Supplementary material

Contents

[Supplement S1. Derivation of Eq. (3) from Egs. (1) and (2) 2](#_Toc128323509)

[Supplement S2: Derivation of the formula that offspring point ends up between parental points in a system with pure PVDI 3](#_Toc128323510)

[Supplement S3. HDBSCAN algorithm description 4](#_Toc128323511)

[Supplement S4. Expected measurement values across the parameter space 6](#_Toc128323512)

[Supplement S5. Example simulation runs 9](#_Toc128323513)

[Supplement S6. Results of a model where agent distance is normalised after each step 18](#_Toc128323514)

[Supplement S7. Results with a random variation along all dimensions of the trait space 24](#_Toc128323515)

All Supplementary animations are deposited in a separate folder at

<https://doi.org/10.17605/osf.io/pvyhe>

# Supplement S1. Derivation of Eq. (3) from Egs. (1) and (2)

We have introduced the Galton-Pearson inheritance model in Eq. (1) and the PVDI model in Eq. (2) in the main article. We can easily combine the two by considering both mutation terms from (1) and (2) together

|  |  |  |
| --- | --- | --- |
|  |  |  |

due to the additivity of variance

|  |  |  |
| --- | --- | --- |
|  |  |  |

To generalise the model to a -dimensional trait space, we consider agents represented by vectors and replace the modulus in the definition of the standard deviation by a -dimensional norm

|  |  |  |
| --- | --- | --- |
|  |  |  |

where are parental agents.

The magnitude of the phenotypic mutation is therefore given by

|  |  |  |
| --- | --- | --- |
|  |  | (S1) |

Because we assume the offspring to be on the line connecting its parent agents and , the magnitude in Eq. (S1) needs to be applied to a unit vector parallel to the connection line. Since normal distribution is symmetric, we may use for instance the unit vector

|  |  |  |
| --- | --- | --- |
|  |  |  |

To obtain Eq. (3), all that remains is to add the resulting vector

|  |  |  |
| --- | --- | --- |
|  |  |  |

to the -dimensional mean . We obtain

|  |  |  |
| --- | --- | --- |
|  |  |  |

Which is Eq. (3).

# Supplement S2: Derivation of the formula that offspring point ends up between parental points in a system with pure PVDI

Recall that the cumulative distribution function of a random variable with normal distribution with mean and standard deviation is

where is the Gauss error function defined as

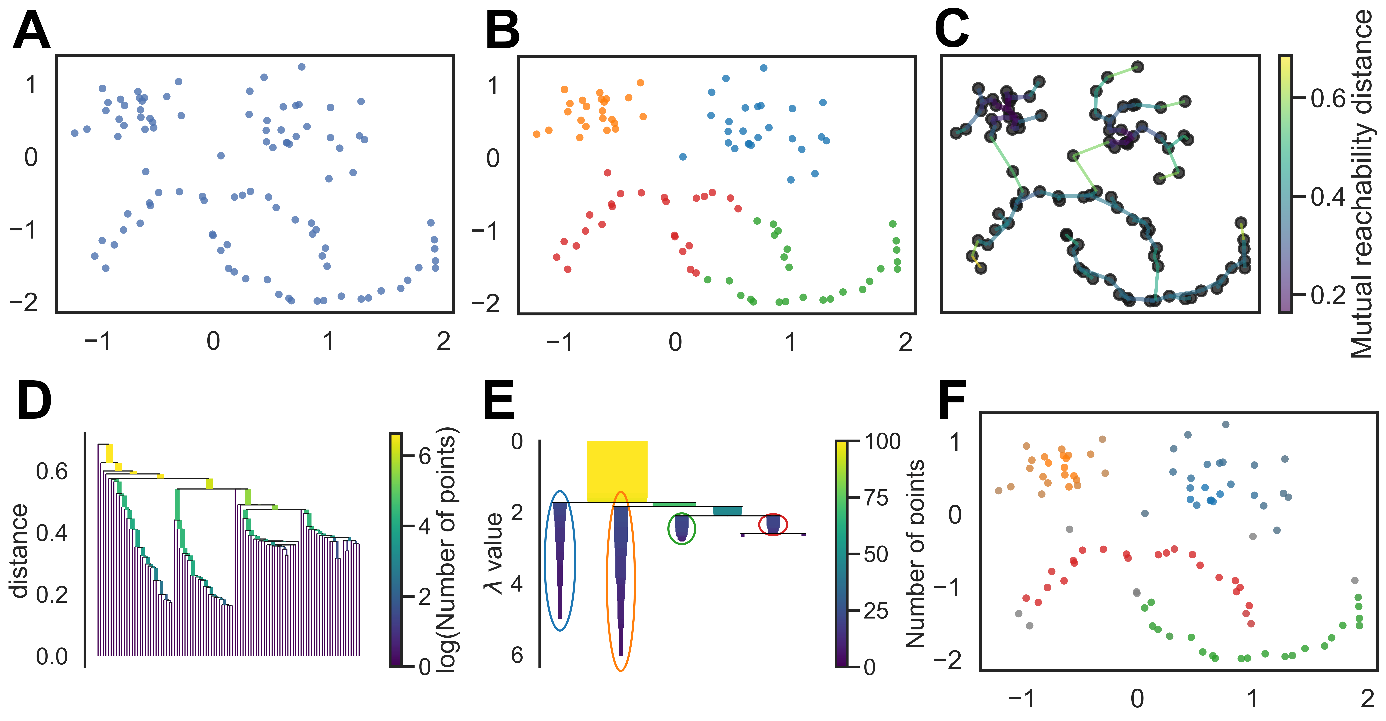
It is well-known that is an odd function, i.e. .  
  
We are interested in the probability that offspring position ends up between parental positions. Without loss of generality, we can translate and rotate the vector connecting the parental positions so that we may consider only one-dimensional culture space and assume that the mean of the parental positions is at . Let us denote the parental distance by , i.e. the parents are located at , respectively. As described in the main text, the position of the offspring, denoted here by a random variable , is normally distributed with standard deviation . The probability that the position of the offspring remains between and can be computed by

In particular, we observe that the probability that the offspring remains in the line segment limited by the positions of the parents depends only on the parameter . Considered as a function of , the probability is a decreasing function.

# Supplement S3. HDBSCAN algorithm description

We will demonstrate the HDBSCAN clustering algorithm on the data in Fig S1A. The data were chosen similarly as in How HDBSCAN works (<https://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html>) from the documentation of the implementation of HDBSCAN in sklearn (Pedregosa et al., 2011). For comparison, we also classified the data using the well-known K-means clusterer with the correct number of clusters. The result in Fig. S1B shows that the K-means algorithm does not perform perfectly on clusters with non-circle geometry.

For two data points and , we denote the distance between and by .   
  
**1. (estimate the density)** Given a positive integer and a point , define the core distance by the distance from to its -th closest neighbour. The core distance corresponds to the density of data around a point by . In the HDBSCAN implementation in sklearn, the parameter is known as .  
  
**2. (construct distance reflecting both density and closeness)** Given as above and two points and , define the mutual reachability distance . The mutual reachability distance measures if two points are close to each other or are in a sufficiently dense region. In effect, points in sparse regions are more distant under the mutual reachability distance than under the original distance while the distance between points in dense regions is unaffected.  
  
**3. (construct minimum spanning tree)** The minimum spanning tree is a tree connecting all the data points with the minimum possible total edge weight, where the weight of the edge is given by the distance of the two points under the mutual reachability distance.  (Fig S1C) The minimum spanning tree can be constructed by Jarník-Prim's algorithm (Rosen, 2011).  
  
**4. (construct cluster hierarchy)** First, convert the minimum spanning tree into a single linkage tree. (Fig S1D) The single linkage tree can be viewed as a hierarchical version of the minimum spanning tree with the hierarchy determined by the weights of the edges in the minimum spanning tree with largest weights being on the top of the tree. Then, condense the single linkage tree into a cluster hierarchy: starting at the top of the tree, remove edges and keep only clusters that are larger than , a given positive integer. (Fig S1E)  
  
**5. (choosing the clusters)** It remains to choose the clusters from the cluster hierarchy. Intuitively, when deciding if a base cluster or its two descendants should be selected, one compares the volume of the clusters in the cluster hierarchy determined by its persistence. If the volume of the base cluster in the cluster hierarchy is larger than the sum of the volumes of the two descendants, choose the base cluster, otherwise choose the two descendants (ellipses in Fig S1E, Fig S1F, the hue of the point corresponds to the likelihood of being a part of a cluster).



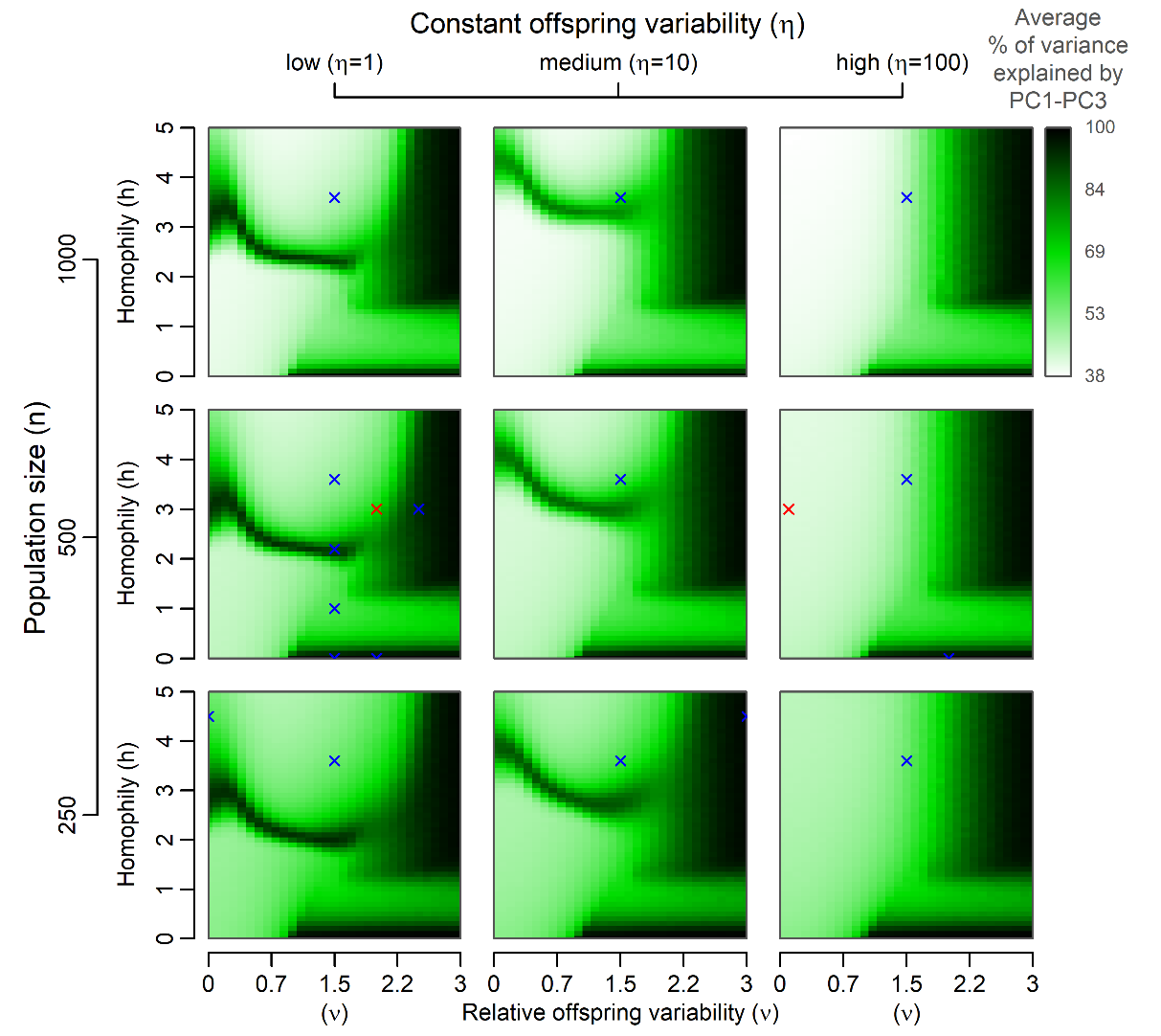
**Figure S1. Visulaization of the HDBSCAN algorithm,** A: exemplar data, B: result of K-means clustering, C: minimum spanning tree, D: cluster hierarchy, E: Selected stable clusters, F: the resulting point affiliation,see the description above for details.

