How does dispersal between urban green spaces and forests shape genetic patterns in European cities? A simulation approach
Supporting information

Savary, Paul*[1], Tannier, Cécile[2], Foltête, Jean-Christophe[2], Bourgeois, Marc[3], Vuidel, Gilles[2], Khimoun, Aurélie[4], Moal, Hervé[5], and Garnier, Stéphane[4]

[1] *Department of Biology, Concordia University, Montreal (QC), Canada*
[2] *UMR 6049 ThéMA, Université de Franche-Comté - CNRS - Besançon, France*
[3] *UMR 5600 Environnement Ville Société, Université Jean Moulin Lyon 3 - CNRS - Lyon, France*
[4] *UMR 6282 Biogéosciences, Université de Bourgogne - CNRS - Dijon, France*
[5] *ARP-Astrance - Paris, France*

1 Supplementary tables

New land cover types	Initial land cover types
Forests	Forests
Wetlands	Wetlands
Grasslands	Pastures, Herbaceous vegetation associations (natural grassland moors)
Semi-natural areas	Orchards at the fringe of urban classes
Urban green spaces	Green urban areas
Agricultural areas	Arable land (annual crops), Permanent crops (vineyards fruit trees olive groves), Complex and mixed cultivation patterns
Water	Water bodies
Roads	Fast transit roads and associated land, Other roads and associated land, Railways and associated land
Artificial areas	Continuous Urban Fabric (S.L. : 80 %), Discontinuous Dense Urban Fabric (S.L. : 50 % - 80 %), Discontinuous Medium Density Urban Fabric (S.L. : 30 % - 50 %), Discontinuous Low Density Urban Fabric (S.L. : 10 % - 30 %), Discontinuous Very Low Density Urban Fabric (S.L. : 10 %), Isolated Structures, Industrial commercial public and private units, Port areas, Airports, Mineral extraction and dump sites, Construction sites, Sports and leisure facilities
Other open areas	Land without current use, Open spaces with little or no vegetations (beaches, dunes, bare rocks, glaciers)

Table S1: Classification of the 27 initial land cover types from the Urban Atlas database into 10 new land cover types. The two land cover types written in bold correspond to the two types of habitat considered for the habitat connectivity analyzes.

*Corresponding author: paul.savary@concordia.ca

Code	Initial land cover types	New land cover types
111	Continuous urban fabric	Artificial
112	Discontinuous urban fabric	Artificial
121	Industrial or commercial units and public facilities	Artificial
122	Road and rail networks and associated land	Road
123	Port areas	Artificial
124	Airports	Artificial
131	Mineral extraction sites	Artificial
132	Dump sites	Artificial
133	Construction sites	Artificial
141	Green urban areas	Artificial
142	Sport and leisure facilities	Artificial
211	Non-irrigated arable land	Agricultural
212	Permanently irrigated land	Agricultural
213	Rice fields	Agricultural
221	Vineyards	Agricultural
222	Fruit trees and berry plantations	Agricultural
223	Olive groves	Agricultural
231	Pastures, meadows and other permanent grasslands under agricultural use	Grassland
241	Annual crops associated with permanent crops	Agricultural
242	Complex cultivation patterns	Agricultural
243	Land principally occupied by agriculture, with significant areas of natural vegetation	Semi-natural
244	Agro-forestry areas	Semi-natural
311	Broad-leaved forest	Forest
312	Coniferous forest	Forest
313	Mixed forest	Forest
321	Natural grasslands	Grassland
322	Moors and heathland	Other open
323	Sclerophyllous vegetation	Forest
324	Transitional woodland-shrub	Forest
331	Beaches, dunes, sands	Other open
332	Bare rocks	Other open
333	Sparsely vegetated areas	Other open
334	Burnt areas	Other open
335	Glaciers and perpetual snow	Other open
411	Inland marshes	Wetland
412	Peat bogs	Wetland
421	Coastal salt marshes	Wetland
422	Salines	Wetland
423	Intertidal flats	Wetland
511	Water courses	Water
512	Water bodies	Water
521	Coastal lagoons	Water
522	Estuaries	Water
523	Sea and ocean	Water

Table S2: Classification of the 44 initial land cover types from the Corine Land Cover database into 10 land cover types.

City	Forest	UGS	Forest.Inter.	UGS.Inter.	EC ratio
Scenario 1					
Plovdiv	7.34	8.74	10.86	10.13	5.59
Konya	4.59	5.73	6.47	6.73	5.02
Sevilla	2.62	4.91	4.54	5.11	2.55
Madrid	4.82	4.6	6.29	5.56	1.54
Logrono	5.53	6.61	7.17	6.82	1.53
Alessandria	5.99	6.41	6.83	6.68	1.42
Scenario 2					
Plovdiv	4.11	2.73	7.47	5.59	5.59
Konya	4.63	2.28	4.62	4.01	5.02
Sevilla	2.15	2.29	2.81	2.71	2.55
Madrid	4.08	2.13	5.22	3.24	1.54
Logrono	3.71	4.27	5.37	4.35	1.53
Alessandria	4.64	3.24	4.28	3.48	1.42
Scenario 3					
Plovdiv	2.08	1.81	3.88	2.49	5.59
Konya	2.86	1.61	3.35	2.29	5.02
Sevilla	1.81	1.61	1.88	1.74	2.55
Madrid	2.55	1.41	3.33	1.5	1.54
Logrono	1.6	1.78	1.62	1.71	1.53
Alessandria	1.79	1.65	2.35	1.53	1.42

Table S3: Mean allelic richness in different habitat types according to the cost scenario in the cities for which the ratio $EC_{UGS.UGS}/EC_{Forest.Forest}$ is above 1. "Forest Interface" corresponds to populations located in the forest patches most connected to UGS according to the $F_{Forest \leftrightarrow UGS}$ metric, whereas "UGS Interface" corresponds to populations located in the UGS patches most connected to forests according to the $F_{UGS \leftrightarrow Forest}$ metric.

2 Supplementary figures

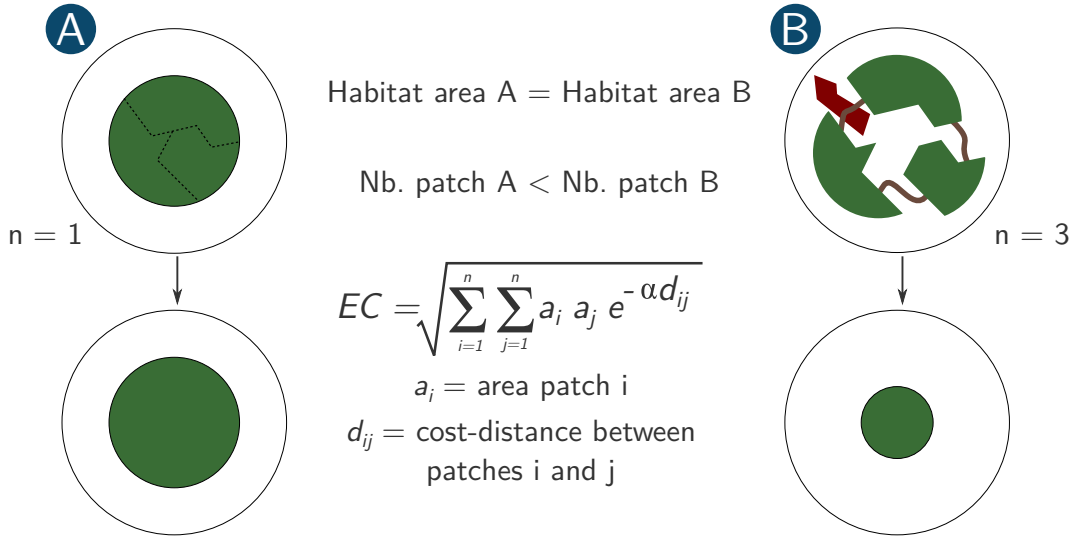

Figure S1: Equivalent Connectivity (EC) formula. This metric is computed from a landscape graph. In both situations A et B, the total amount of habitat is equivalent. However, this habitat area is subdivided into 3 patches in situation B and two patches are separated by a unfavourable land cover type for dispersal movements (red patch). For these reasons, EC is larger in situation A as compared with situation B. In the formula, n is the total number of habitat patches, $e^{-\alpha \times d_{ij}}$ is equivalent to the dispersal probability between patches i and j given the cost-distance d_{ij} between them. α is the parameter determining the shape of the decrease of dispersal probability when d_{ij} increases.

