corresponding to the analysis in Fig. 2.

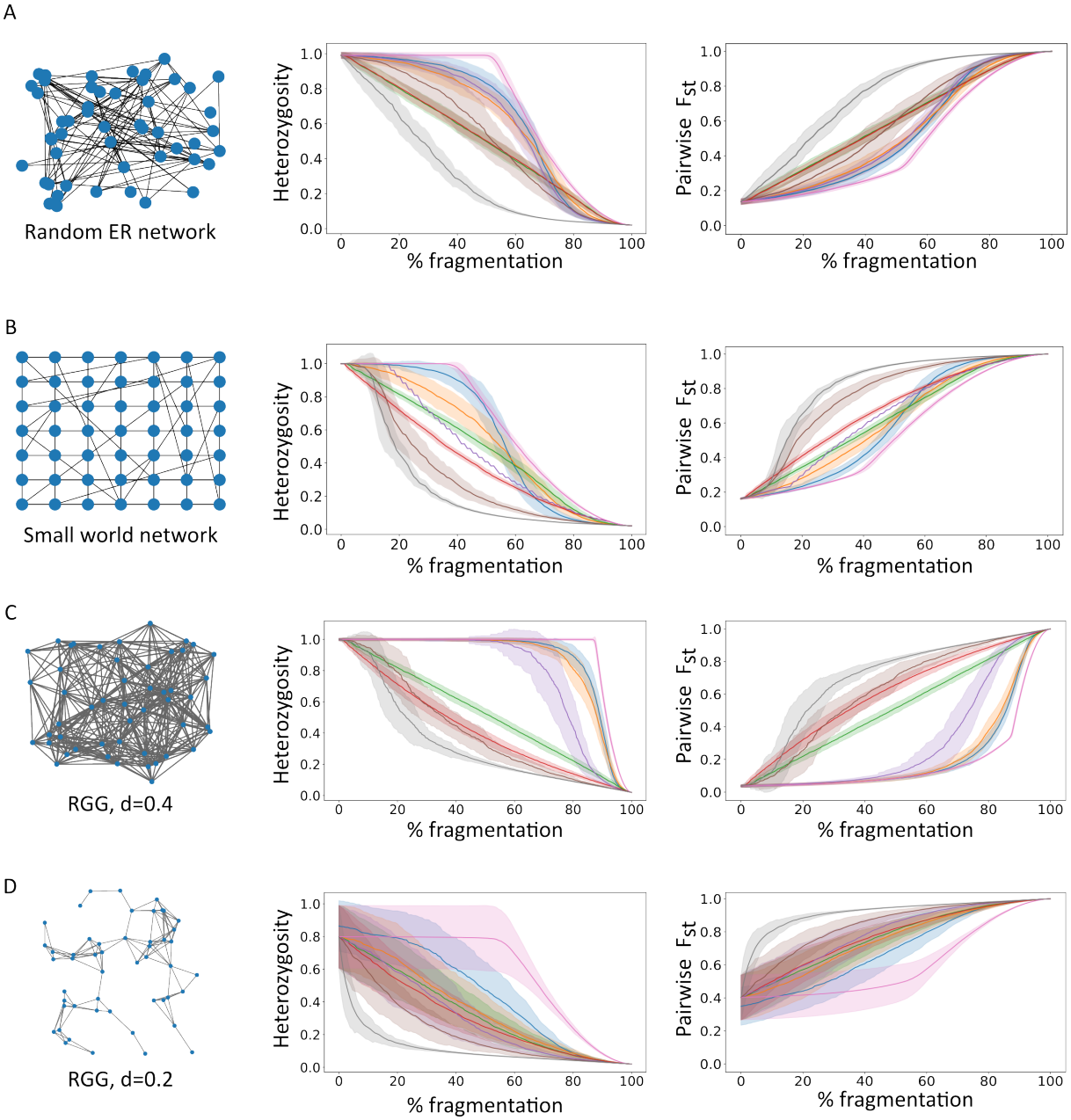

Fig. S3. Changes in genetic measures during fragmentation under alternative initial networks. (A) Mean heterozygosity and F_{ST} for an ER network in which nodes are connected to each other at random with probability $p = 0.3$. (B) Mean heterozygosity and F_{ST} for spatial small-world networks in which nodes are connected to their closest neighbors ($g = 4$) with additional random connections at probability $p = 0.015$. (C) Mean heterozygosity and F_{ST} for RGG networks with $d = 0.4$. (D) Mean heterozygosity and F_{ST} for RGG networks with $d = 0.2$. All networks shown on the left depict a single realization, for demonstration of the topology. The curves denote the mean across 100 simulation replicates (both networks and fragmentation were simulated in each repeat), and the shaded areas denote the standard deviation. Colors as in Fig. 2.