**References:**

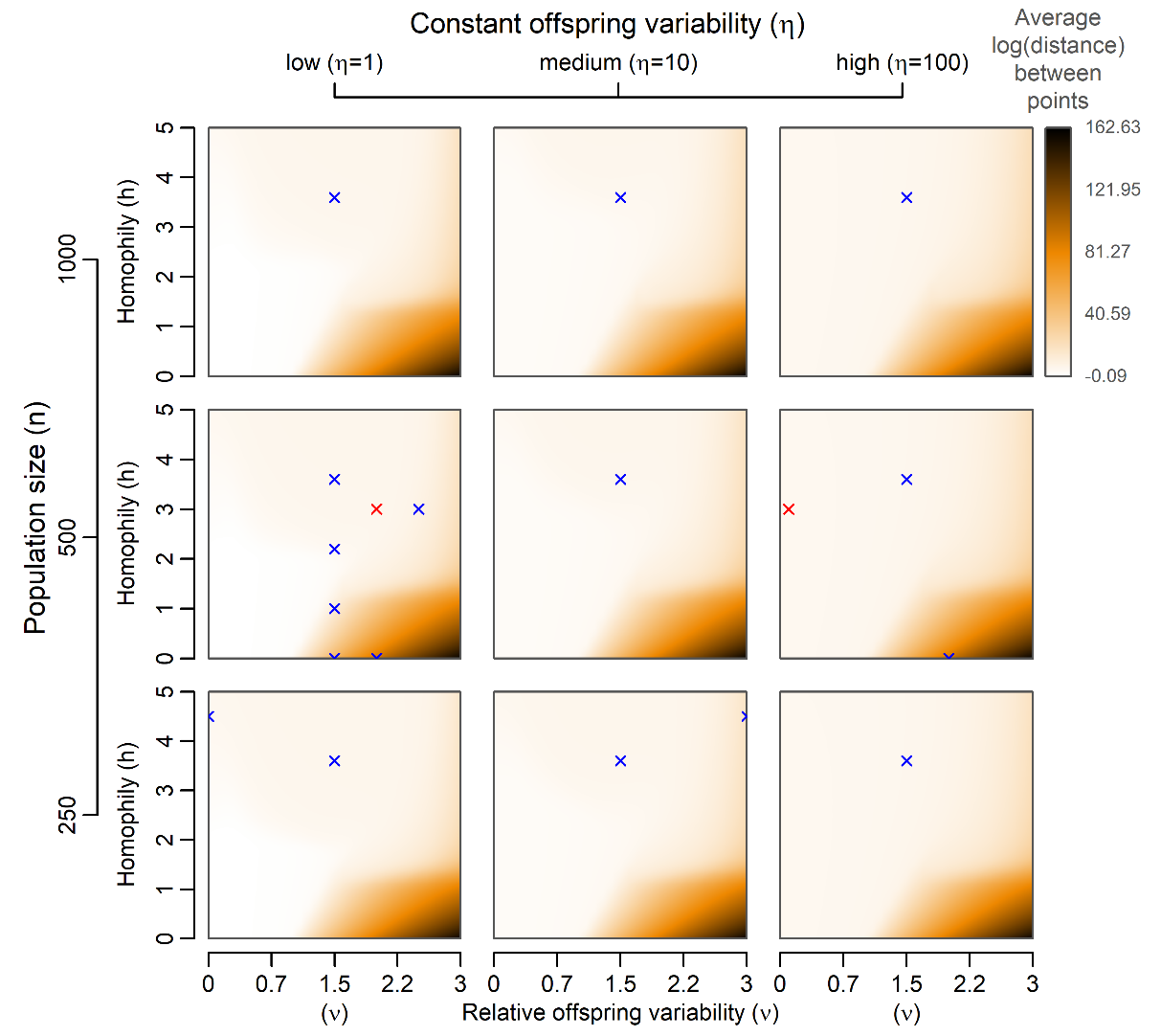
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rosen, K. (2011). *Discrete Mathematics and Its Applications* (7th ed.). New York: McGraw Hill.

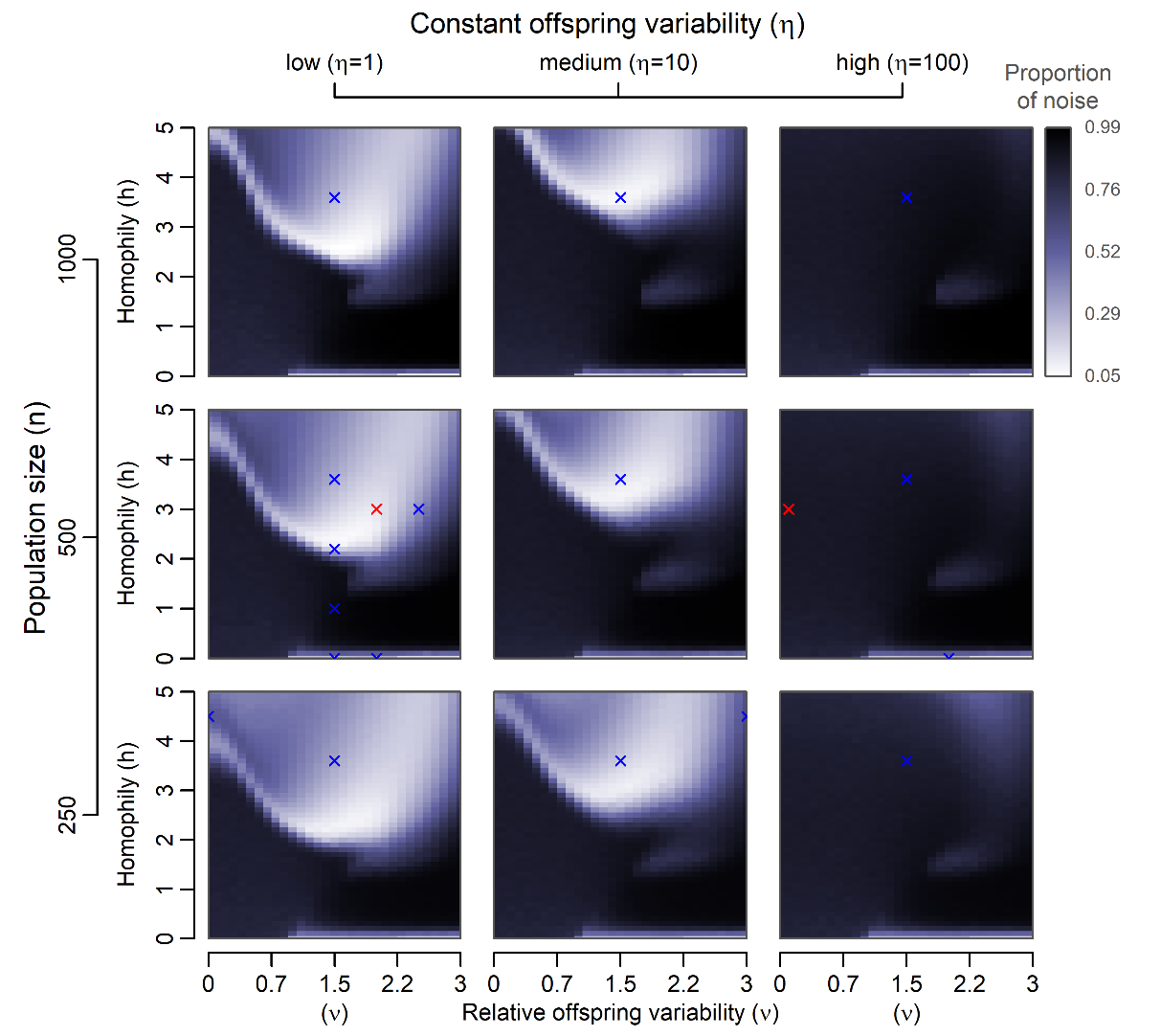
# Supplement S4. Expected measurement values across the parameter space



**Figure S2. A graphical summary of expected reduction in the number of effective dimensions** after 200 model generations. The same set of simulation runs as in Figure 2 was used to generate the image. PC1–PC3 stand for the first three principal components. Red crosses indicate the values of parameters used for single-run examples included in the main article. Blue crosses indicate examples available in the Supplementary material. The extreme values of explained variance (close to 100%) can stem from two distinct, yet not mutually exclusive, reasons. The “black bridge” on the left side of each panel corresponds to scenarios, where exactly two clusters are likely to emerge and exist for a substantial amount of the simulation run (see Figure S13). The dimension that separates the two clusters the best (PC1 after the rotation) obviously monopolizes almost all variance. The black region on the right of each panel is caused by an unconstrained variance growth ( so two points are expected to get further apart if they are mutually selected as role models). Under such circumstances, variance along all effective dimensions of the point configuration is expected to grow exponentially, so the dimensions that are more prominent at the beginning by chance grow disproportionally faster in absolute terms and soon contain most of the extreme variance that emerges (Figure S7). Negative or low homophily can lead to a similar scenario even at lower values of , because the variance is expected to grow for the population as a whole even if every two interacting points are expected to get closer to each other (Figure S9).