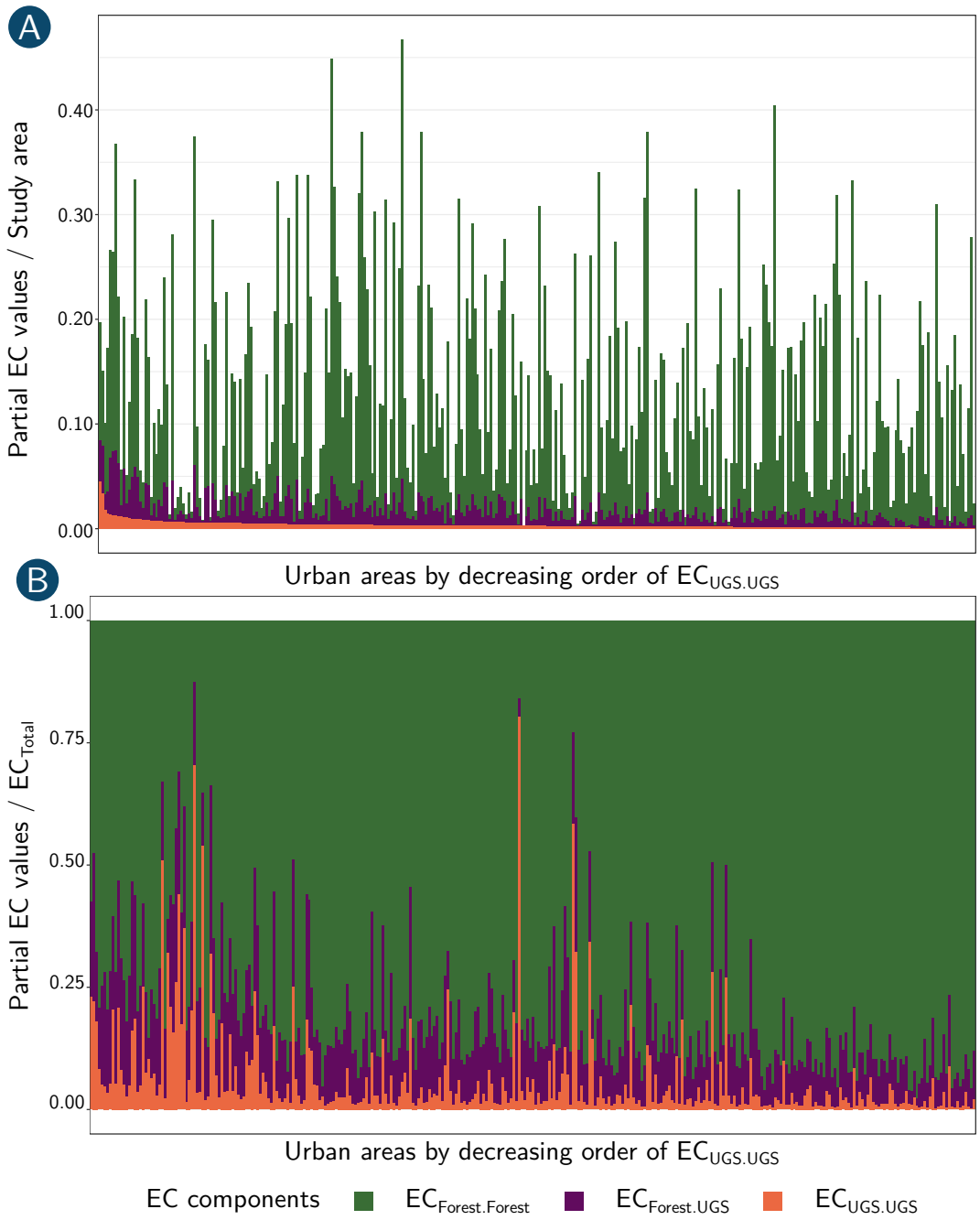

Figure S2: Absolute and relative values of the different EC components computed in the 325 cities according to cost scenario 2. (A) $EC_{Forest.Forest}$ (green), $EC_{Foret.UGS}$ (purple) and $EC_{UGS.UGS}$ (orange) divided by the total study area, for each city. (B) Respective contributions of $EC_{Forest.Forest}$ (green), $EC_{Forest.UGS}$ (purple) and $EC_{UGS.UGS}$ (orange) to the connectivity of the habitat network. The total connectivity value of the network is the sum of the three EC components, which is slightly different from the global EC value because of square root properties.