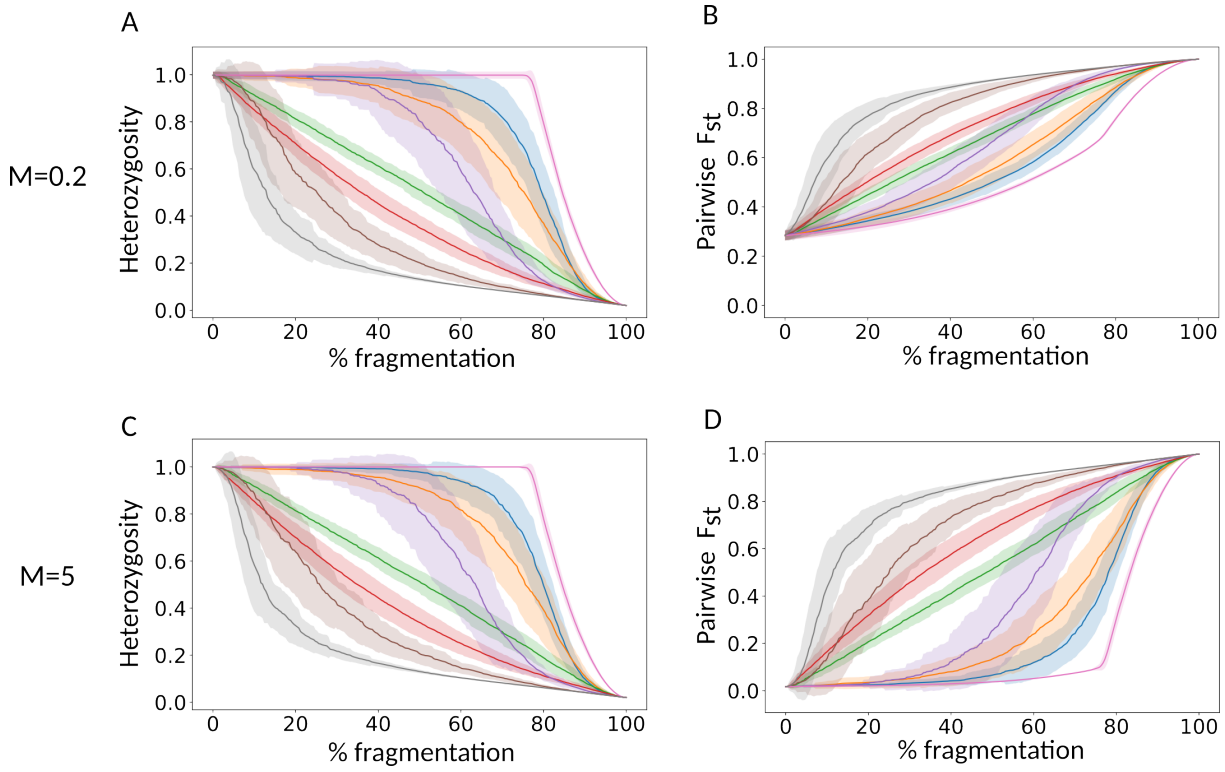

Fig. S4. Changes in genetic measures during fragmentation under alternative migration rates (M). (A–B) Mean heterozygosity and F_{ST} under low migration rates $M = 0.2$. (C–D) Mean heterozygosity and F_{ST} under high migration rates $M = 5$. Results in the main text are for $M = 1$. The curves denote the mean across 100 simulation replicates, and the shaded areas denote the standard deviation. Colors as in Fig. 2.

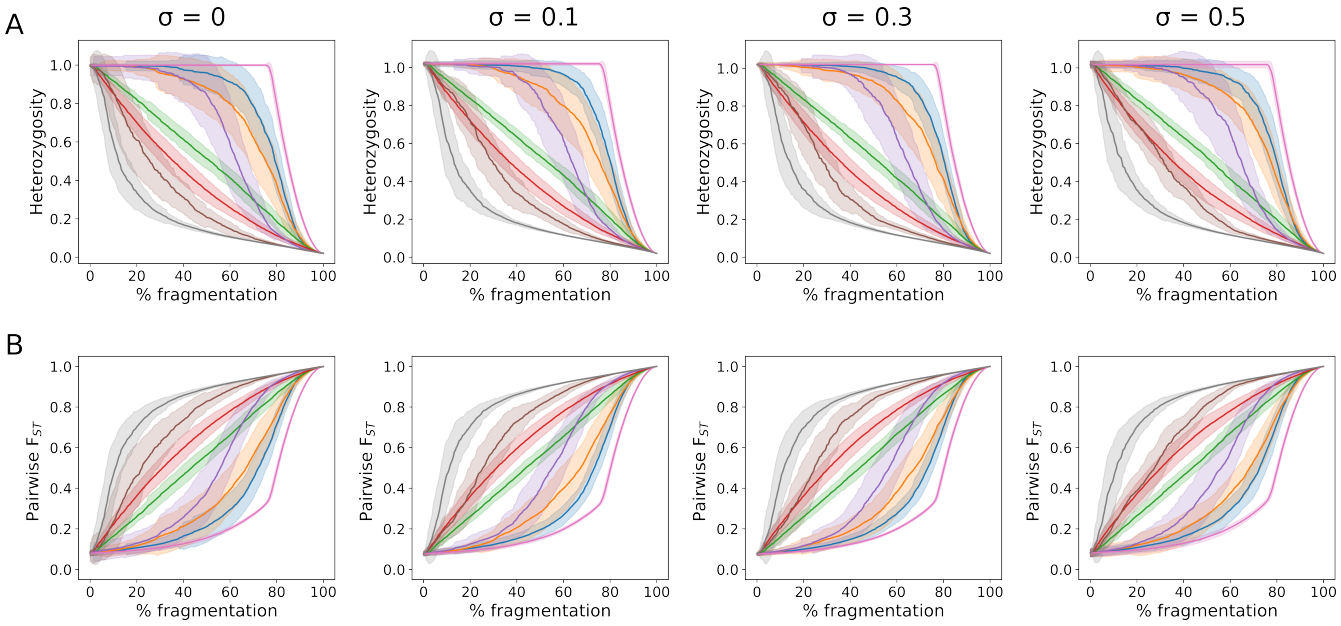

Fig. S5. Robustness analysis for migration networks with unequal migration rates. Migration rates are sampled from a Normal distribution with mean $M = 1$ and different standard deviations (σ). The curves denote the mean across 100 simulation replicates and the shaded areas denote the standard deviation. Colors as in Fig. 2 in the main text. The $\sigma = 0$ column is identical to Figure 2 in the main text. (A) Mean heterozygosity along fragmentation. (B) Mean pairwise F_{ST} along fragmentation.