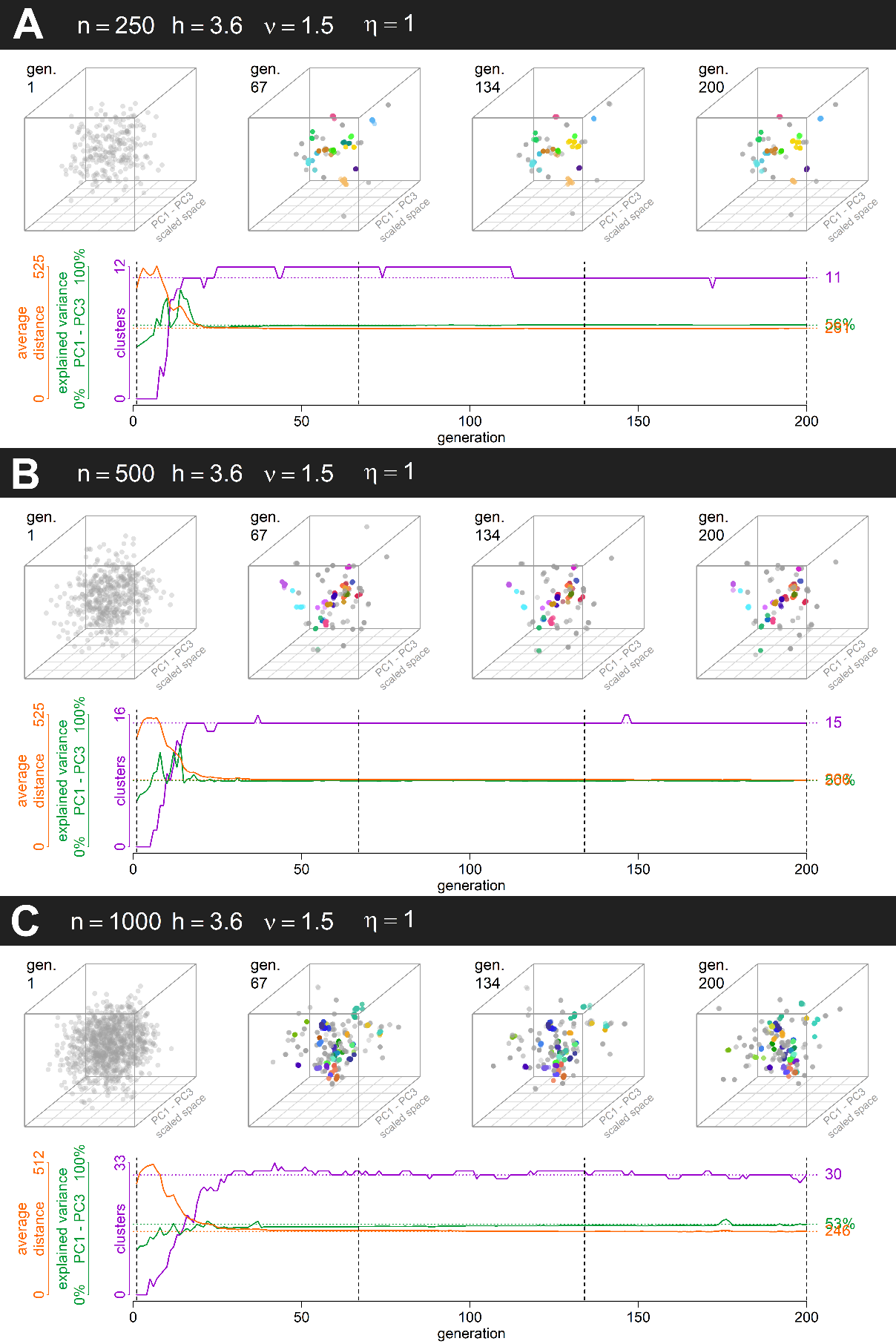


**Figure S3. A graphical summary of the expected average distance between agent positions in a culture-space** after 200 model generations. The same set of simulation runs as in Figure 2 was used to generate the image. Red crosses indicate the values of parameters used for single-run examples included in the main article. Blue crosses indicate examples available in the Supplementary materials. The extreme distance between points at ends of simulations in the bottom right corner of each panel of parameter space is due to the combination of three things: 1. There is no selection in the model that could trim extreme values. 2. High values of cause points to move further apart. 3. Low values of cause formation of more distance pairs of agents and, in turn, an increased speed of the variance growth.

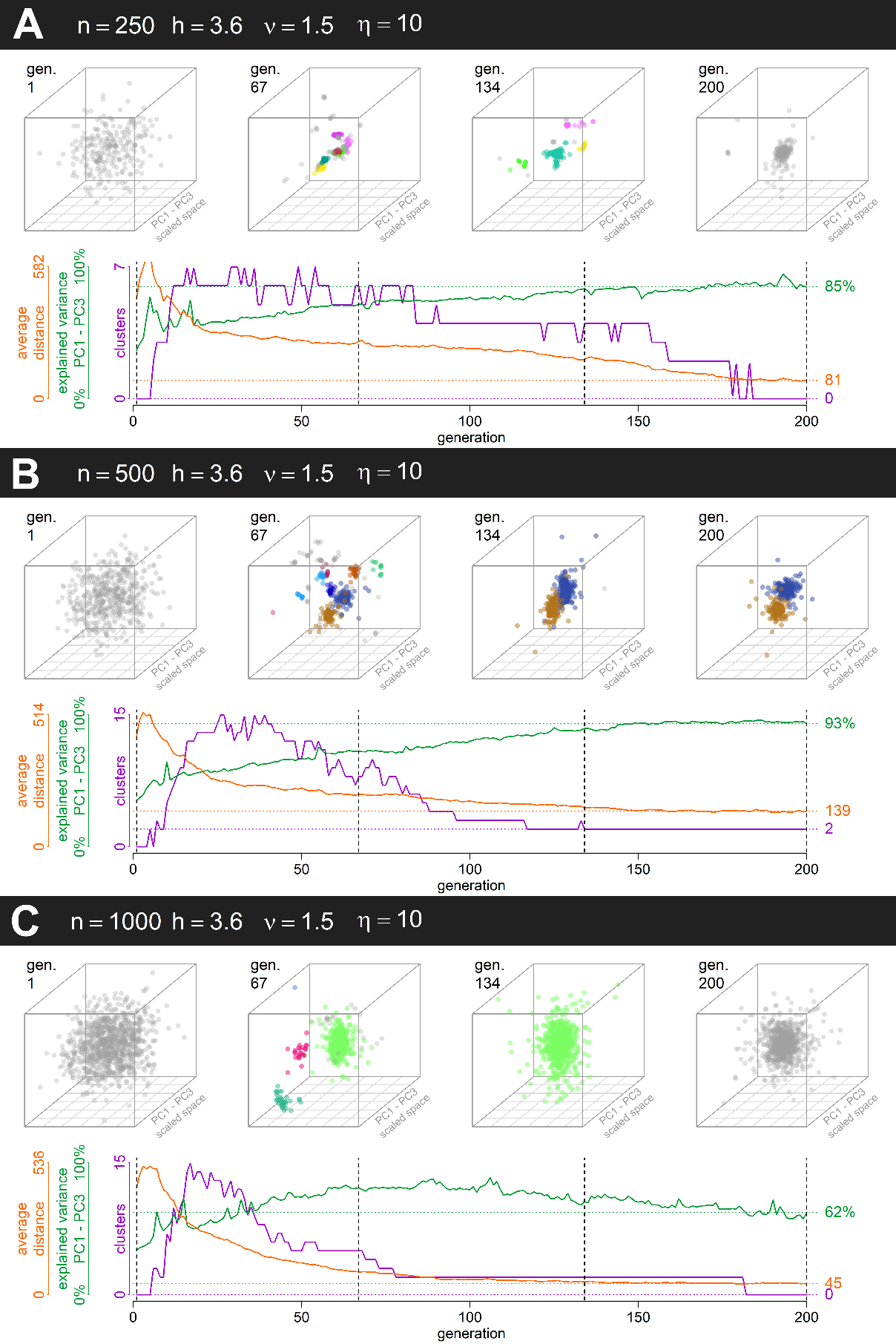


**Figure S4. A graphical summary of the proportion of noise (points that are not a part of any cluster)** after 200 model generations. The same set of simulation runs as in Figure 2 was used to generate the image. Red crosses indicate the values of parameters used for single-run examples included in the main article. Blue crosses indicate examples available in the Supplementary materials. The image is very similar to Figure 2. Where clustering conclusively happens, the noise proportion goes to 0. Where clustering is rare, noise proportion goes to 100%. The figure makes a region at low values of and high values of (it is darker here) stand out more than it does in Figure 1. This is the region where constant standard deviation matters more. Since noise generated due to is independent of the variance within clusters, more observations can end up between the tightly packed groups. This region is outlined by a lighter strip. The number of clusters in this part of parameter space is consistently low. Low number of clusters means higher total size of each cluster and hence a larger total variance. Larger clusters exhibit lower probability that their members that deviate more in absolute terms end up outside of the cluster and are identified as noise.

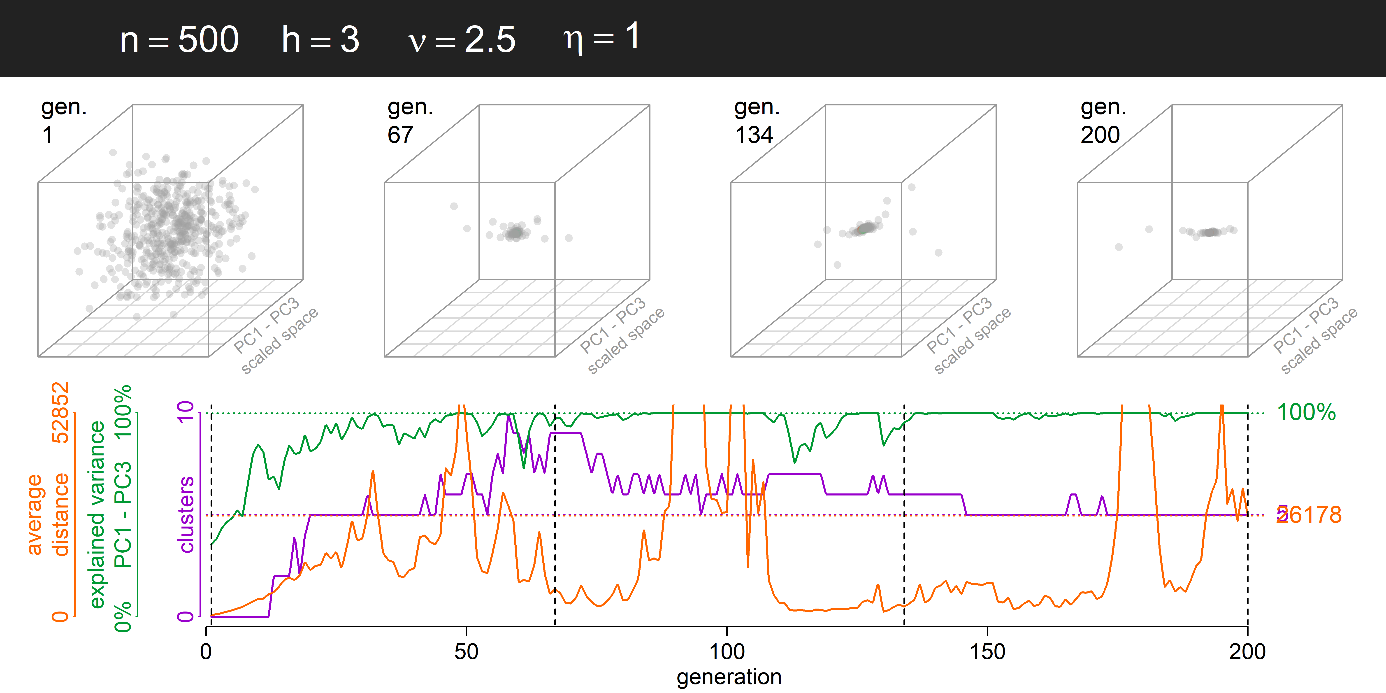
# Supplement S5. Example simulation runs