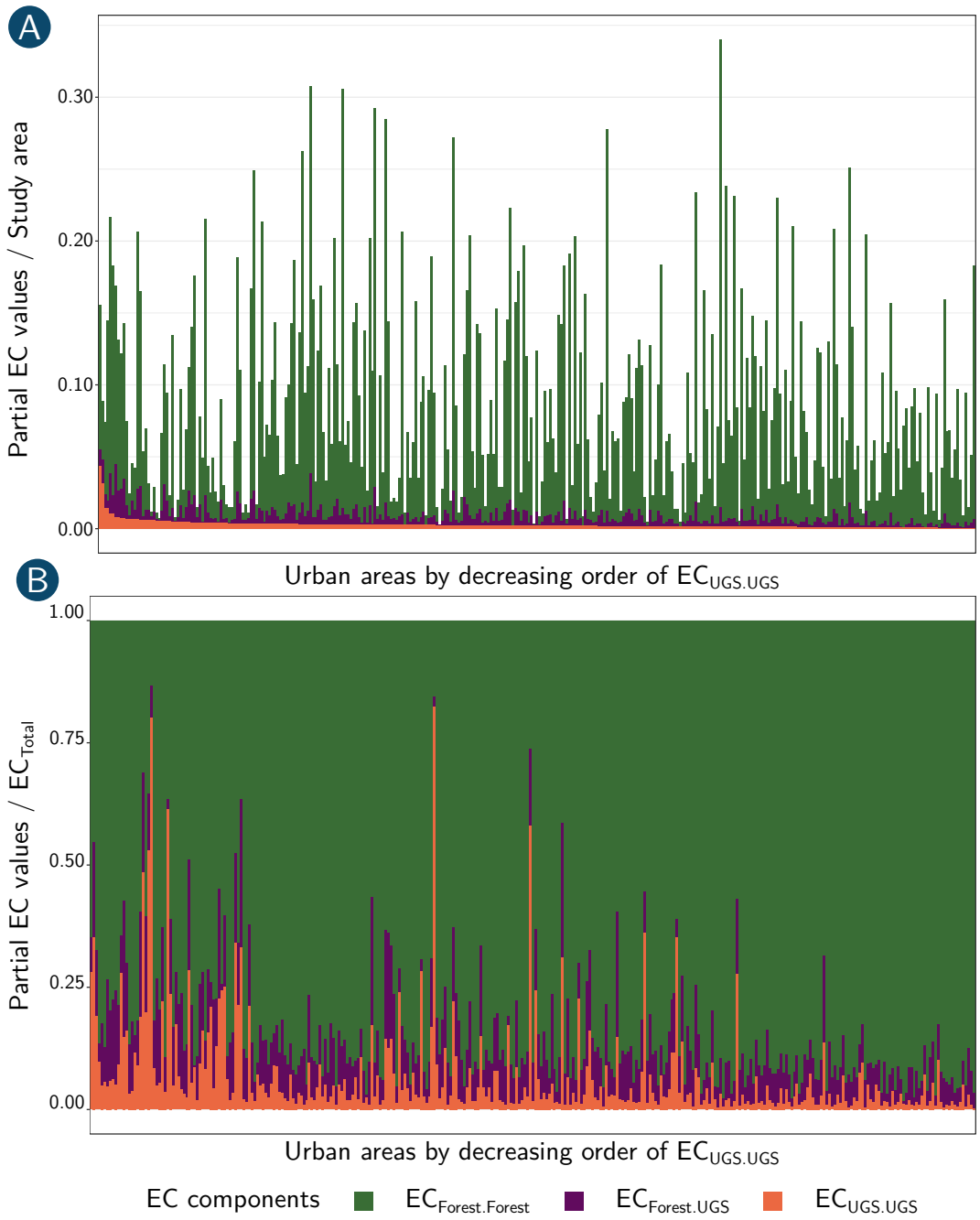

Figure S3: Absolute and relative values of the different EC components computed in the 325 cities according to cost scenario 3. (A) $EC_{Forest.Forest}$ (green), $EC_{Foret.UGS}$ (purple) and $EC_{UGS.UGS}$ (orange) divided by the total study area, for each city. (B) Respective contributions of $EC_{Forest.Forest}$ (green), $EC_{Forest.UGS}$ (purple) and $EC_{UGS.UGS}$ (orange) to the connectivity of the habitat network. The total connectivity value of the network is the sum of the three EC components, which is slightly different from the global EC value because of square root properties.

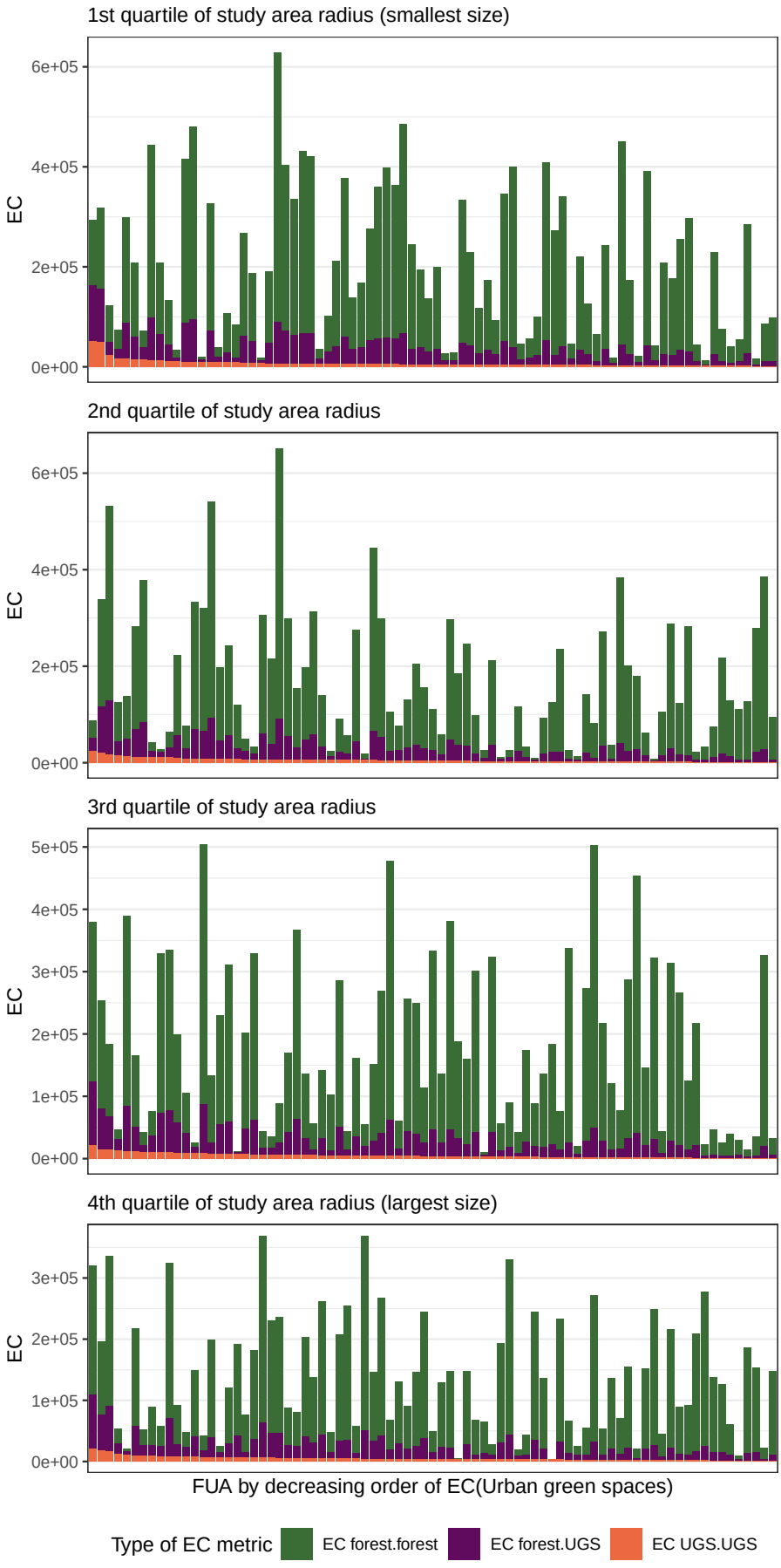

Figure S4: Absolute values of the different EC components computed in the 325 cities according to cost scenario 1. These figures are equivalent to the top panel of Figure 2 in the main document, but plotted by separating cities according to their quartile of radius size (from smallest [top] to largest [bottom]). $EC_{Forest.Forest}$ (green), $EC_{Foret.UGS}$ (purple) and $EC_{UGS.UGS}$ (orange) divided by the total study area, for each urban area. See formula in the main text.