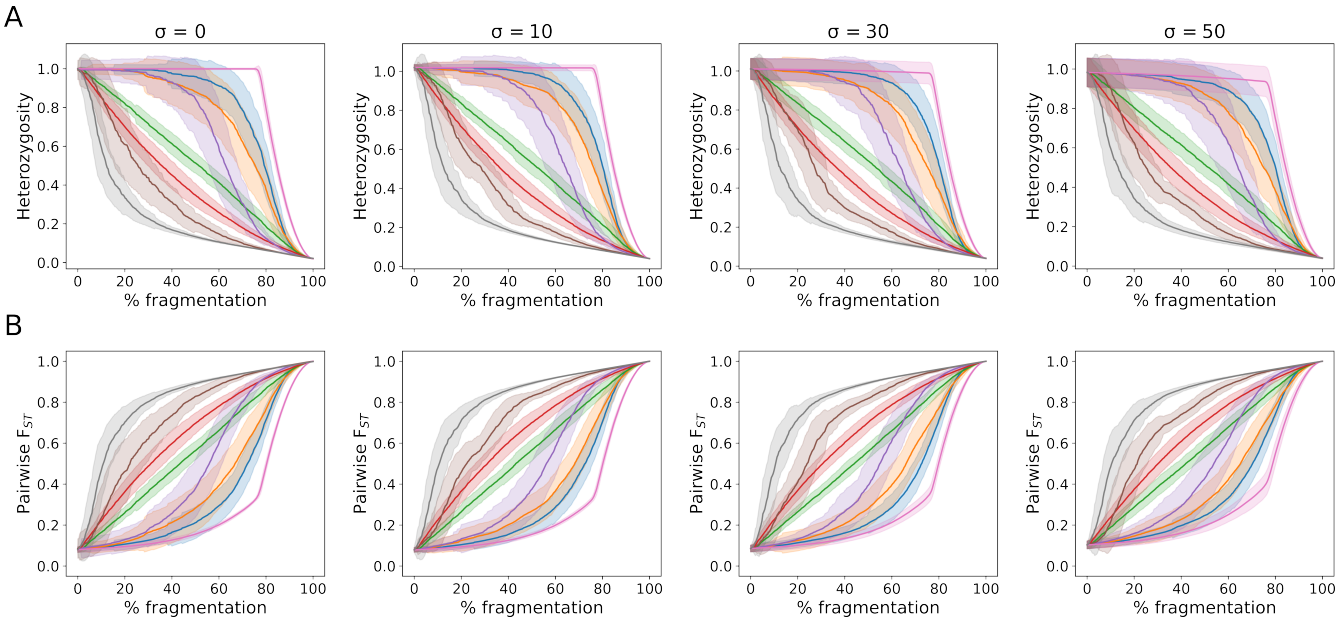

Fig. S6. Robustness analysis for migration networks with unequal population sizes. Population sizes are sampled from a Normal distribution with mean $N = 100$ and different standard deviations (σ). The curves denote the mean across 100 simulation replicates and the shaded areas denote the standard deviation. Colors as in Fig. 2 in the main text. The $\sigma = 0$ column is identical to Figure 2 in the main text. (A) Mean heterozygosity along fragmentation. (B) Mean pairwise F_{ST} along fragmentation.

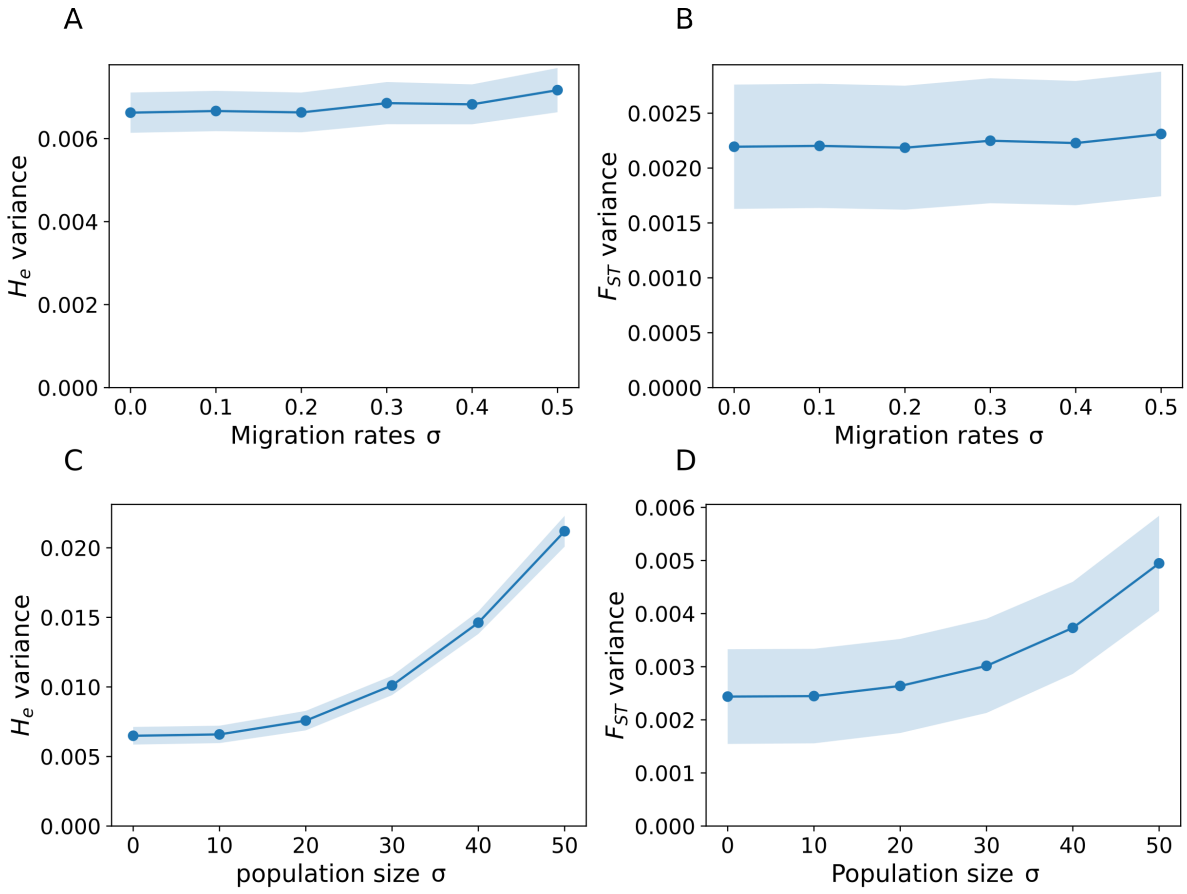

Fig. S7. The effect of standard deviation (σ) in migration rates and population sizes on the variance of genetic measures. (A) The variance in H_e among populations for different levels of migration rate σ. (B) The variance in pairwise F_{ST} among population pairs for different levels of migration rates σ. (C) The variance in H_e among populations for different levels of population size σ. (D) The variance in pairwise F_{ST} among population pairs for different levels of population size σ. Blue dots show the mean variance from 100 simulation replicates, and the shaded area shows the standard error of the variance from the simulation replicates. The analysis was conducted on the same RGG model ($n = 50$ and $d = 0.3$) used in the main text .