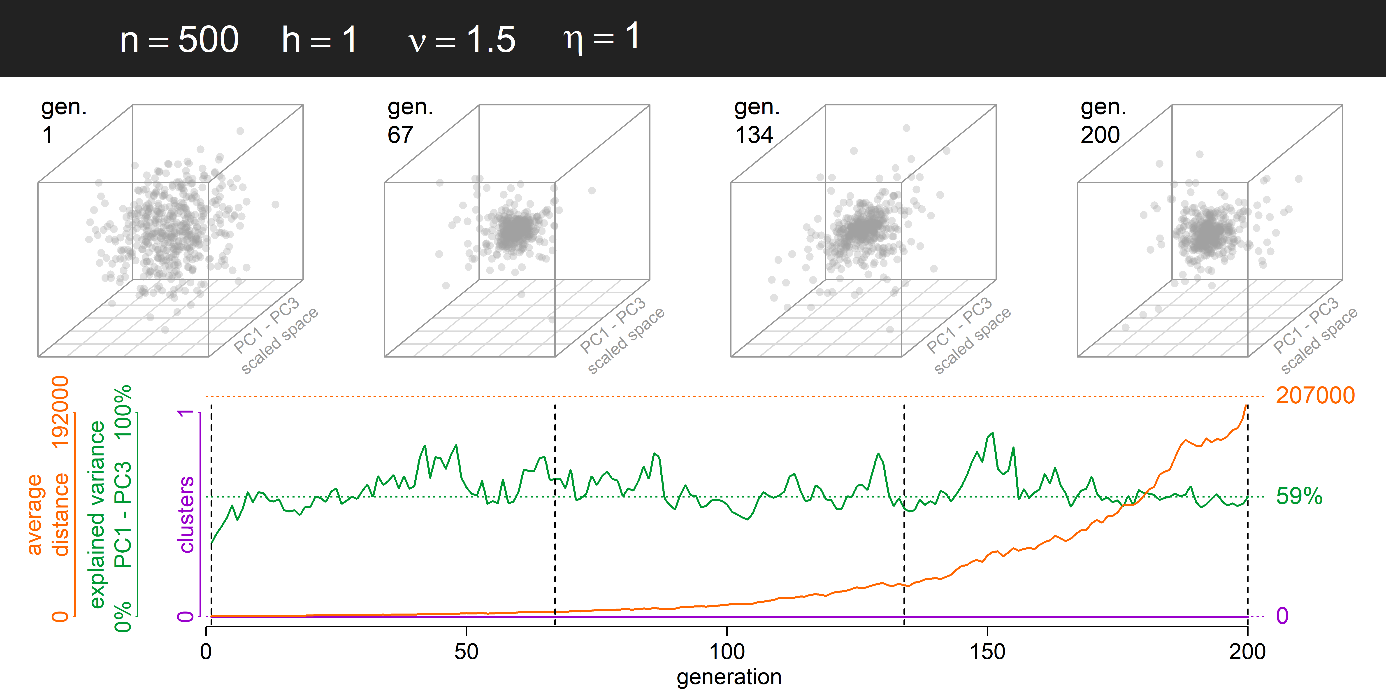
**Figure S5. Example simulations leading to stable distinct clusters, one simulation per small (A: n = 250), medium (B: n = 500), and large (C: n=1000) population.** All other parameters remain constant across the simulation runs (see Supplementary animations S05A–S05C).



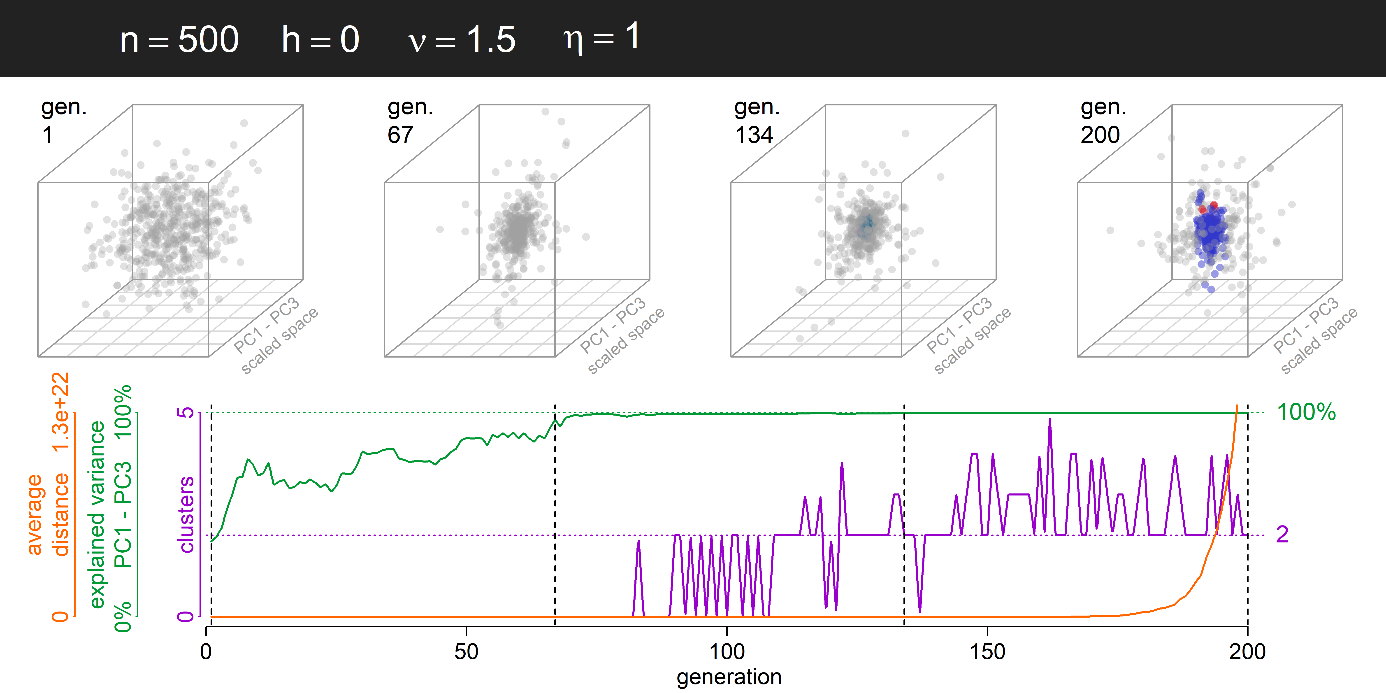
**Figure S6. Example simulations leading to stable distinct clusters, one simulation per small (A: n = 250), medium (B: n = 500), and large (C: n=1000) population** (see Supplementary animations S06A–S06C). We can clearly see that with a higher constant offspring standard deviation (), clusters are less stable. They tend to merge and eventually disappear. (Here, we see regress to a single cluster in A and C, but a similar scenario is possible also for n = 500).



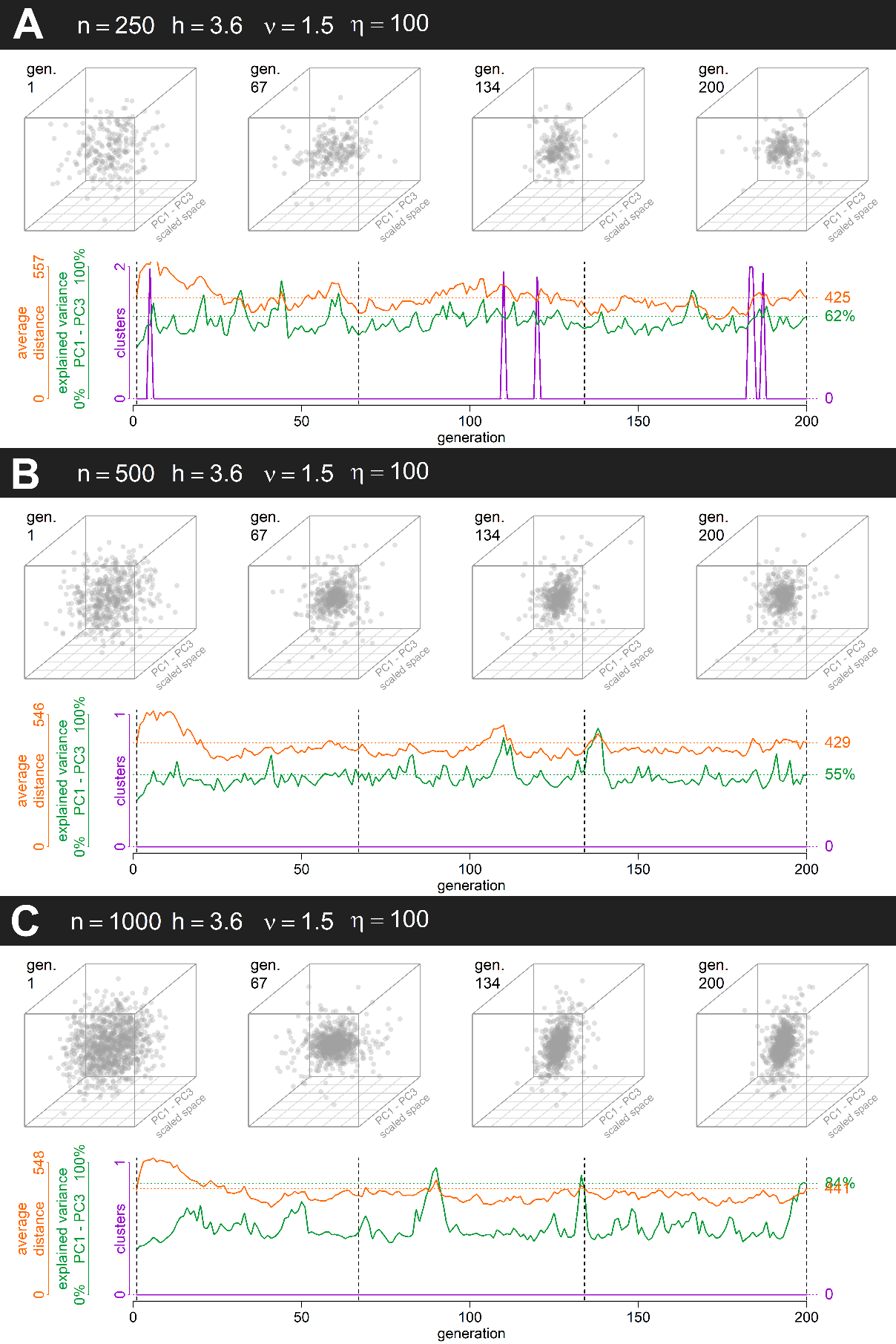
**Figure S7. Example simulation leading to a lower number of distinct clusters due to lower homophily (h = 3) and larger relative offspring variance (ν = 2.5)** (see supplementary animation S07).