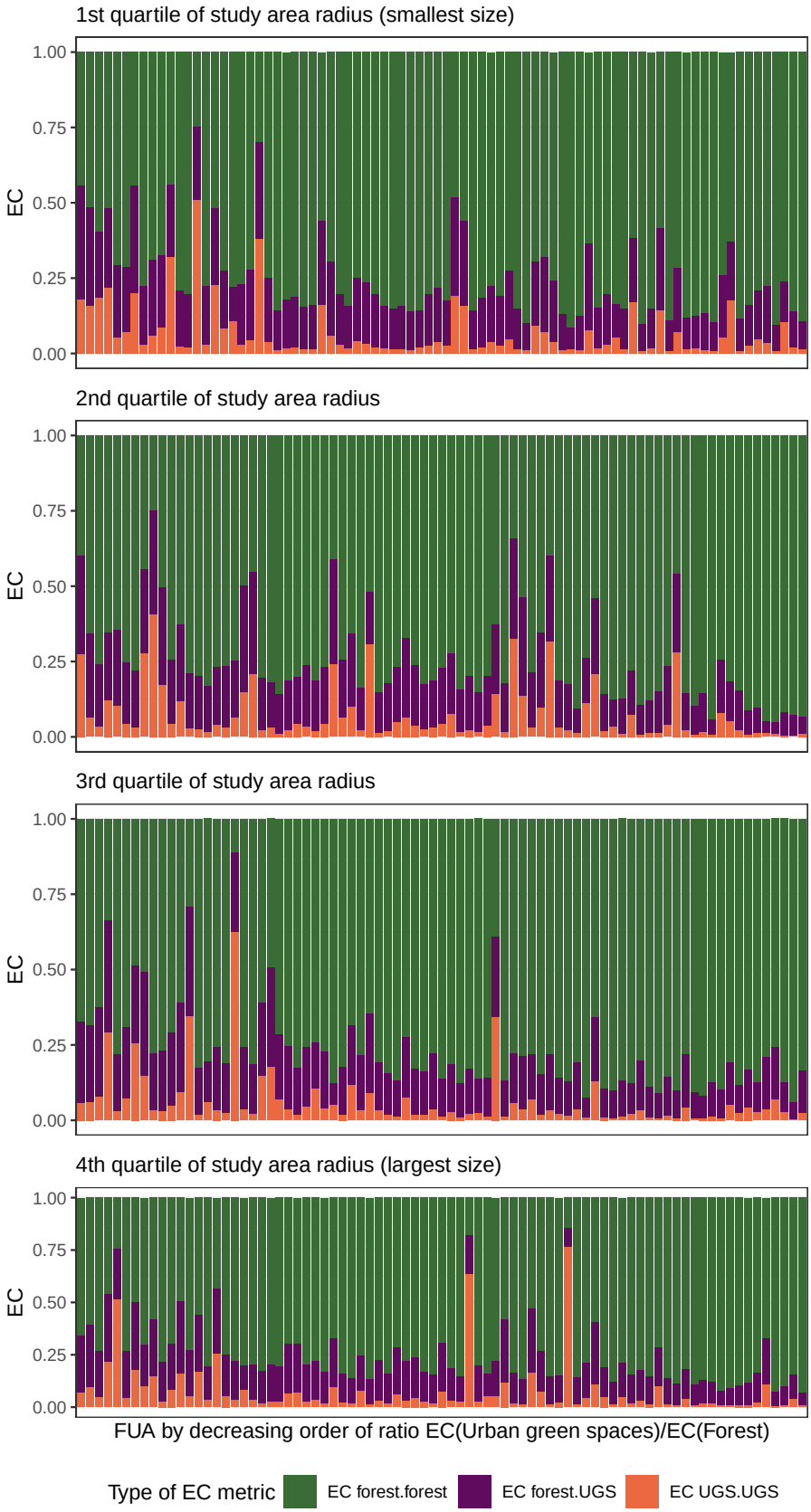

Figure S5: Relative values of the different EC components computed in the 325 cities according to cost scenario 1. These figures are equivalent to the bottom panel of Figure 2 in the main document, but plotted by separating cities according to their quartile of radius size (from smallest [top] to largest [bottom]). Respective contributions of $EC_{Forest.Forest}$ (green), $EC_{Foret.UGS}$ (purple) and $EC_{UGS.UGS}$ (orange) to the connectivity of the habitat network. The total value to which each value is compared is not the global EC value itself but the sum of the three components. Square root properties explain for the difference between these values. See formula in the main text.

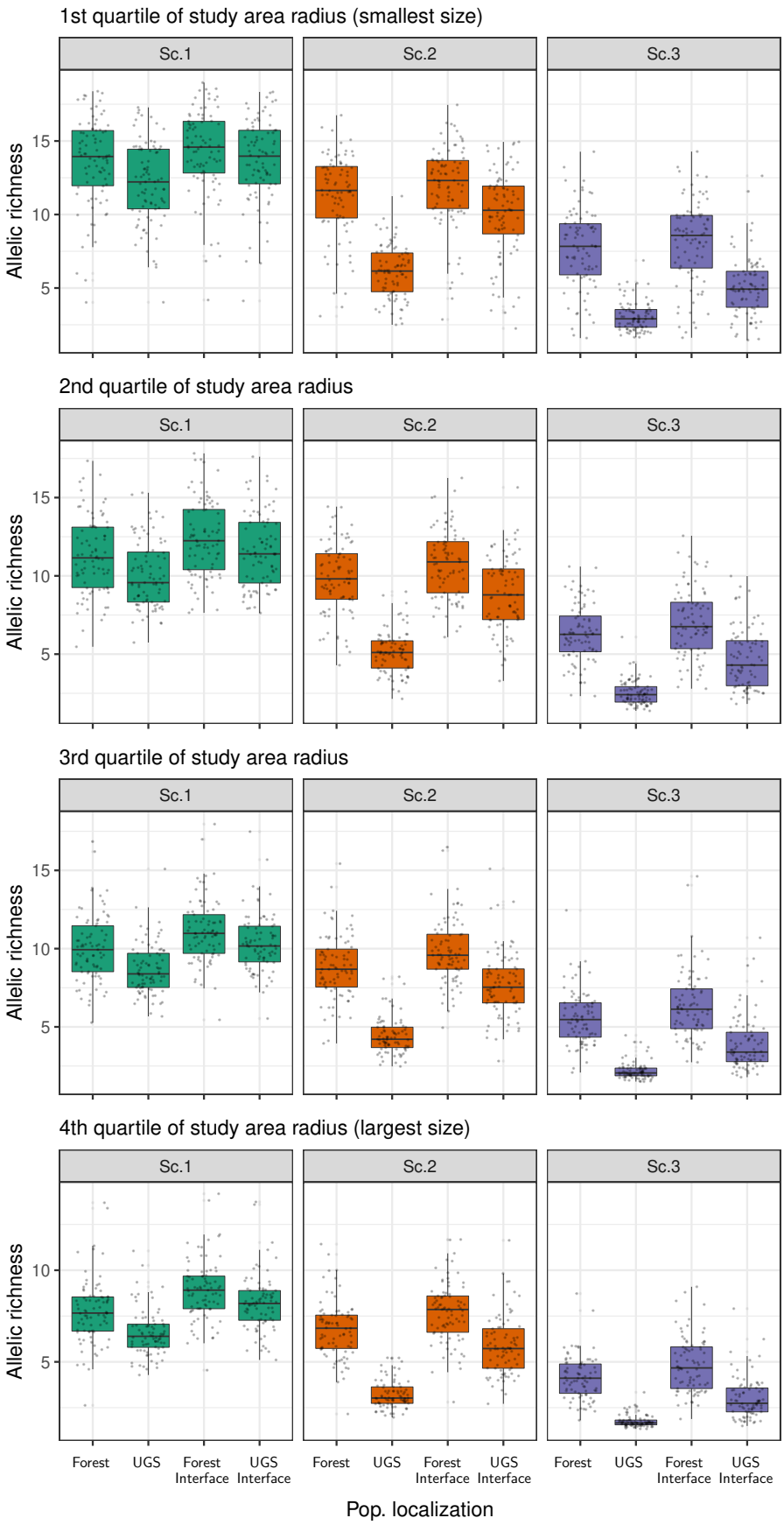

Figure S6: Distribution of the mean allelic richness of "Forest", "Forest Interface", "UGS" and "UGS Interface" populations in the 325 cities across the three dispersal cost scenarios. These figures are equivalent to Figure 3 in the main document, but plotted by separating cities according to their quartile of radius size (from smallest [top] to largest [bottom]). "Forest Interface" corresponds to populations located in the forest patches most connected to UGS according to the $F_{Forest \leftrightarrow UGS}$ metric, whereas "UGS Interface" corresponds to populations located in the UGS patches most connected to forests according to the $F_{UGS \leftrightarrow Forest}$ metric. $n = 325$ values per box. Habitat patch types are mutually exclusive, and each population is represented once per scenario.