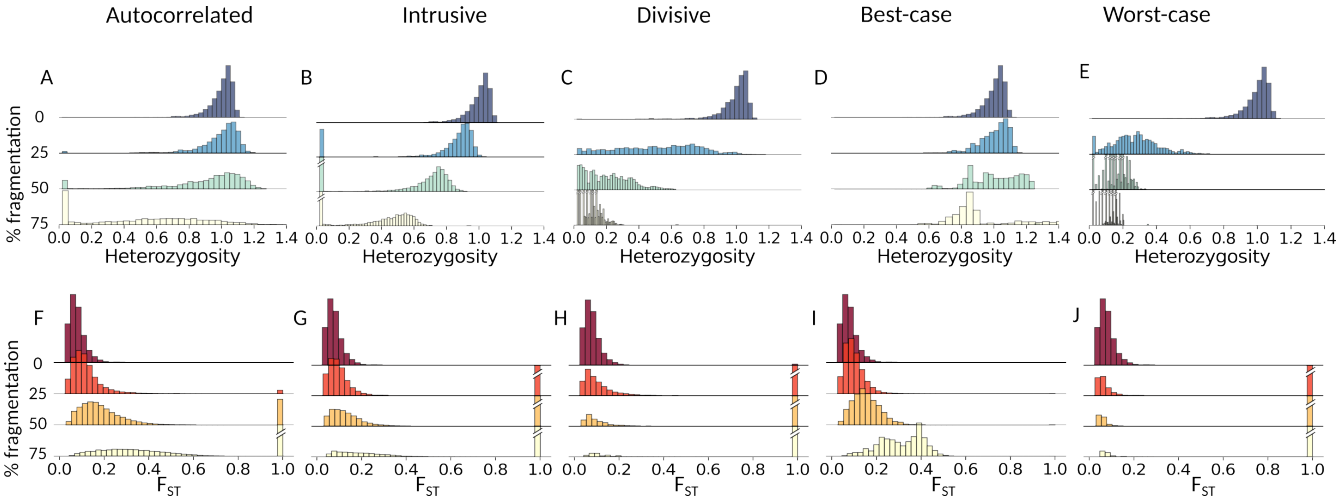

Fig. S8. Changes in distribution of genetic measures along fragmentation. Presented here are the remaining five fragmentation scenarios that are not included in the main text: autocorrelated, intrusive, divisive, best-case, and worst-case. Shown are four snapshots during the process, with 0, 25, 50, and 75% of the edges removed. Diagonal lines on bars represent high values that were truncated for better visualization. (A–E) Distribution of expected heterozygosity (H_e) within populations. (F–J) Distribution of pairwise F_{ST} values. $F_{ST} = 1$ indicates pairs of nodes that are not connected by any path. In all panels, distributions are pooled across 100 simulation replicates.

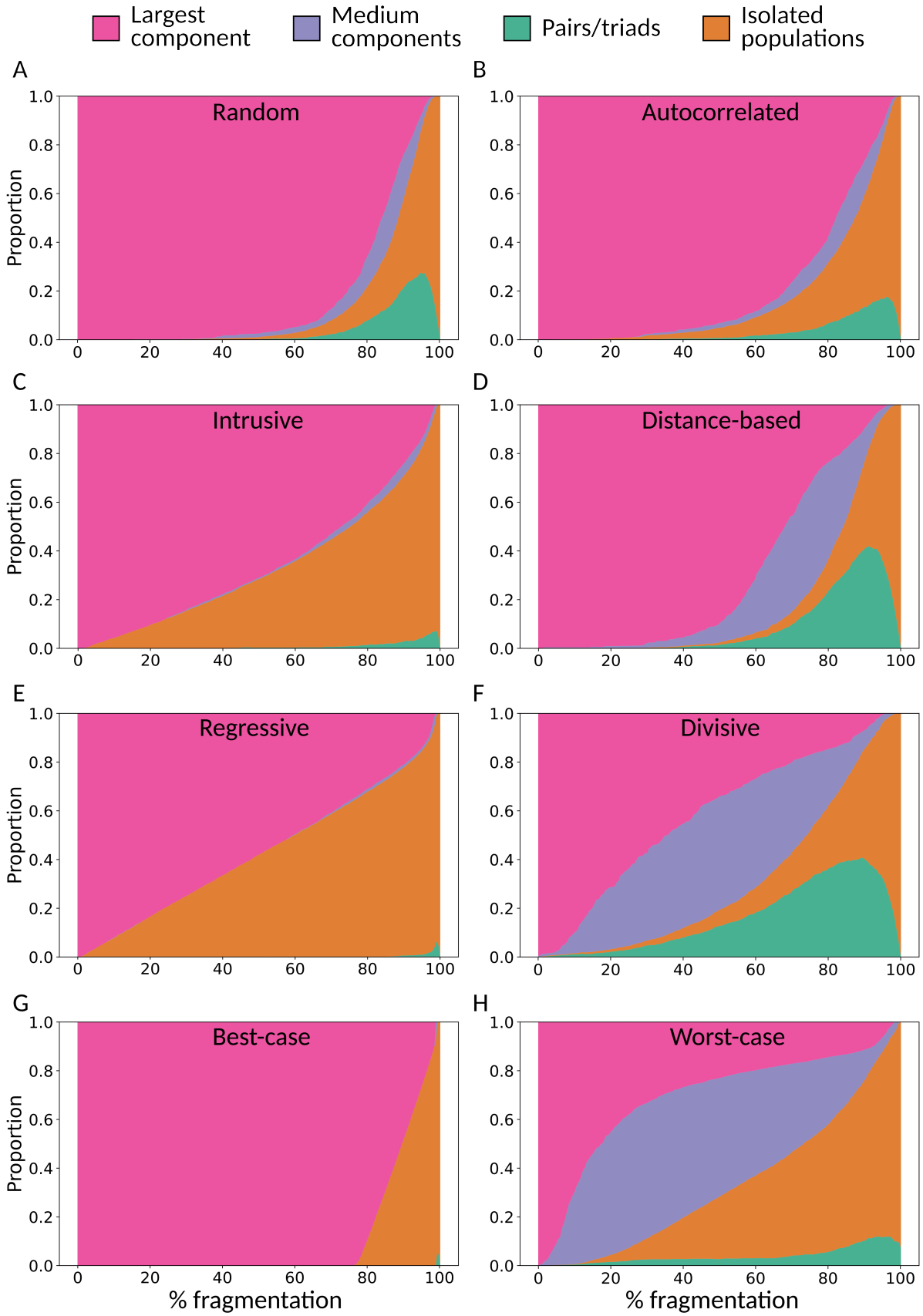

Fig. S9. The proportions of different network structures throughout eight fragmentation scenarios. Largest component–the largest set of connected nodes in the network; medium components–a connected set of $>= 4$ nodes, excluding the largest component; pairs/triads–two or three connected nodes; isolated populations–nodes without edges. For each scenario A–H we show the mean across all simulation replicates.