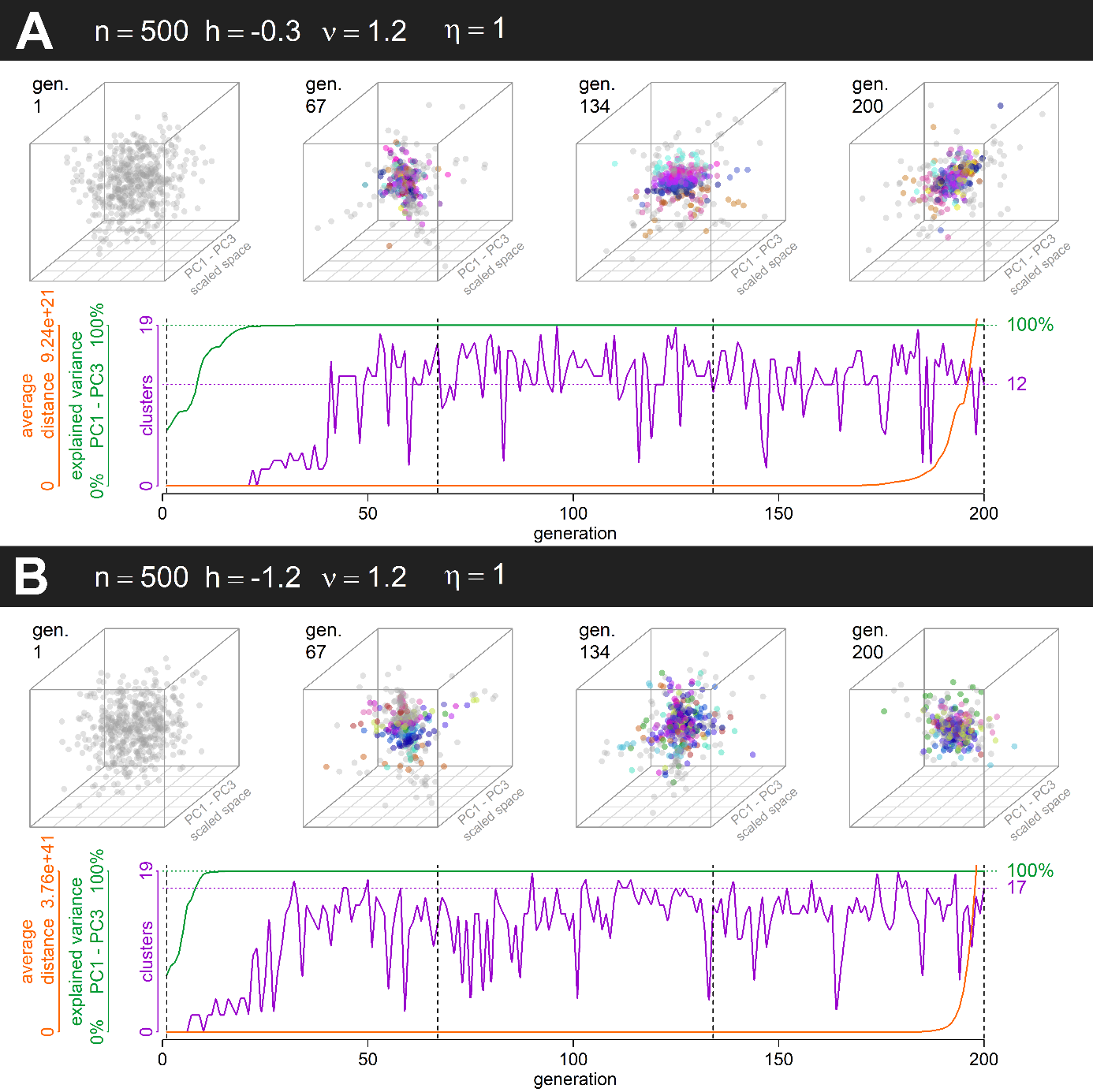
**Figure S8. Example simulation leading to a very low number of distinct clusters due to weak assortment between agents (h = 1)** (see supplementary animation S08).



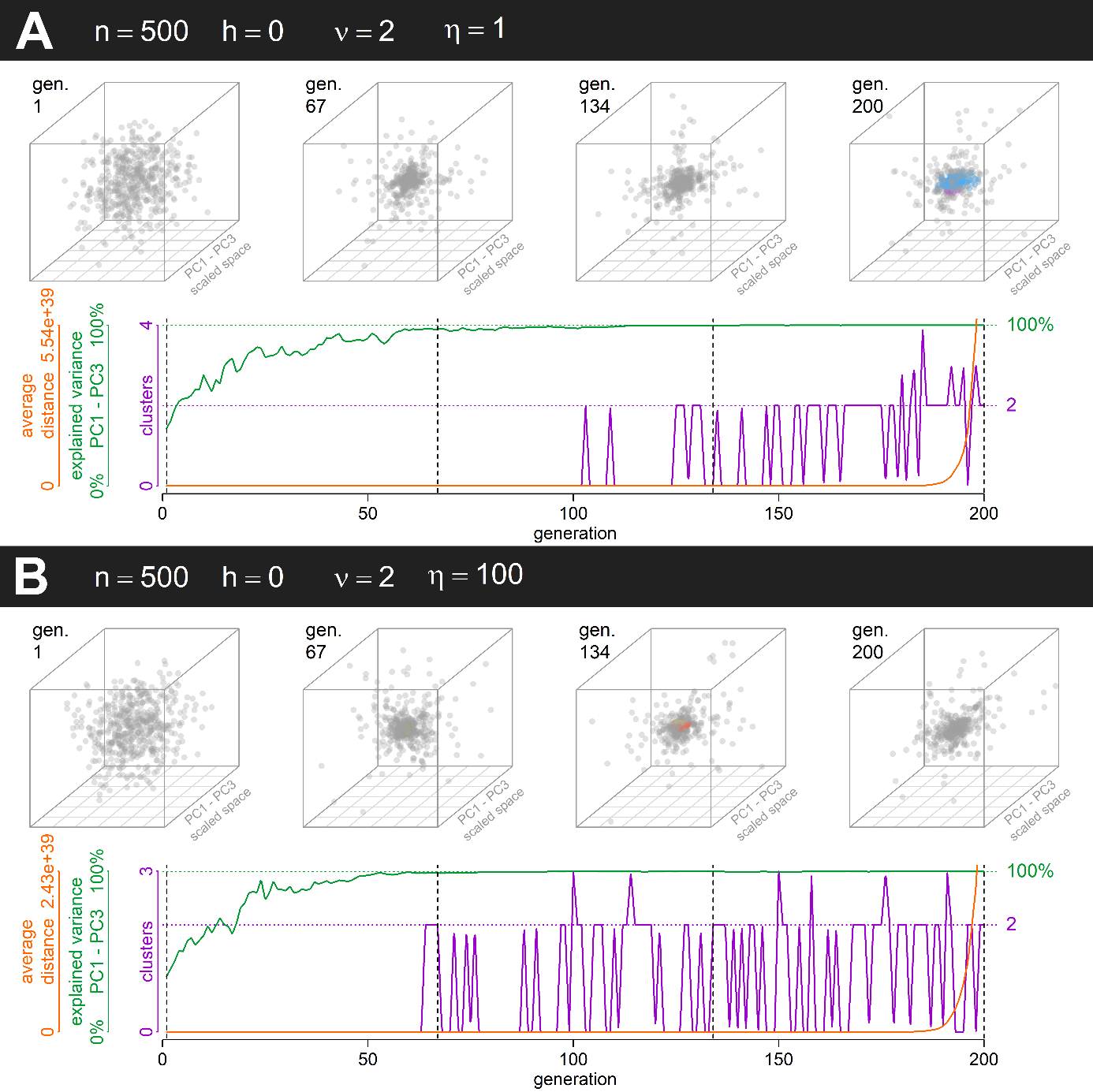
**Figure S9. Example simulation leading to variability explosion and weak clustering along a single dimension due to random interactions between agents (h = 0)** (see supplementary animation S09). Due to exponential variance explosion, it is in this case problematic to speak about the formation of subcultures, although the HDBSCAN algorithm is able to detect distinct clusters aligned along a single prominent culture-space dimension. In real life, of course, it is hard to imagine a trait, or a vector constituted by a combination of traits, along which variance (or average distance between points) could grow unrestricted to magnitude 1022. Results such as this should be interpreted in more general way: If homophily is low and coefficient for relative variance is high, we expect cultural variance to grow and the society to polarise. Notice, that the reduction of effective culture space dimensionality is present also in the modified version of the model, where the population configuration is normalised after every generation to hold the average distance between individuals constant (Figure S16).



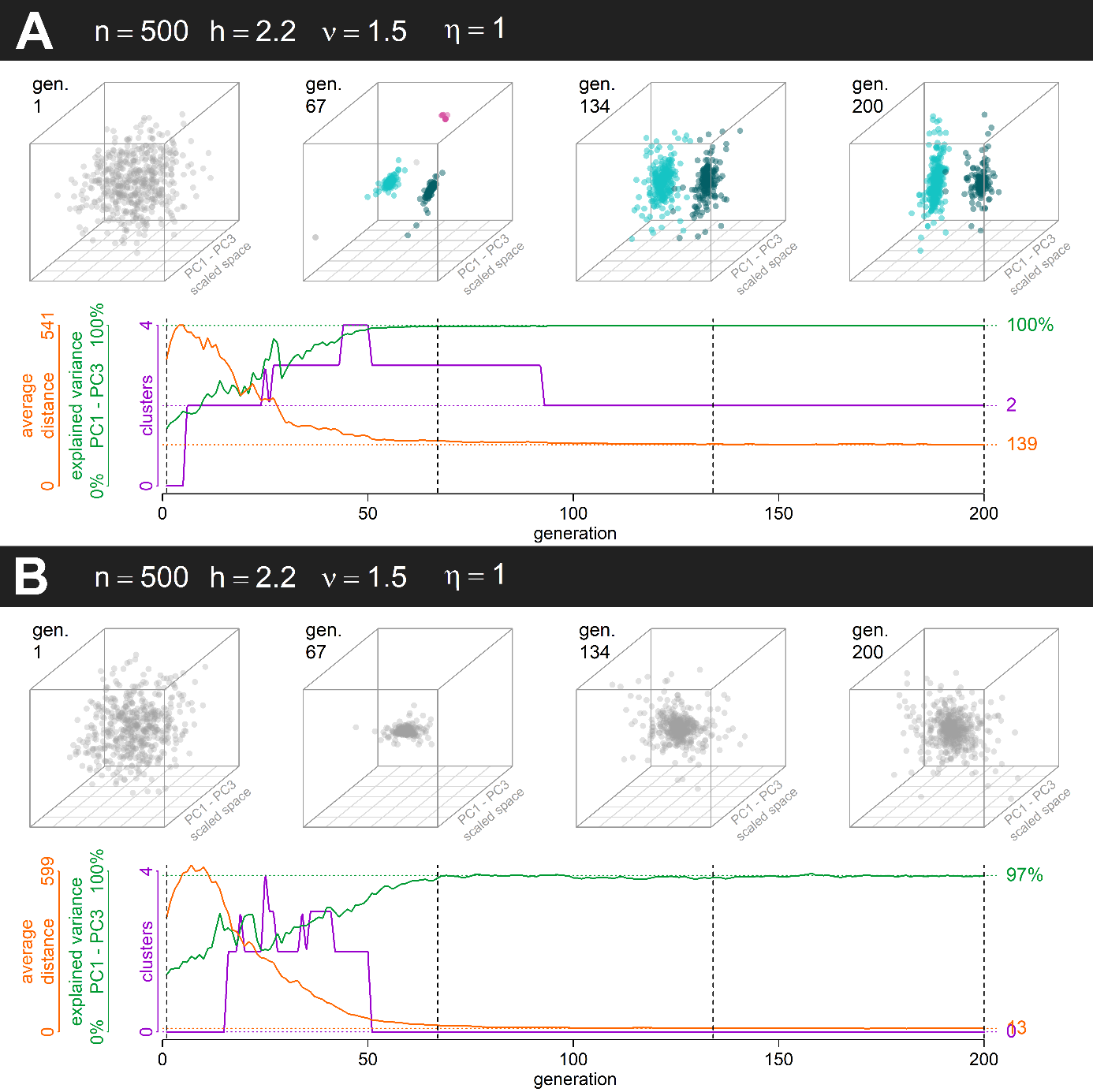
**Figure S10. Example simulations with a single stable cluster due to a dominant influence of Galton-Pearson inheritance (η = 100) despite high homophily (h = 3.6) and high relative offspring variance (ν = 1.5); one simulation per small (A: n = 250), medium (B: n = 500), and large (C: n=1,000) population** (see Supplementary animations S10A-S10C).