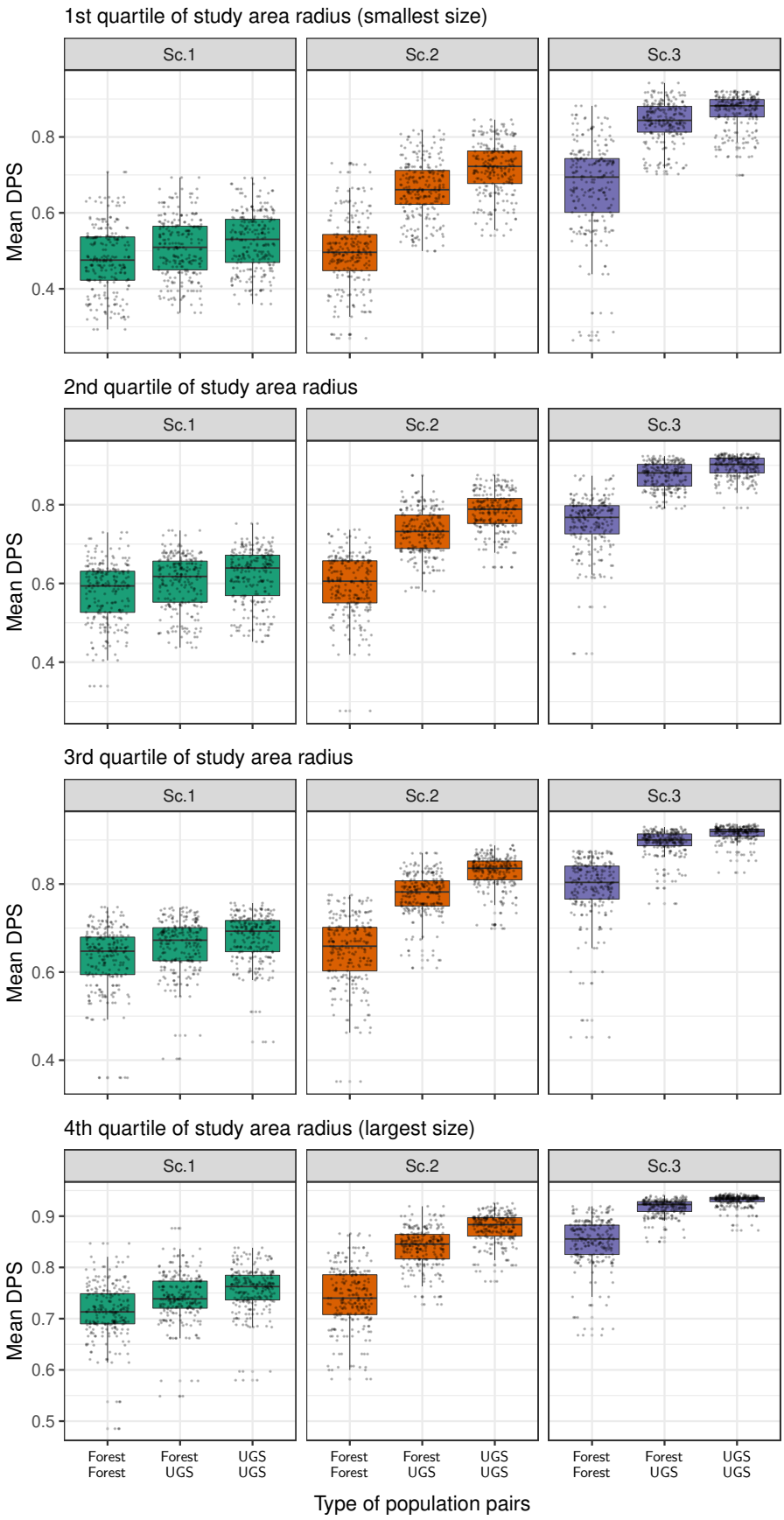

Figure S7: Distribution of the mean genetic differentiation (D_{PS}) computed between forest patches (Forest.Forest), forest and UGS patches (Forest.UGS) or between UGS patches (UGS.UGS) in the 325 cities across the three dispersal cost scenarios. These figures are equivalent to Figure 4 in the main document, but plotted by separating cities according to their quartile of radius size (from smallest [top] to largest [bottom]).

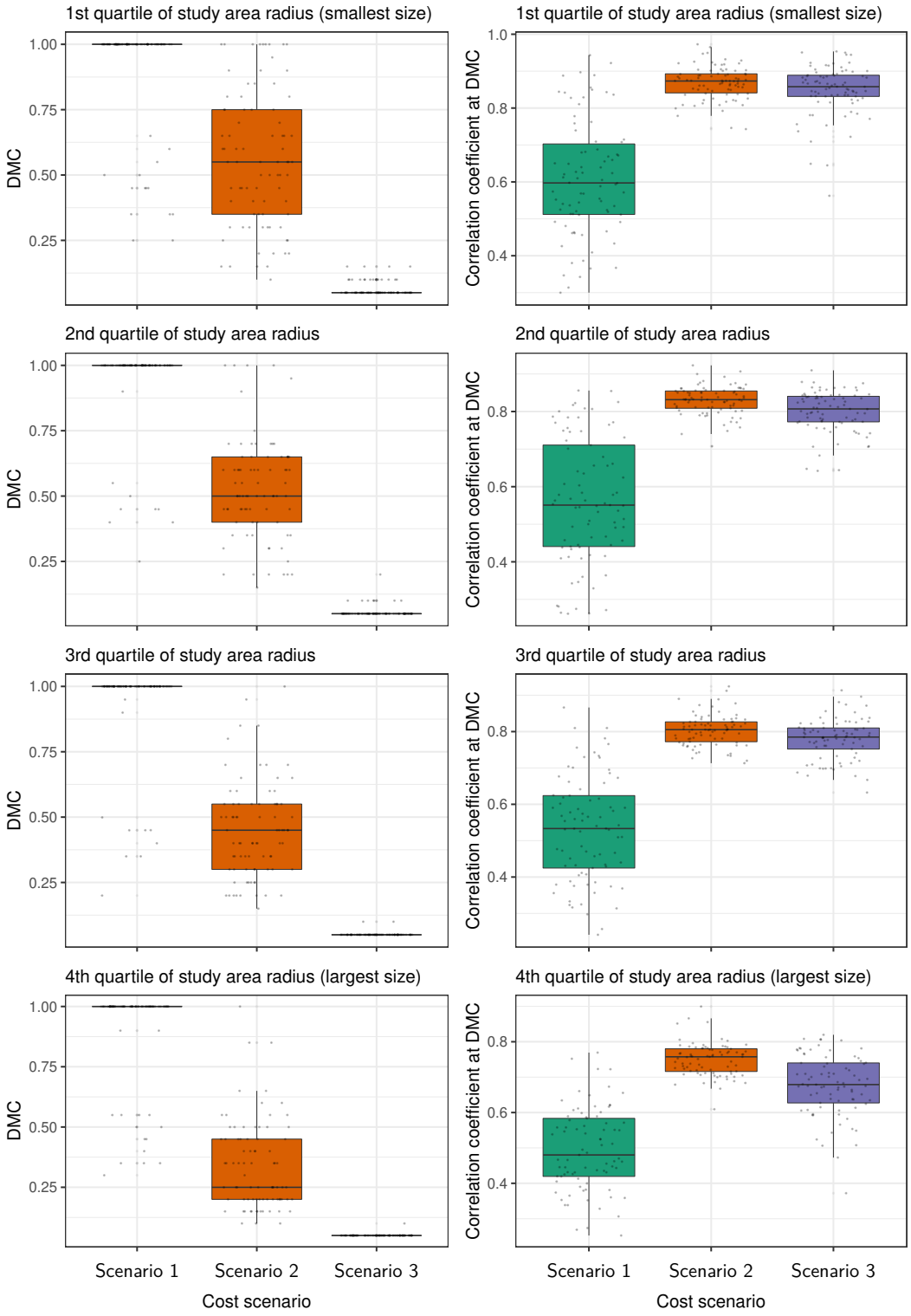

Figure S8: Distribution of the Distance of Maximum Correlation (left panels) and of the Mantel correlation coefficients measured at the DMC (right panels) according to the radius size of the urban study areas. The DMC is computed as the threshold distance used for selecting the subset of population pairs giving the maximum Mantel correlation coefficient between genetic distances (D_{PS}) and cost-distances, in the cities and across the three dispersal cost scenarios. The DMC is divided by the maximum cost-distances between populations in the corresponding urban area and cost scenario and therefore ranges from 0 to 1. These figures are equivalent to Figure 5 in the main document, but plotted by separating cities according to their quartile of radius size (from smallest [top] to largest [bottom]).