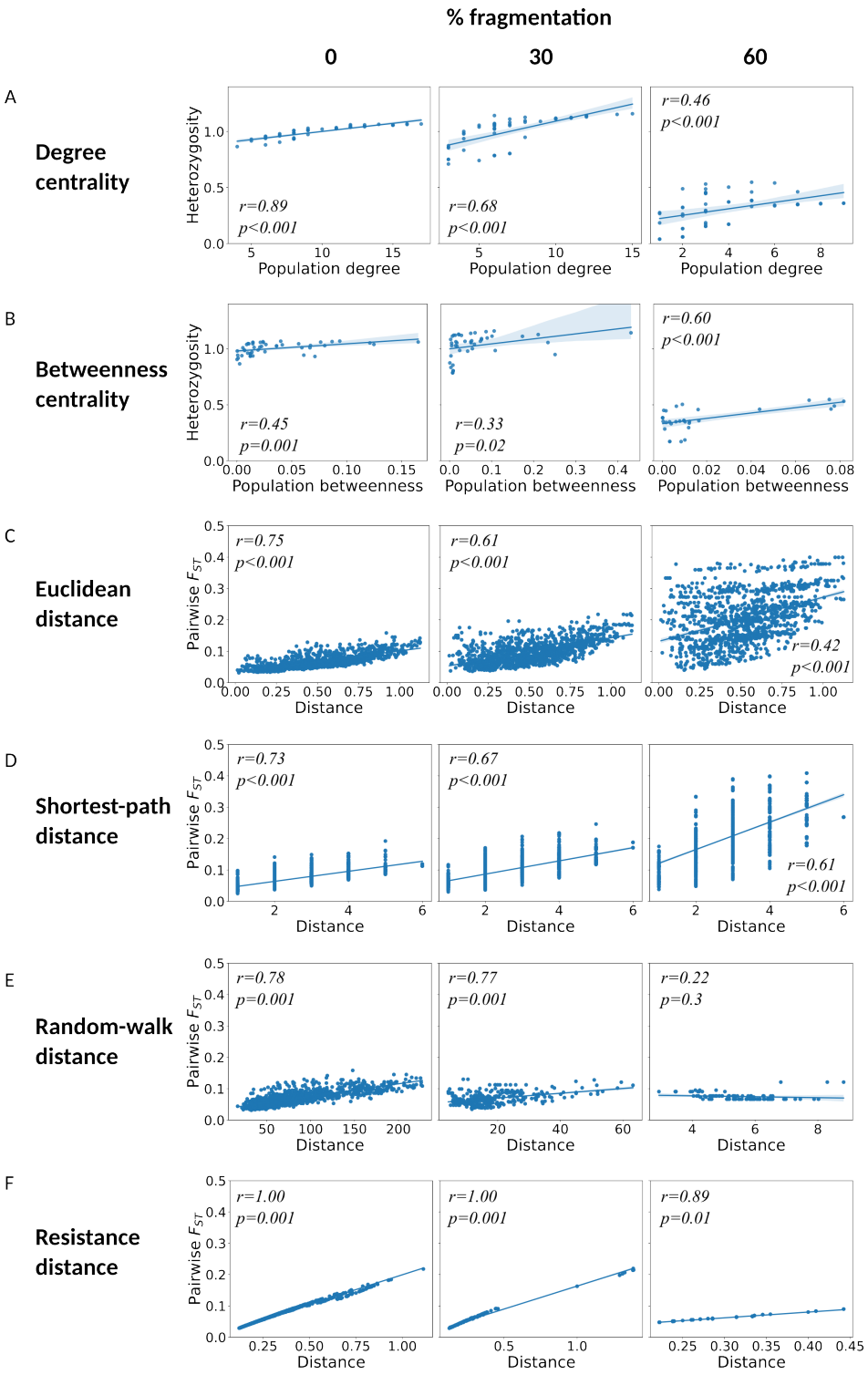

Fig. S10. Snapshots of correlation between population genetic measures and network metrics. The Regression plots show snapshots at 0%, 30%, and 60% level of fragmentation. r and $p - value$ are shown inline. (A–B) Correlation under distance-based scenario, (C–D) Correlation under best-case scenario. (E–F) Correlation under worst-case scenario.