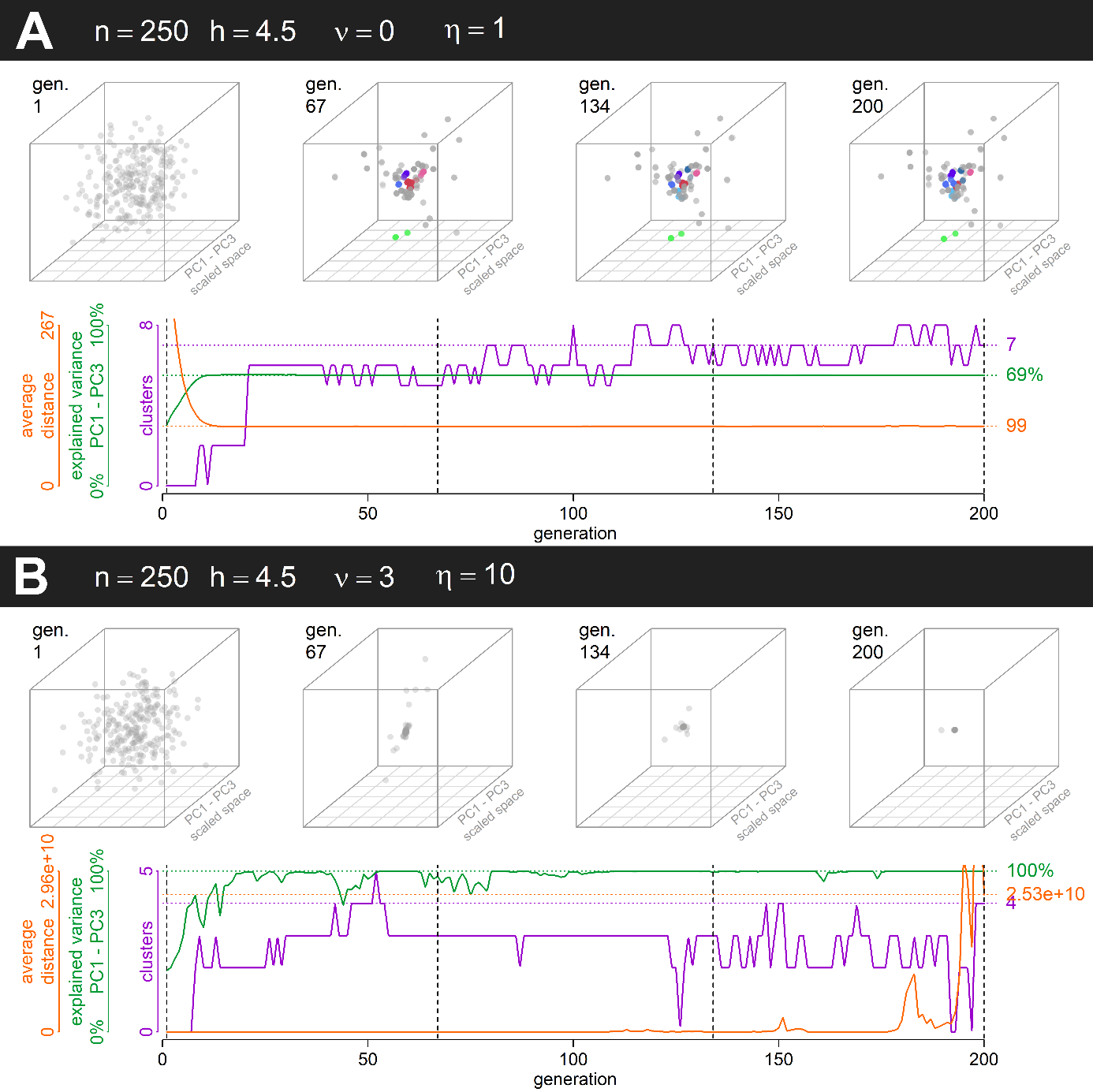
**Figure S11. Example simulations showing what happens when homophily is negative (A: h = -0.3, B: h = -1.2) and relative offspring variance is higher than 1 (ν = 1.2).** Due to the absence of stabilising selection (e.g., Gaussian fitness function decreasing the probability that an individual is selected as a role model proportionally to its distance from the coordinates origin), we see a rapid spread along a single effective dimension of the culture space(see Supplementary animations S11A and S11B). Due to the exponential variance explosion, it is in this case problematic to speak about the formation of subcultures, although HDBSCAN algorithm is able to detect distinct clusters aligned along a single prominent culture-space dimension.



**Figure S12. Example simulations that show a rapid spread along a single effective culture space dimension despite random assortment.** This explosion is driven by a high relative offspring variance (ν = 2). In such case, the role of constant offspring variance (A: η = 1, B: η = 100) is negligible(see supplementary animation S12A and S12B). Due to exponential variance explosion, it is in this case problematic to speak about the formation of subcultures, although HDBSCAN algorithm is able to detect distinct clusters.



**Figure S13. Example simulation of a system with parameter combination that leads to a general loss of cultural variance but has a chance to form a small number of polarised clusters.** Such parameter combinations typically lie on the ‘black bridge’ apparent on panels for and in Figure S2. The reduction in dimensions is due to the emergence of just two stable clusters. This ‘polarisation of the population’ is typically reflected in a PC1 that monopolises most of the variance.In A, there emerged two distinct clusters which helped preserve some of the overall variance (see also the high proportion of variance explained by PC1–PC3). That was not, however, the universal outcome. All variance was lost where the population failed to form stable clusters (regardless of identical initial conditions). In such case, the population collapsed into a single point (see Supplementary animations S13A and S13B).



**Figure S14. Extreme simulation runs which demonstrate that a small does not always prevent the emergence of clusters (A) when constant variance is small, and that excessively large can in effect limit the number of clusters (B), because the overall new variance in each generation is huge despite being reasonably small.** (See Supplementary animations S14A and S14B.)

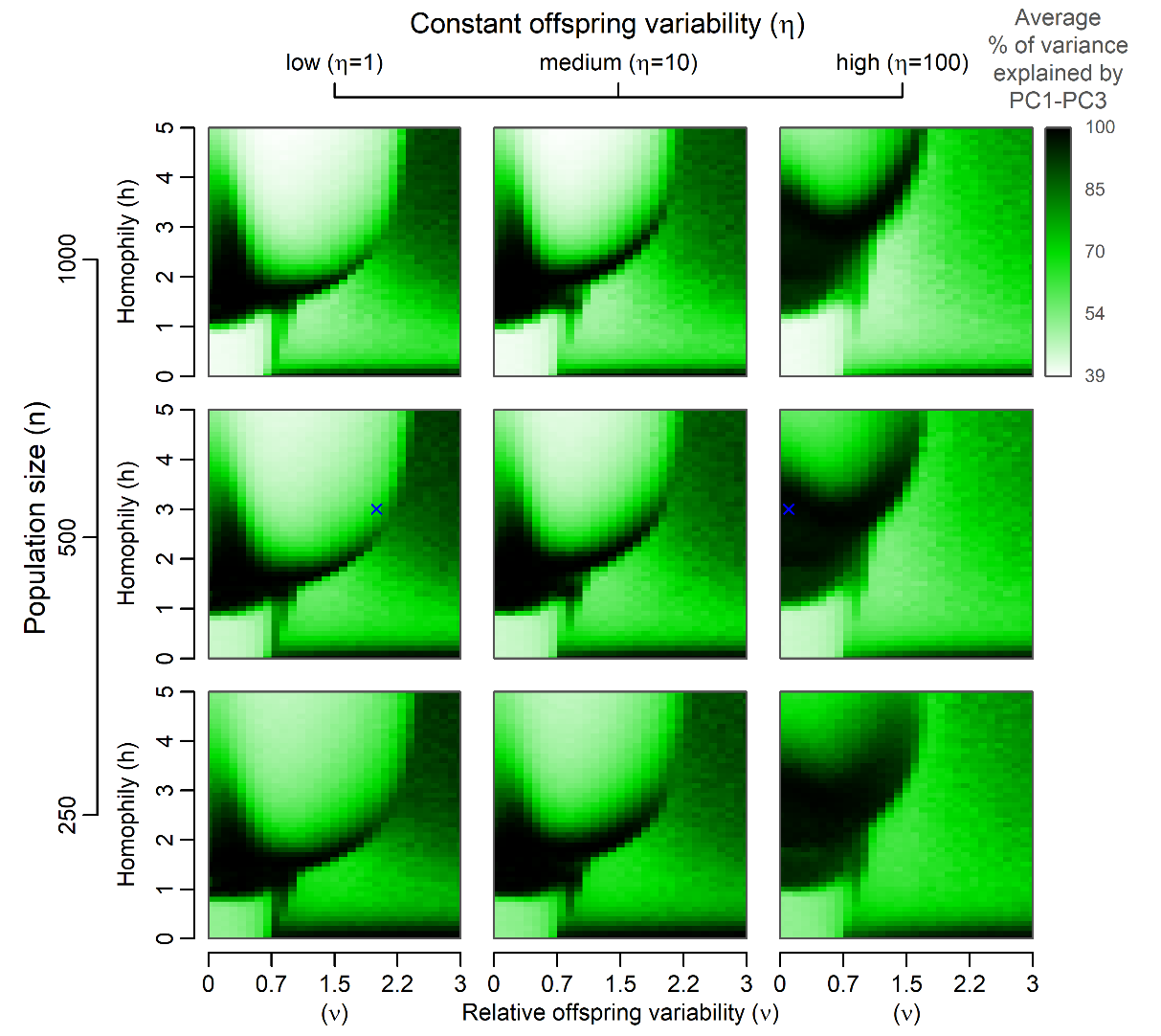
# Supplement S6. Results of a model where agent distance is normalised after each step

The model presented in the main manuscript has the vast advantage of being simple and straightforward. It does not introduce any additional assumptions besides the pairing and inheritance algorithm. This simplicity, however, does have some side effects. For example, for many combinations of parameter values (when the configuration of agents’ position in the trait space is unconstrained), the basic system leads to one of two extreme configurations: in the first, all positions collapse into a tiny area within the trait space, where the pairing is as good as random (in case both and are small), or, in the second case, the configuration explodes to infinity along a single effective dimension of the trait space (in case the is excessively large or is positive and negative). Both of these extreme outcomes can be avoided if the entire configuration is normalised without distortion to maintain a constant average distance between agents in the trait space. This allows us to focus on changes in the relative configuration of agents without having to deal with variability loss or explosion. (In fact, though, a loss of all cultural variability may be a real possibility for the natural phenomena in question but, for the moment being, we can leave it aside.) Most findings, however, remain, unchanged regardless of the model.