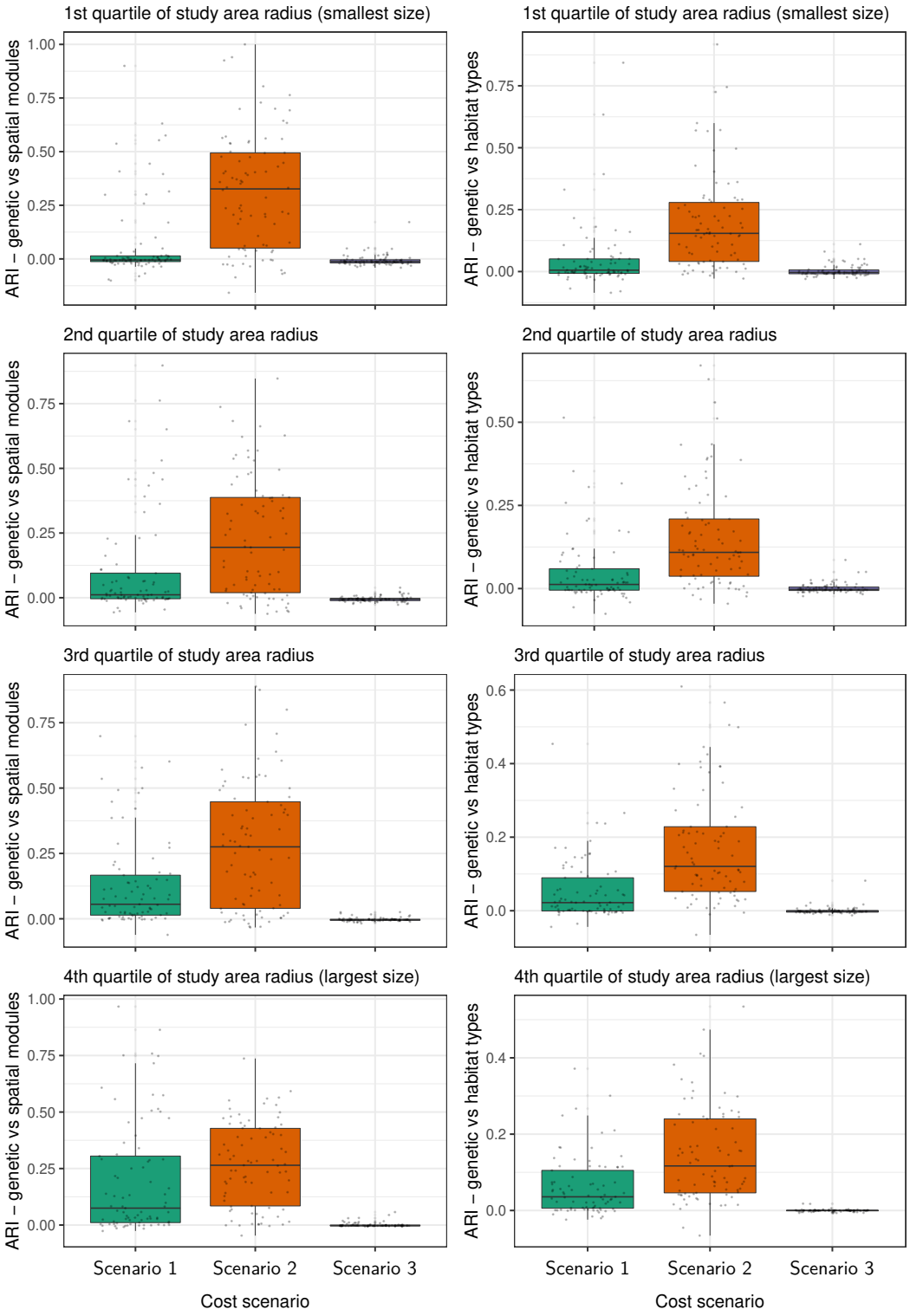

Figure S9: Distribution of the Adjusted Rand Index (ARI) comparing the partitions into modules of the genetic graph with links weighted by genetic distances (D_{PS}) with module partitions obtained from similar graphs with links weighted by cost-distances (left panels) or with the classifications of populations into forest or UGS populations according to the type of patch they occupy (right panels). These figures are equivalent to Figure 6 in the main document, but plotted by separating cities according to their quartile of radius size (from smallest [top] to largest [bottom]). An ARI value is computed in every urban area for each cost scenario.

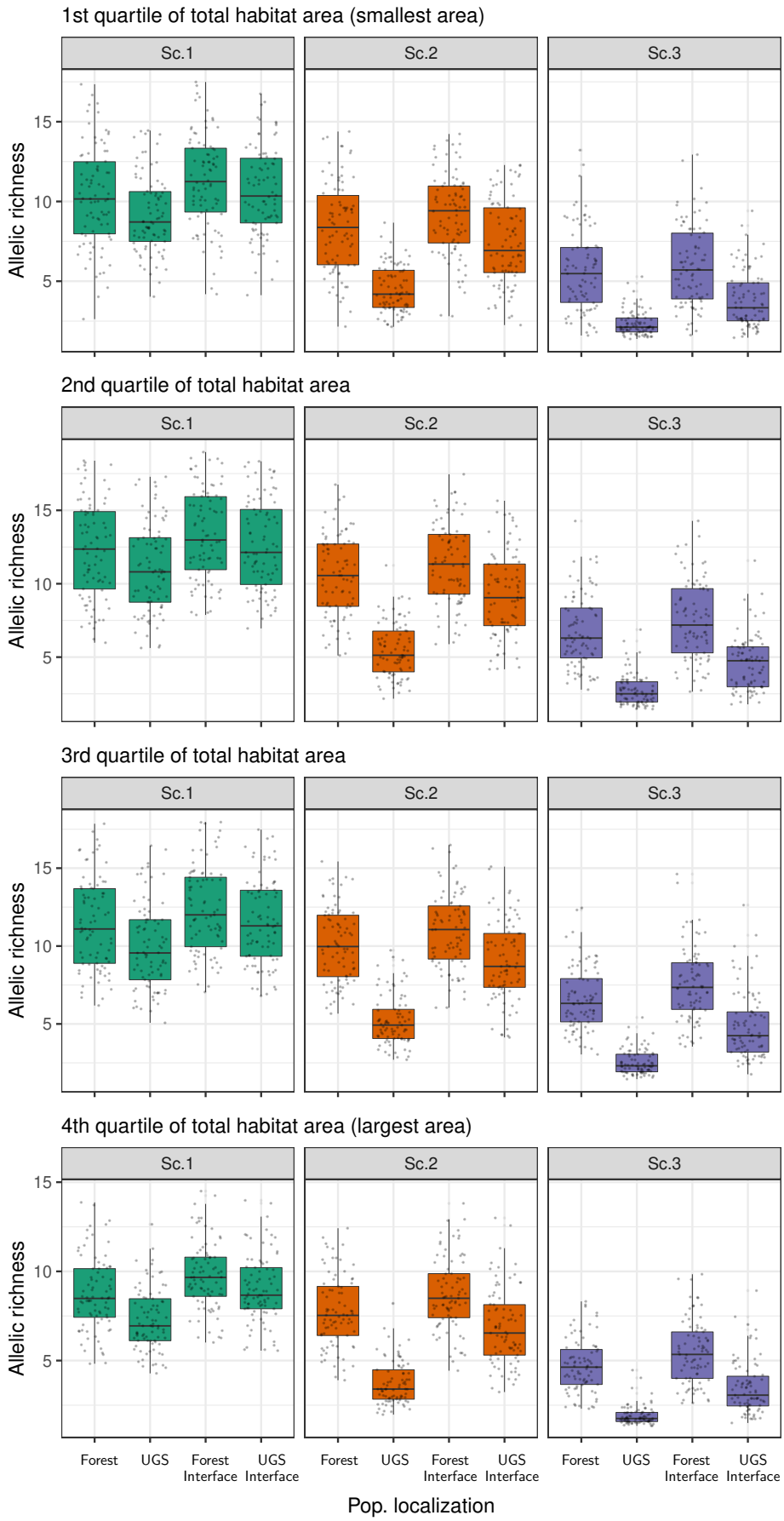

Figure S10: Distribution of the mean allelic richness of "Forest", "Forest Interface", "UGS" and "UGS Interface" populations in the 325 cities across the three dispersal cost scenarios. These figures are equivalent to Figure 3 in the main document, but plotted by separating cities according to their quartile in terms of total area of forest and urban green spaces (from smallest [top] to largest [bottom]). "Forest Interface" corresponds to populations located in the forest patches most connected to UGS according to the $F_{Forest \leftrightarrow UGS}$ metric, whereas "UGS Interface" corresponds to populations located in the UGS patches most connected to forests according to the $F_{UGS \leftrightarrow Forest}$ metric. $n = 325$ values per box. Habitat patch types are mutually exclusive, and each population is represented once per scenario.