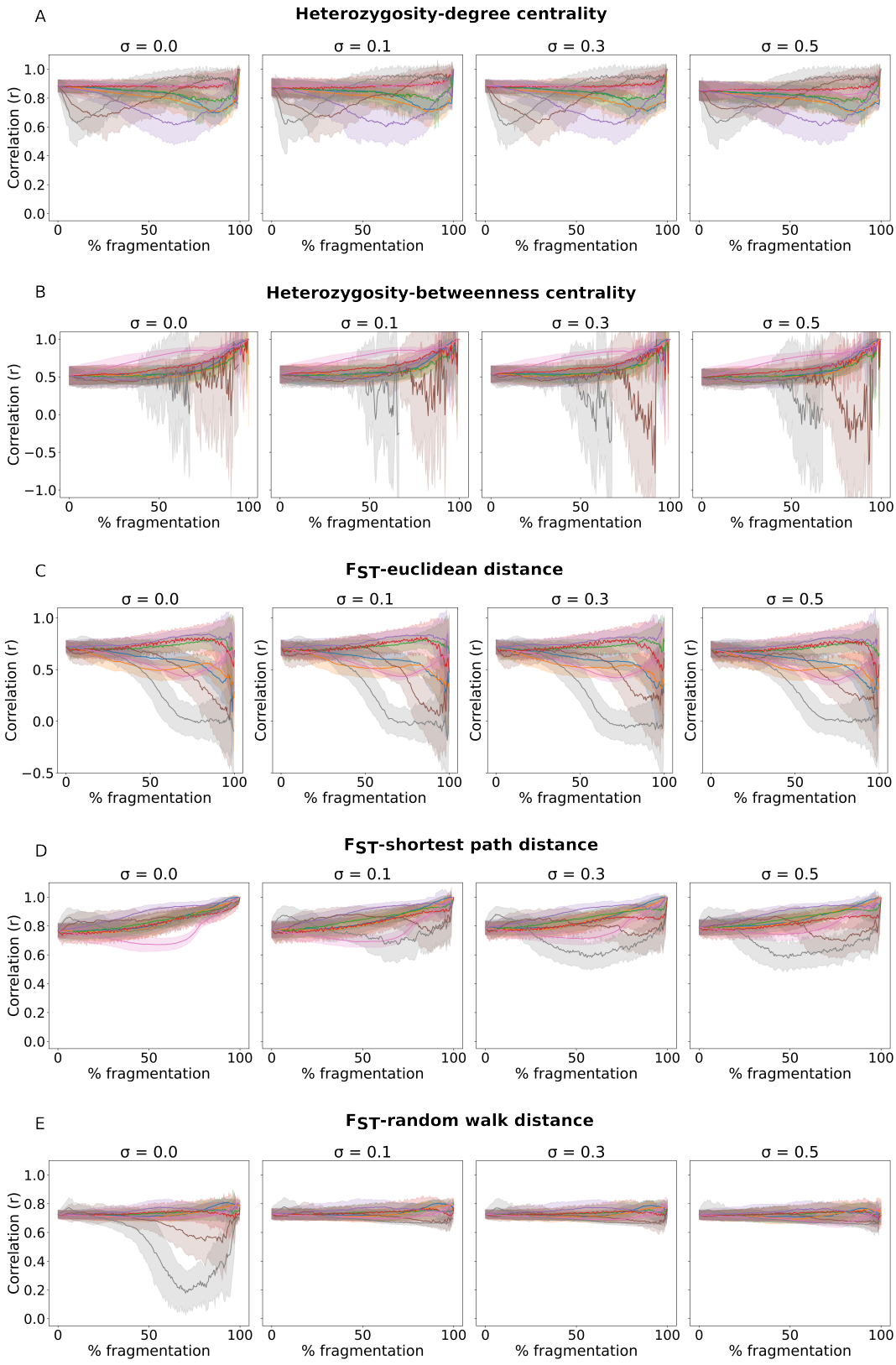

Fig. S11. Robustness analysis of changes in correlations along fragmentation for unequal migration rates. Migration rates are sampled from a Normal distribution with mean $M = 1$ and different standard deviations (σ). The $\sigma = 0$ column shows the results presented in Figure 5 in the main text. (A) Correlation between a population's H_e and its degree centrality. (B) Correlation between a population's H_e and its betweenness centrality. (C) Correlation between the F_{ST} of a pair of populations and their Euclidean distance. (D) Correlation between the F_{ST} of a pair of populations and their shortest-path distance. (E) Correlation between the F_{ST} of a pair of populations and their random-walk distance. Robustness analysis of the resistance metric under unequal migration rates was not performed because the theory from which this metric is derived requires balanced migration ((13)). Colors as in Fig. 2 in the main text.

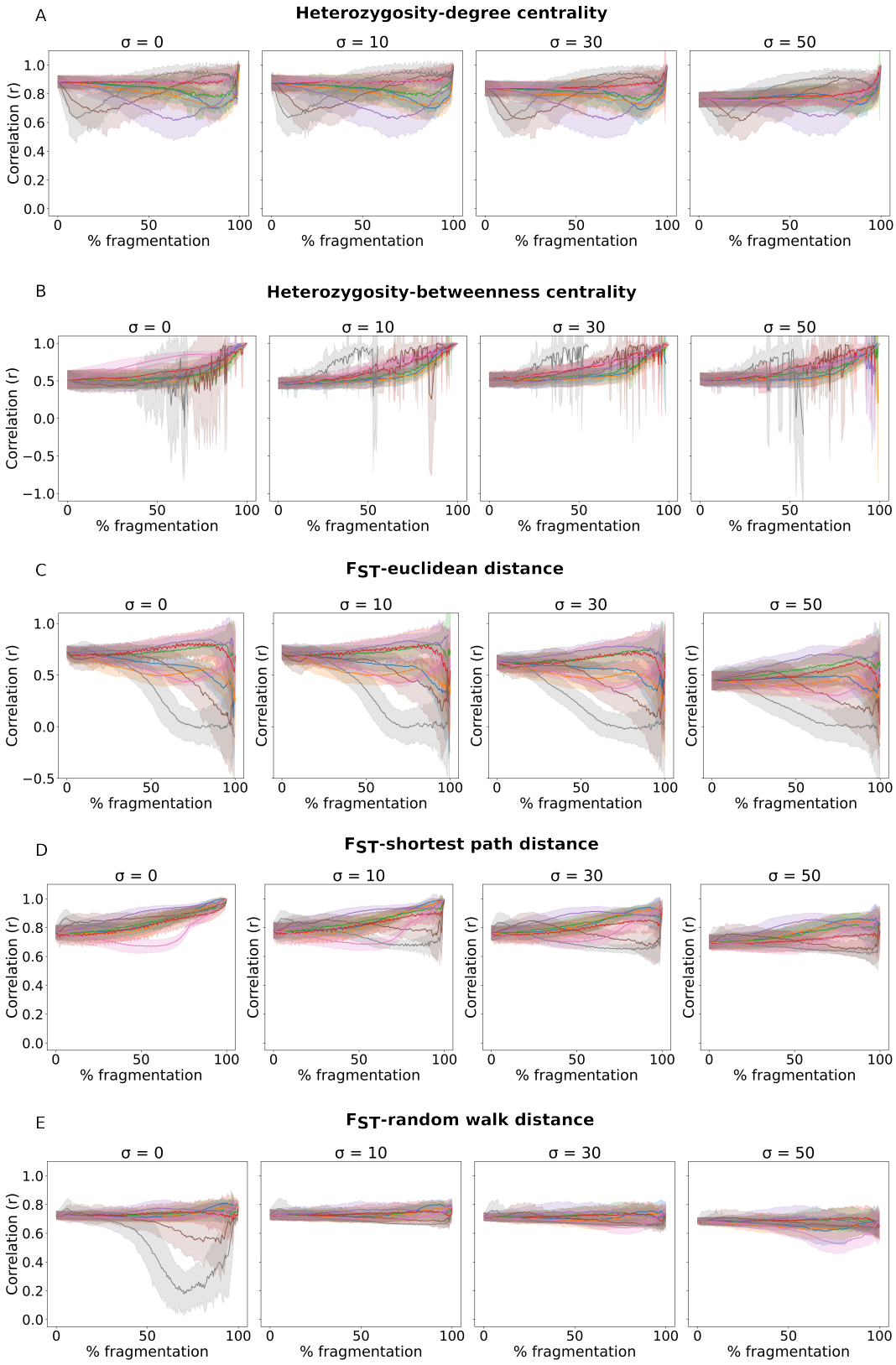

Fig. S12. Robustness analysis of changes in correlations along fragmentation for unequal population sizes. Population sizes are sampled from a Normal distribution with mean $N = 100$ and different standard deviations (σ). The $\sigma = 0$ column shows the results presented in Figure 5 in the main text. (A) Correlation between a population's H_e and its degree centrality. (B) Correlation between a population's H_e and its betweenness centrality. (C) Correlation between the F_{ST} of a pair of populations and their Euclidean distance. (D) Correlation between the F_{ST} of a pair of populations and their shortest-path distance. (E) Correlation between the F_{ST} of a pair of populations and their random-walk distance. Robustness analysis of the resistance metric under unequal migration rates was not performed because the theory from which this metric is derived requires balanced migration ((13)). Colors as in Fig. 2 in the main text.