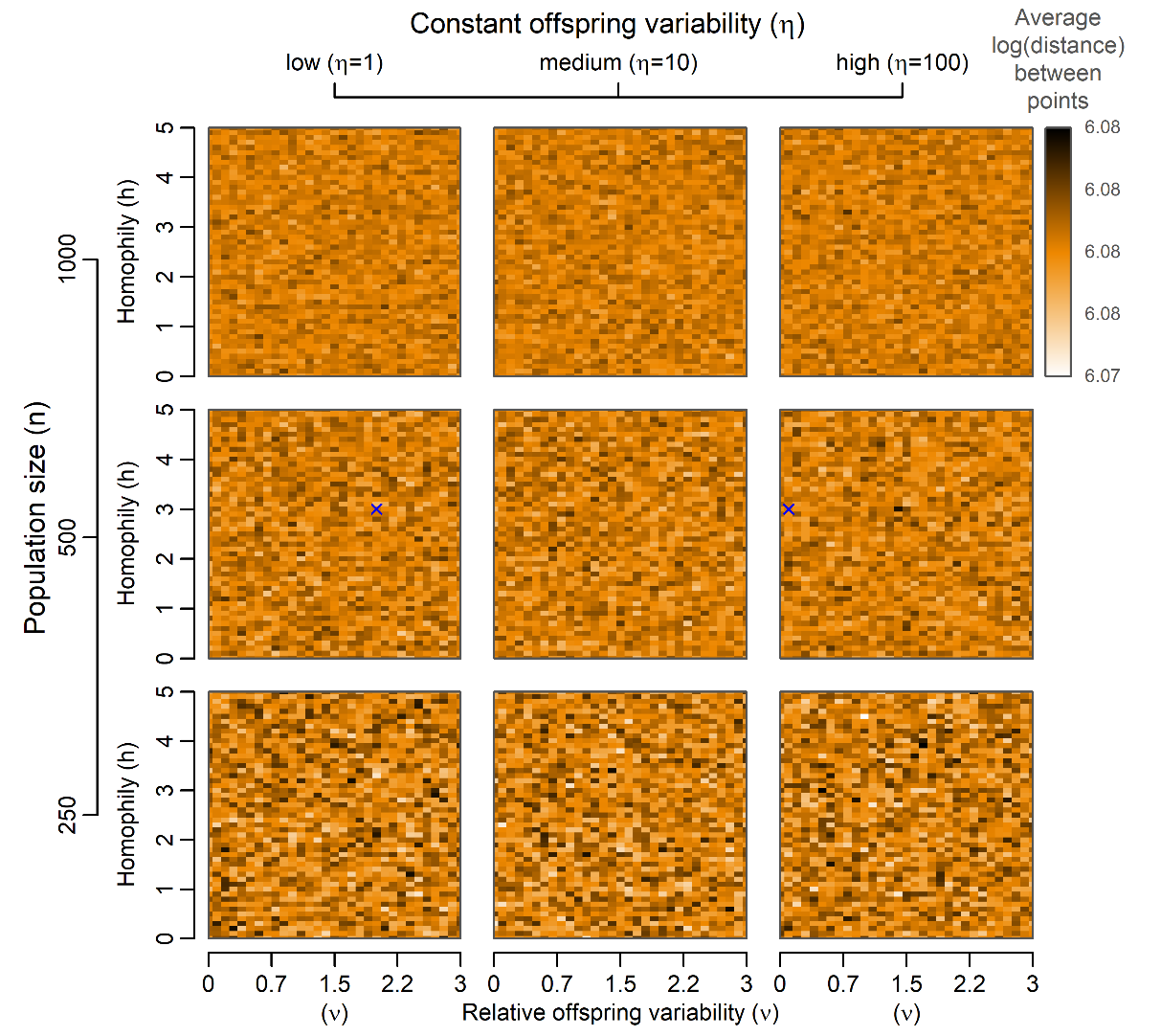
In the results below, the methods are the same as in the main manuscript, only after each generation, the point coordinates are normalised to average distance of 436, which is approximately the average distance between normally distributed points in 10 dimensions with sd = 100 along each dimension.

****

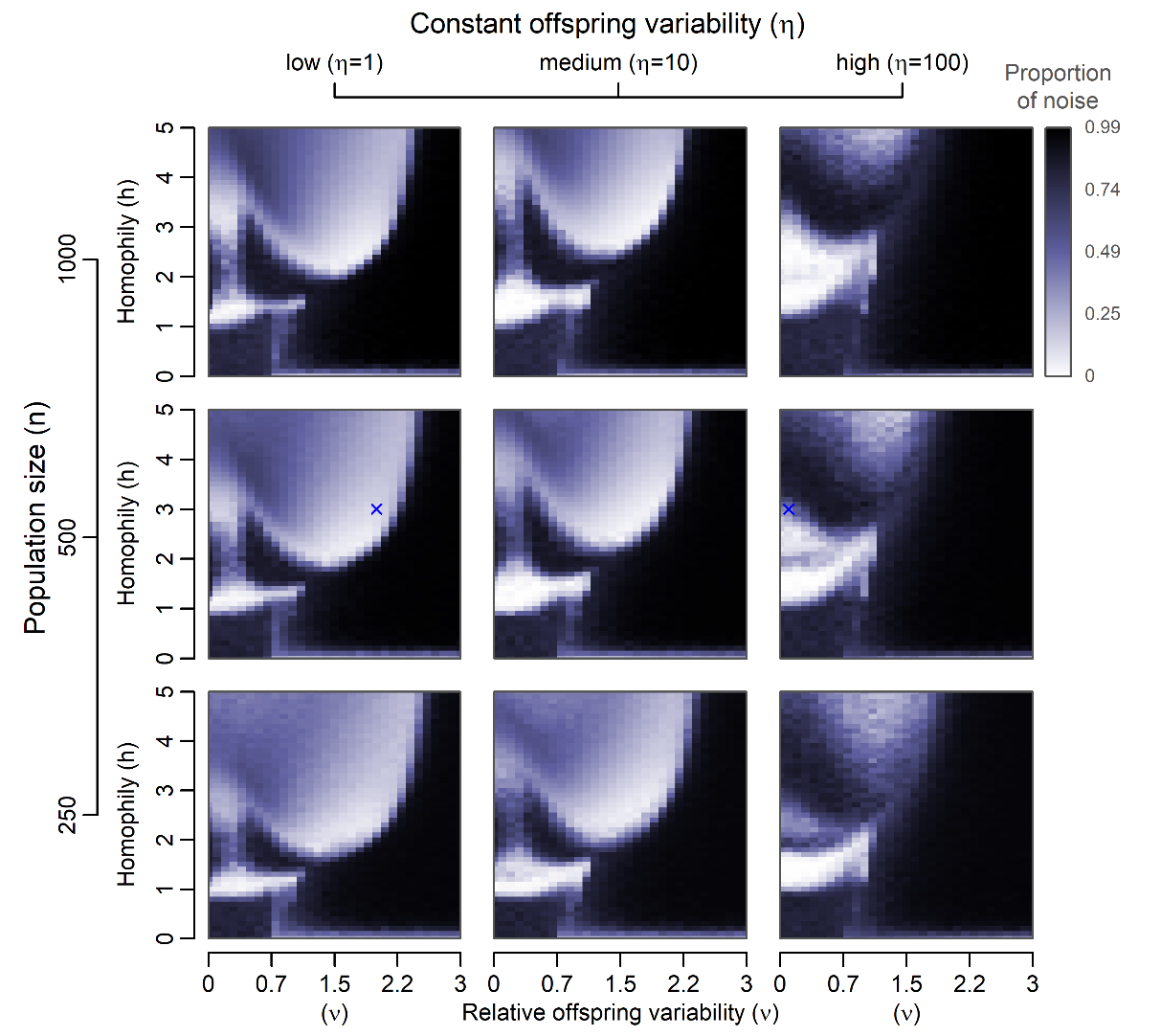
**Figure S15. A graphical summary of the tendency to form subcultures after 200 model generations in a system with normalised average distance**. The points in the 10-dimensional culture space were normally distributed across all dimensions at the beginning of each simulation run and 100 simulation runs were executed for each combination of parameters. Blue crosses indicate examples available below. Notice the new region in which clustering is possible: Region with intermediate homophily and low coefficient of proportional variability . This is a region where polarisation along one dimension is expected (Figure S16) and there is a very little noise in between clusters (Figure S18), which means that this is an inflated “black bridge” from Figure S2, where emergence of two large clusters influences the outcome substantially. With normalisation occurring after each simulated generation, distances change proportionally to their magnitude. Similar to a system with PVDI and without normalisation, larger distances become larger, which may cause polarisation along a single dimension. Where intermediate homophily keeps points mostly within their clusters but scarce interactions across two clusters allow for the emergence of new variation along the dimension that connects them, the similarity between a system with PVDI and a system with low and normalisation is most prominent.