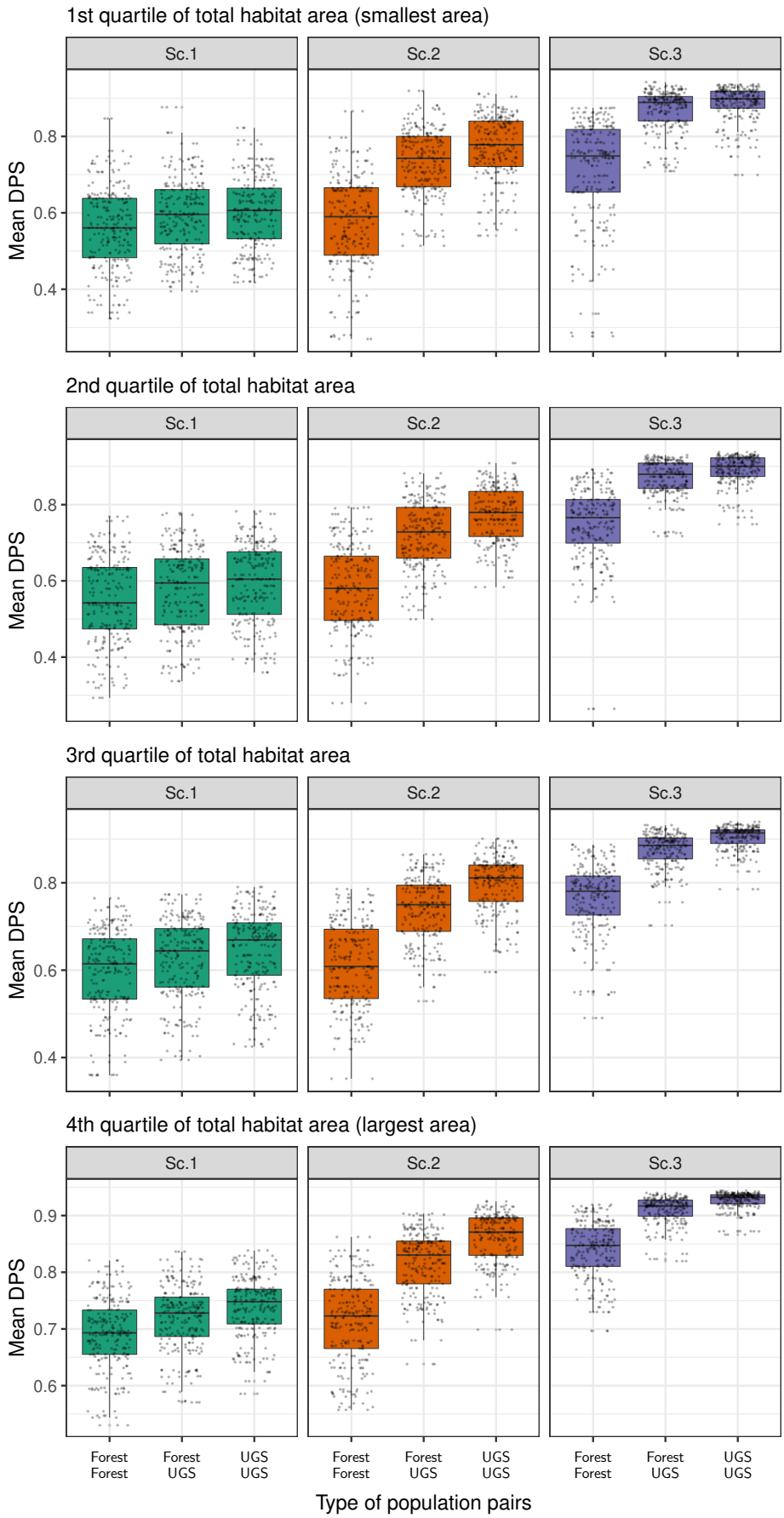

Figure S11: Distribution of the mean genetic differentiation (D_{PS}) computed between forest patches (Forest.Forest), forest and UGS patches (Forest.UGS) or between UGS patches (UGS.UGS) in the 325 cities across the three dispersal cost scenarios. These figures are equivalent to Figure 4 in the main document, but plotted by separating cities according to their quartile in terms of total area of forest and urban green spaces (from smallest [top] to largest [bottom]).