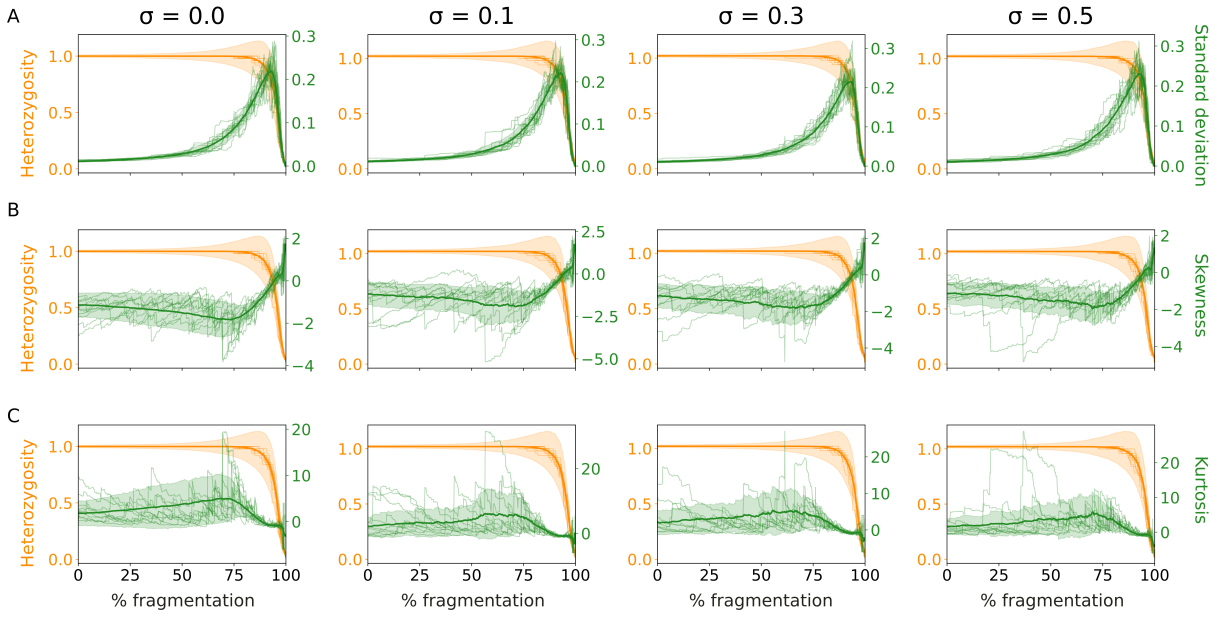

Fig. S13. Robustness analysis for early warning signals of monitoring a metapopulation with unequal migration rates. Shown is the mean metapopulation heterozygosity (in orange) and three early warning statistics (in green; standard deviation in (A), skewness in (B), and kurtosis in (C)) along fragmentation. Solid lines show the mean across 100 simulation replicates, shaded areas show the standard deviation, and thin lines show ten individual replicates. The $\sigma = 0$ column shows the results presented in Figure 6 in the main text.

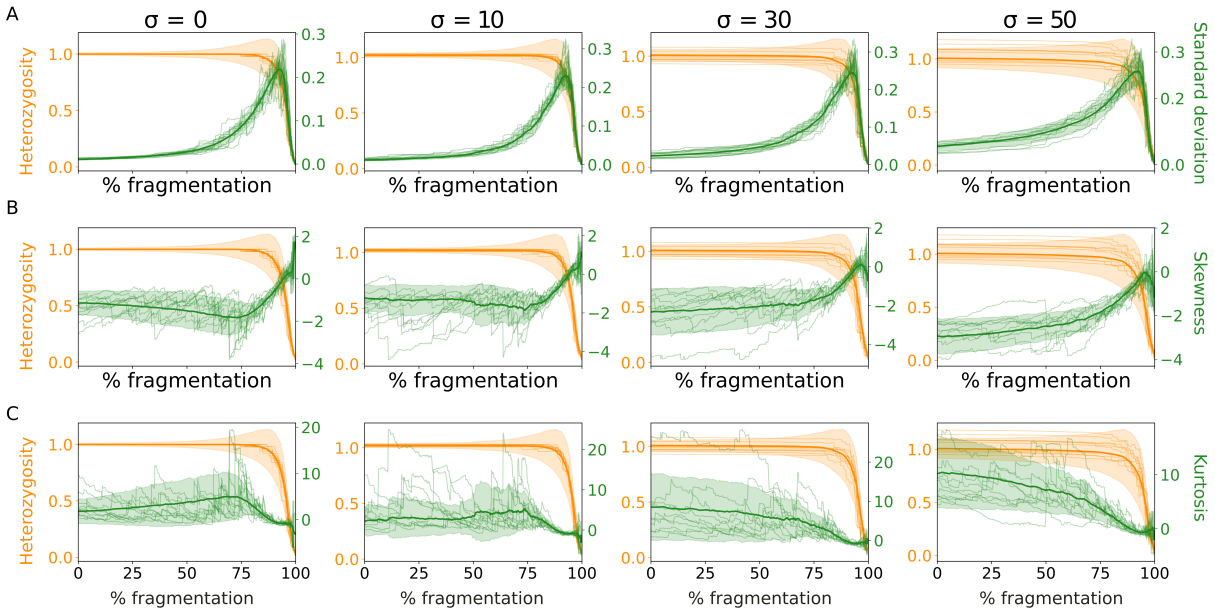

Fig. S14. Robustness analysis for early warning signals of monitoring a metapopulation with unequal population sizes. Shown is the mean metapopulation heterozygosity (in orange) and three early warning statistics (in green; standard deviation in (A), skewness in (B), and kurtosis in (C)) along fragmentation. Solid lines show the mean across 100 simulation replicates, shaded areas show the standard deviation, and thin lines show ten individual replicates. The $\sigma = 0$ column shows the results presented in Figure 6 in the main text.