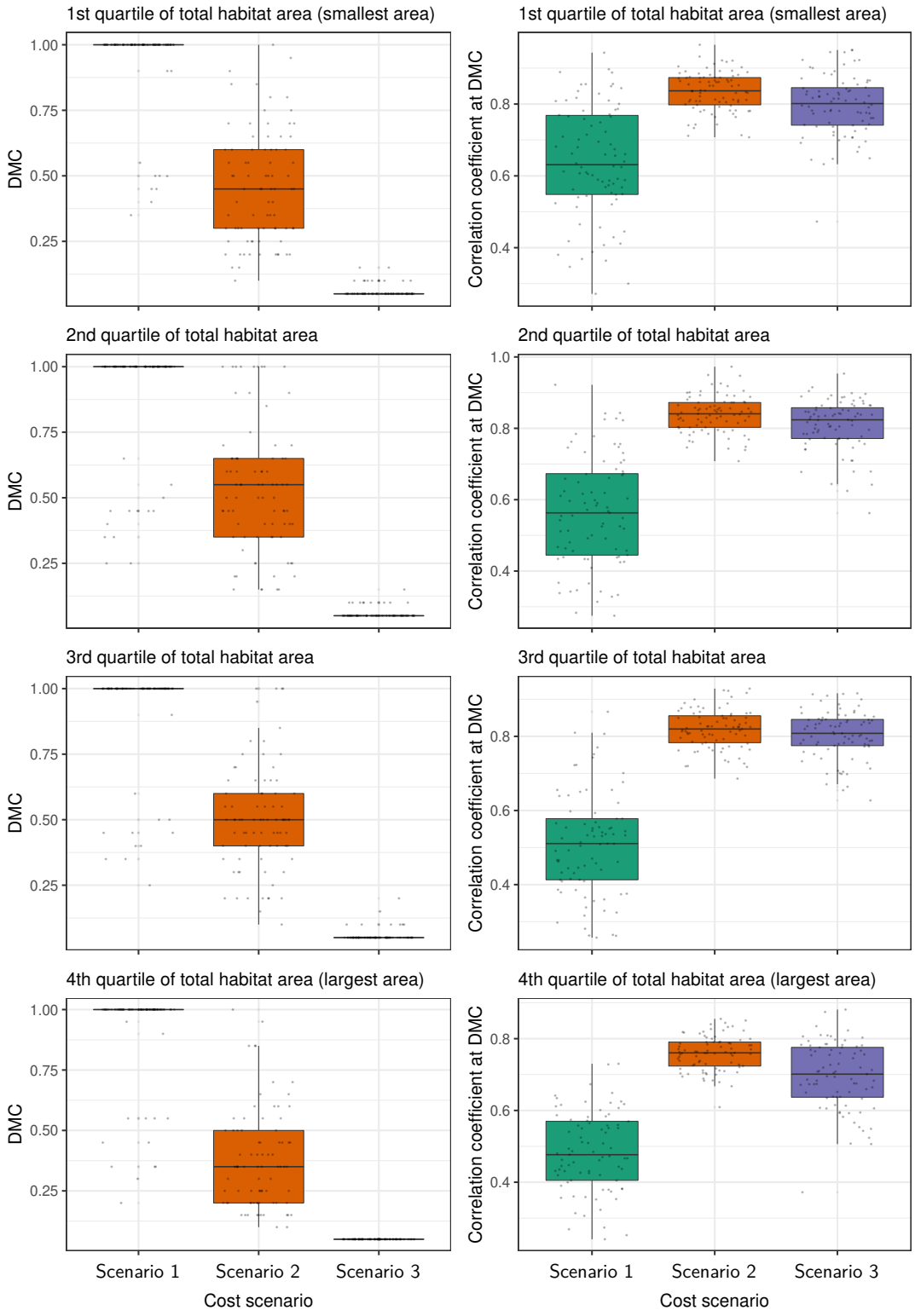

Figure S12: Distribution of the Distance of Maximum Correlation (left panels) and of the Mantel correlation coefficients measured at the DMC (right panels) according to their quartile in terms of total area of forest and urban green spaces. The DMC is computed as the threshold distance used for selecting the subset of population pairs giving the maximum Mantel correlation coefficient between genetic distances (D_{PS}) and cost-distances, in the cities and across the three dispersal cost scenarios. The DMC is divided by the maximum cost-distances between populations in the corresponding urban area and cost scenario and therefore ranges from 0 to 1. These figures are equivalent to Figure 5 in the main document, but plotted by separating cities according to their quartile in terms of total area of forest and urban green spaces (from smallest [top] to largest [bottom]).

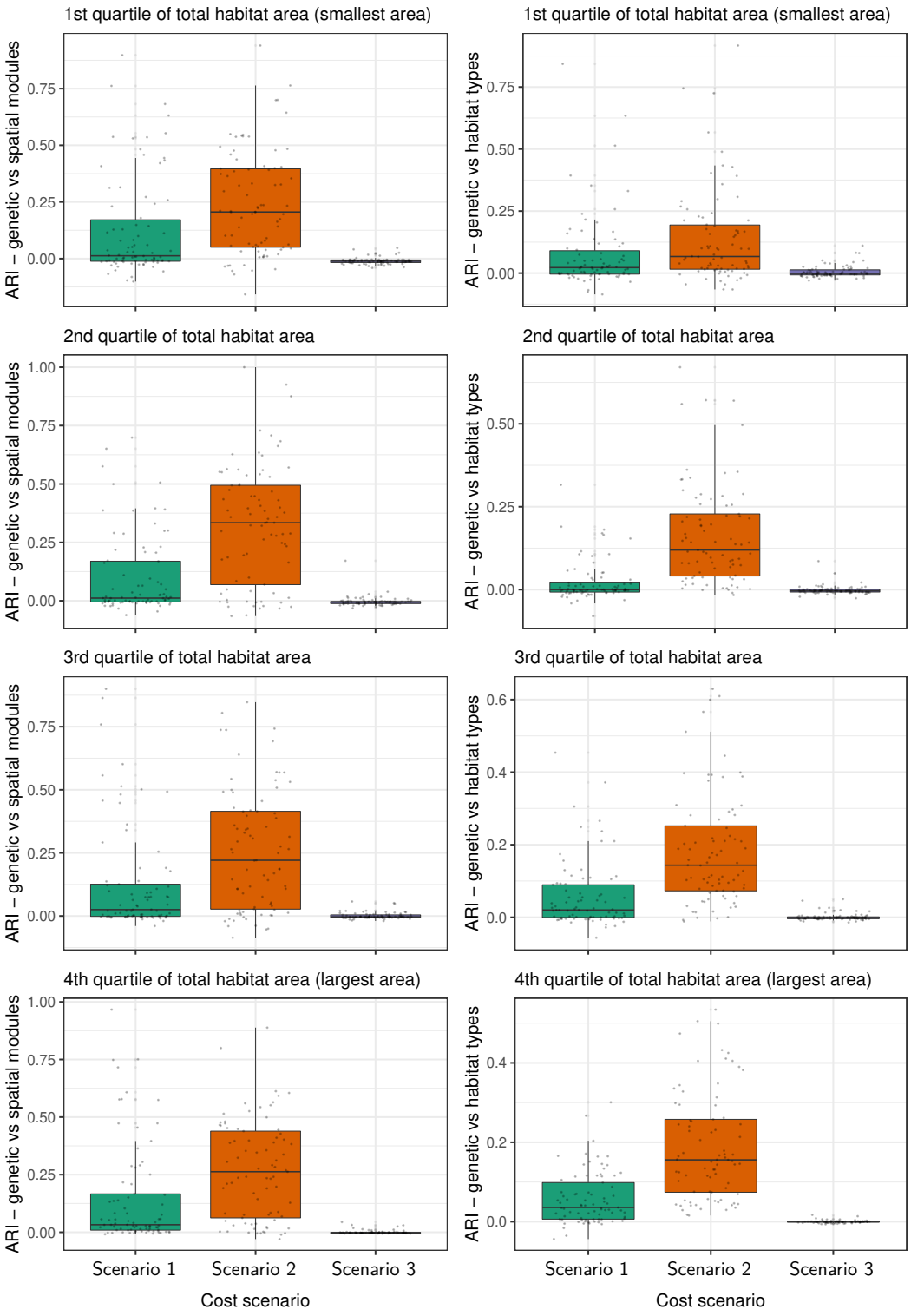

Figure S13: Distribution of the Adjusted Rand Index (ARI) comparing the partitions into modules of the genetic graph with links weighted by genetic distances (D_{PS}) with module partitions obtained from similar graphs with links weighted by cost-distances (left panels) or with the classifications of populations into forest or UGS populations according to the type of patch they occupy (right panels). These figures are equivalent to Figure 6 in the main document, but plotted by separating cities according to their quartile in terms of total area of forest and urban green spaces (from smallest [top] to largest [bottom]). An ARI value is computed in every urban area for each cost scenario.

3 Mixed-effects modeling supplementary results

3.1 Distance of Maximum Correlation and corresponding correlations

The Distance of Maximum Correlation was computed as the ratio of the pruning distance threshold to the maximum distance between two populations for which we obtain the maximum correlation between genetic distances (DPS) and cost-distances. We modelled it as a function of the dispersal cost scenario only, as we had one value per city and scenario in that case. We included a city level random intercept. The DMC ranges from 0 to 1, and includes both 0 and 1 values, which precludes us from using a GLMM based on a beta distribution. This response was highly heteroscedastic and we did not find an ideal model fitting its distribution adequately. However, the differences observed on Figure 5 are sufficiently neat for a conservative interpretation of the results (see main document).

We modelled the correlation coefficient between genetic distances (DPS) and cost-distances with a GLMM assuming a beta distribution of the response and a logit link function, and with a LMM. In both cases, the mixed-effects model had the following structure:

$$Coef.corr.DMC \sim Cost.scenario + (1|City)$$

Both models had reasonable fit and matched residual assumptions. In both cases, the distributions assumed by the models are however suboptimal given that a correlation coefficient varies between -1 and 1. The beta distribution is strictly positive and excludes 0 and 1, whereas the LMM does not consider this limited range. We comment the results obtained from both models.

In the LMM, the random effects explained about a quarter of the variance (ICC: 0.277) and its overall fit was good (conditional $R^2 = 0.66$, marginal $R^2 = 0.53$). According to both models, the maximum correlation coefficient between genetic distances and cost-distances (i.e., when pruning distances matrices at a distance equal to the DMC) was lowest in scenario 1 (LMM: 0.55, 95% CI [0.54, 0.56], GLMM: 0.55, 95% CI [0.54, 0.57]), intermediate in scenario 3 (LMM: 0.78, 95% CI [0.76, 0.79], GLMM: 0.78, 95% CI [0.77, 0.79]) and highest in scenario 2 (LMM: 0.81, 95% CI [0.80, 0.83], GLMM: 0.81, 95% CI [0.80, 0.82]).

3.2 Adjusted Rand Index

No model had a satisfactory residual distribution, and all the models we intended to fit revealed that the variations observed on the Figure 6 are similar after accounting for city level variations among ARI values (see codes).