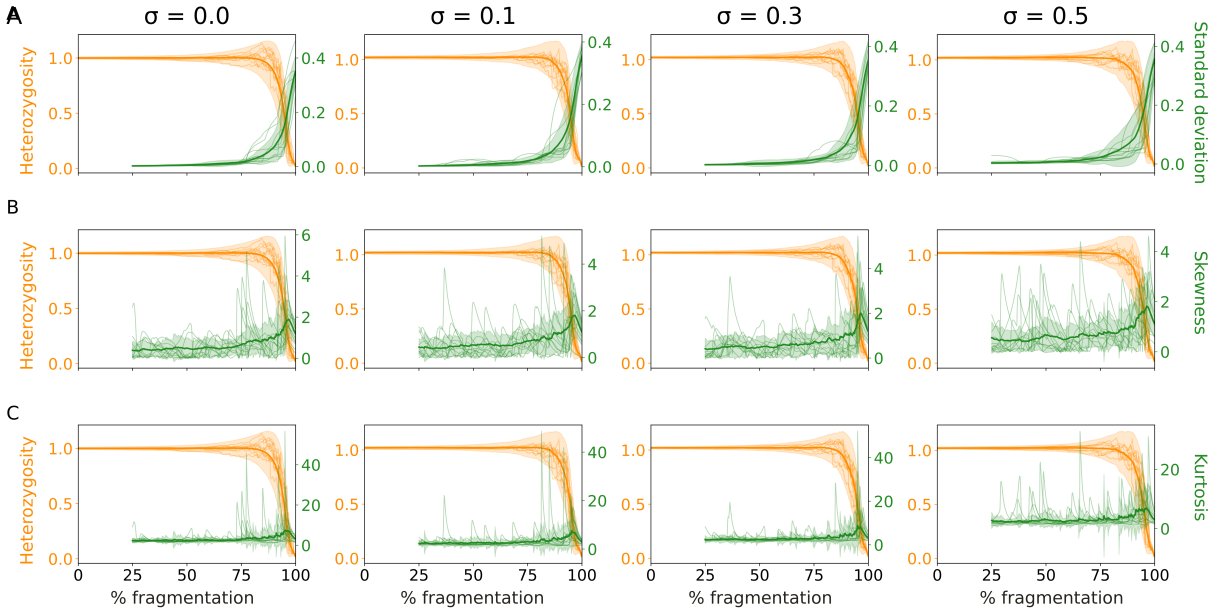

Fig. S15. Robustness analysis for early warning signals of monitoring a single population with unequal migration rates. Shown is the mean single population heterozygosity (in orange) and three early warning statistics (in green; standard deviation in (A), skewness in (B), and kurtosis in (C)) along fragmentation. Solid lines show the mean across 100 simulation replicates, shaded areas show the standard deviation, and thin lines show ten individual replicates. The $\sigma = 0$ column shows the results presented in Figure 6 in the main text.

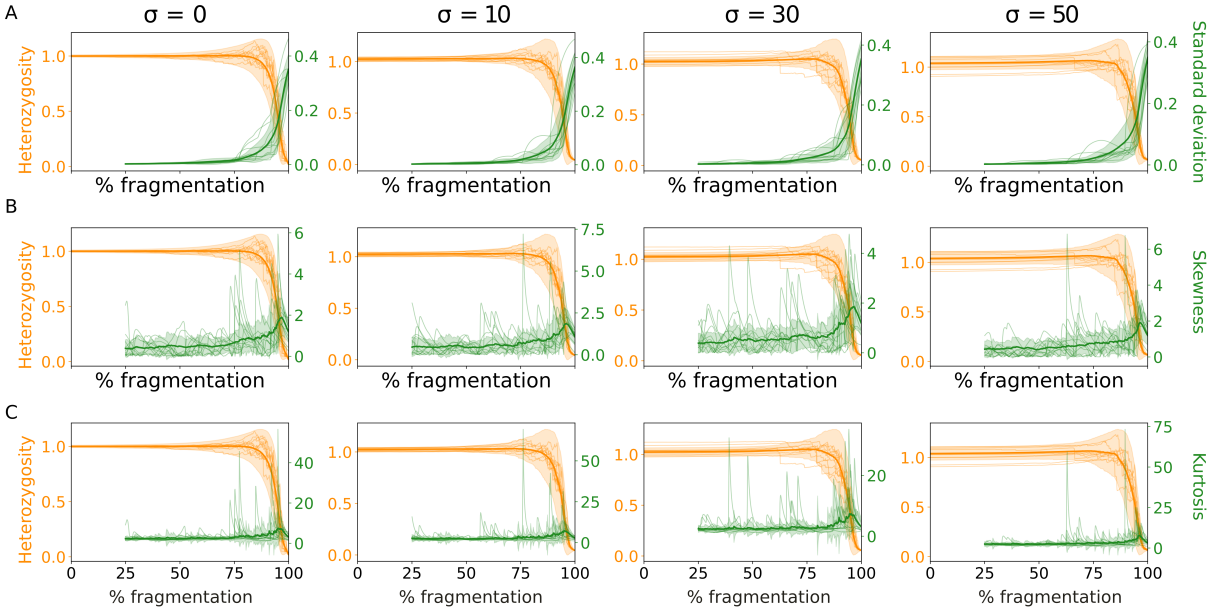

Fig. S16. Robustness analysis for early warning signals of monitoring a single population with unequal population sizes. Shown is the mean single population heterozygosity (in orange) and three early warning statistics (in green; standard deviation in (A), skewness in (B), and kurtosis in (C)) along fragmentation. Solid lines show the mean across 100 simulation replicates, shaded areas show the standard deviation, and thin lines show ten individual replicates. The $\sigma = 0$ column shows the results presented in Figure 6 in the main text.

References

1. HM Wilkinson-Herbots, Genealogy and subpopulation differentiation under various models of population structure. *J. Math. Biol.* **37**, 535–585 (1998).
2. N Alcala, A Goldberg, U Ramakrishnan, NA Rosenberg, Coalescent theory of migration network motifs. *Mol. Biol. Evol.* **36**, 2358–2374 (2019).
3. M Kimura, The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893 (1969).
4. S Wright, The genetical structure of populations. *Annals Eugen.* **15**, 323–354 (1949).
5. M Slatkin, Inbreeding coefficients and coalescence times. *Genet. Res.* **58**, 167–175 (1991).
6. M Slatkin, Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264–279 (1993).
7. P Erdős, A Rényi, , et al., On the evolution of random graphs. *Publ. Math. Inst. Hungarian Acad. Sci.* **5**, 17–60 (1960).
8. P Sen, K Banerjee, T Biswas, Phase transitions in a network with a range-dependent connection probability. *Phys. Rev. E* **66**, 037102 (2002).
9. ME Newman, DJ Watts, Scaling and percolation in the small-world network model. *Phys. Rev. E* **60**, 7332 (1999).
10. A Hagberg, PJ Swart, DA Schult, Exploring network structure, dynamics, and function using networkx, (Los Alamos National Laboratory (LANL), Los Alamos, NM (United States)), Technical report (2008).
11. M Kimura, GH Weiss, The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561 (1964).
12. HM Wilkinson-Herbots, Coalescence times and fst values in subdivided populations with symmetric structure. *Adv. Appl. Probab.* **35**, 665–690 (2003).
13. BH McRae, Isolation by resistance. *Evolution* **60**, 1551–1561 (2006).