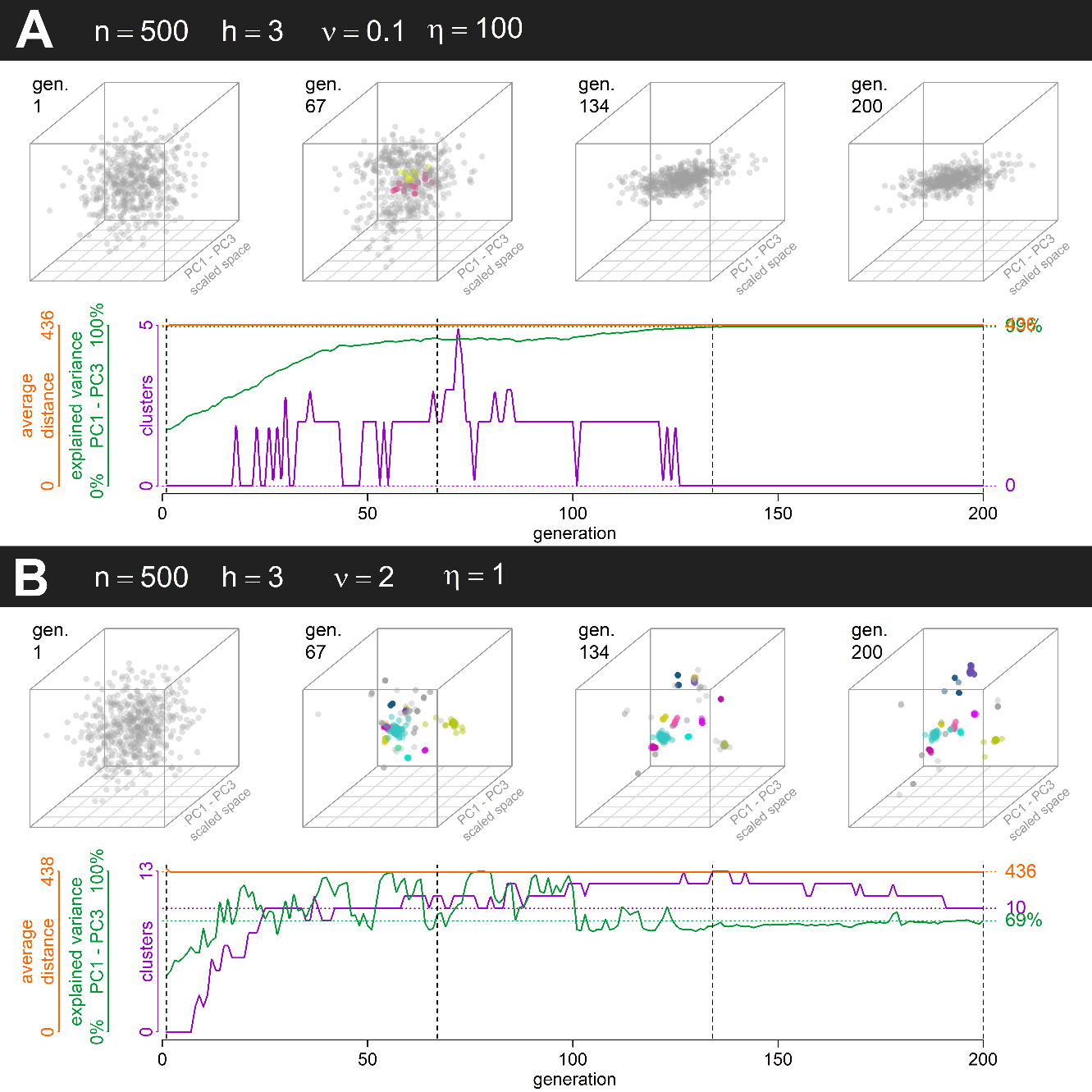
**Figure S16. A graphical summary of expected reduction in the number of effective dimensions after 200 model generations in a system with normalised average distance**. The same set of simulation runs as in Figure S15 was used to generate the image. PC1–PC3 stand for the first three principal components. See the explanation of dimension reduction in a system with large and small in the caption of Figure S19. Blue crosses indicate examples available below.



**Figure S17. A graphical summary of the expected average distance between agent positions in a culture space** **after 200 model generations in a system with normalised average distance**. The same set of simulation runs as in Figure S15 was used to generate the image. Blue crosses indicate examples available below.



**Figure S18. A graphical summary of the proportion of noise (points that are not a part of any cluster)** after 200 model generations. The same set of simulation runs as in Figure 15 was used to generate the image. Blue crosses indicate examples available below.



**Figure S19. Two simulation runs in a system with normalised average distance, one with a strong influence of Galton-Pearson inheritance (A) and one with a strong influence of PVDI (B).** For detailed description, see the caption of Figure 3 in the main article. (See Supplementary animations S19A and S19B.) Simulation run parameters were chosen to match those of Figure 3. For the sake of brevity, equivalents of Figures S5–S14 are not included but can be easily obtained using the code at <https://doi.org/10.17605/osf.io/pvyhe>

Notice the clustering tendencies and dimension reduction in a system with a large influence of Galton-Pearson inheritance and homophily. The polarisation is due to an inflation of the originally small differences that come up due to the normalisation after every generation (see also the explanation in the caption of Figure S15 and notice that this parameter combination appears on the edge of the new region of clustering which can be best seen in Figure S18).

# Supplement S7. Results with a random variation along all dimensions of the trait space

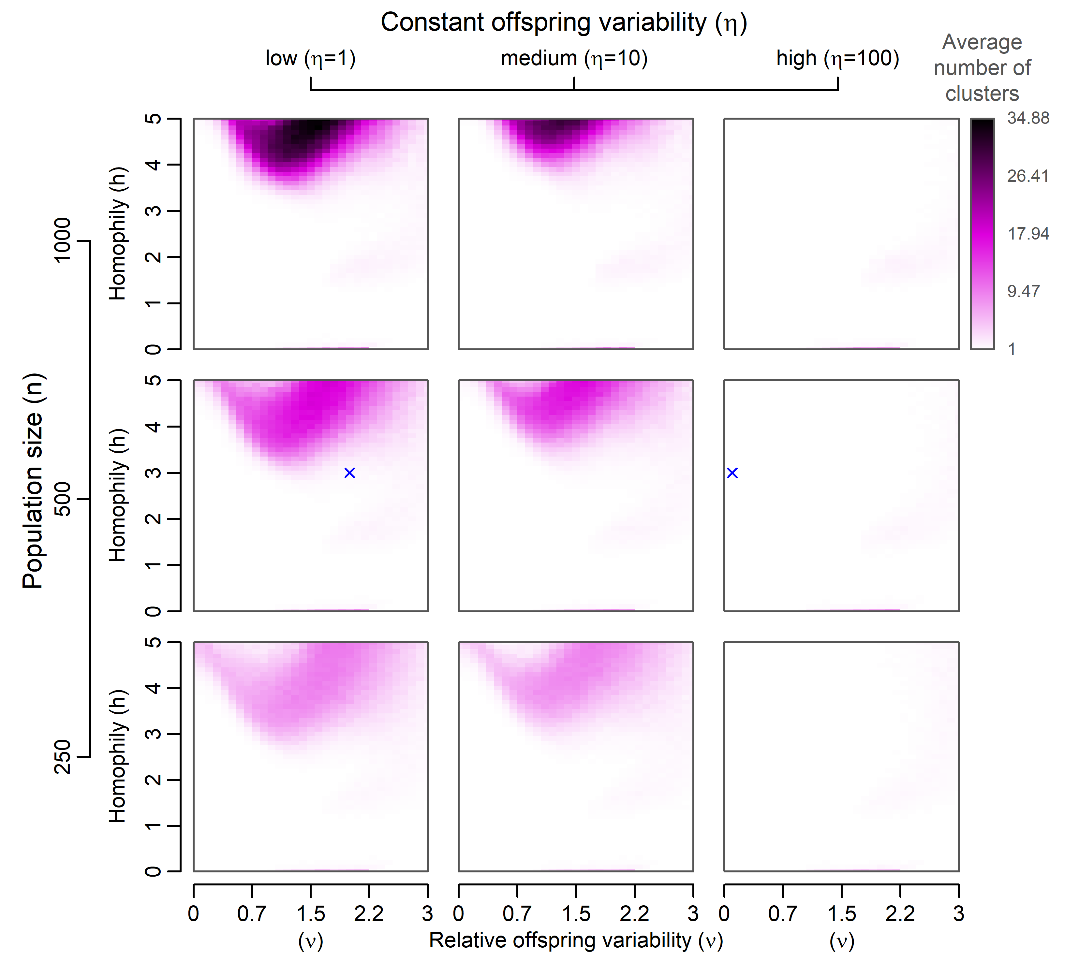
One could argue that a model where all random variation is realised along the vector linking parental positions is not realistic. It is trivial to modify the inheritance part of the model summarised in Eq. (3) of the main manuscript and add random variation along the D dimensions of the trait space, i.e.,

|  |  |  |
| --- | --- | --- |
|  |  |  |

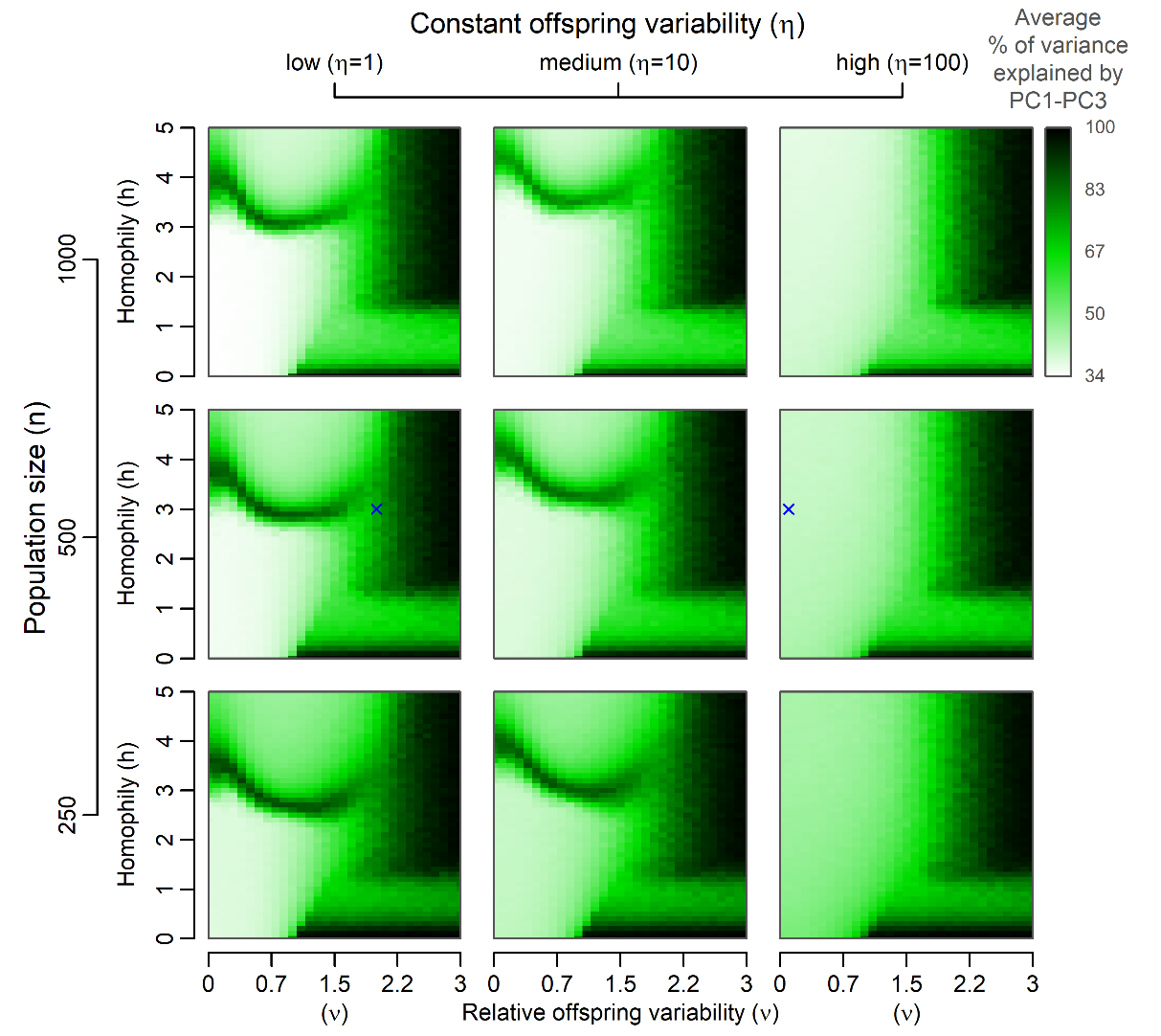
where with being a random phenotypic mutation of the offspring independent of parental positions.

The assumption of a larger variation (proportional to the parental distance for ) along the vector connecting parental positions is preserved.

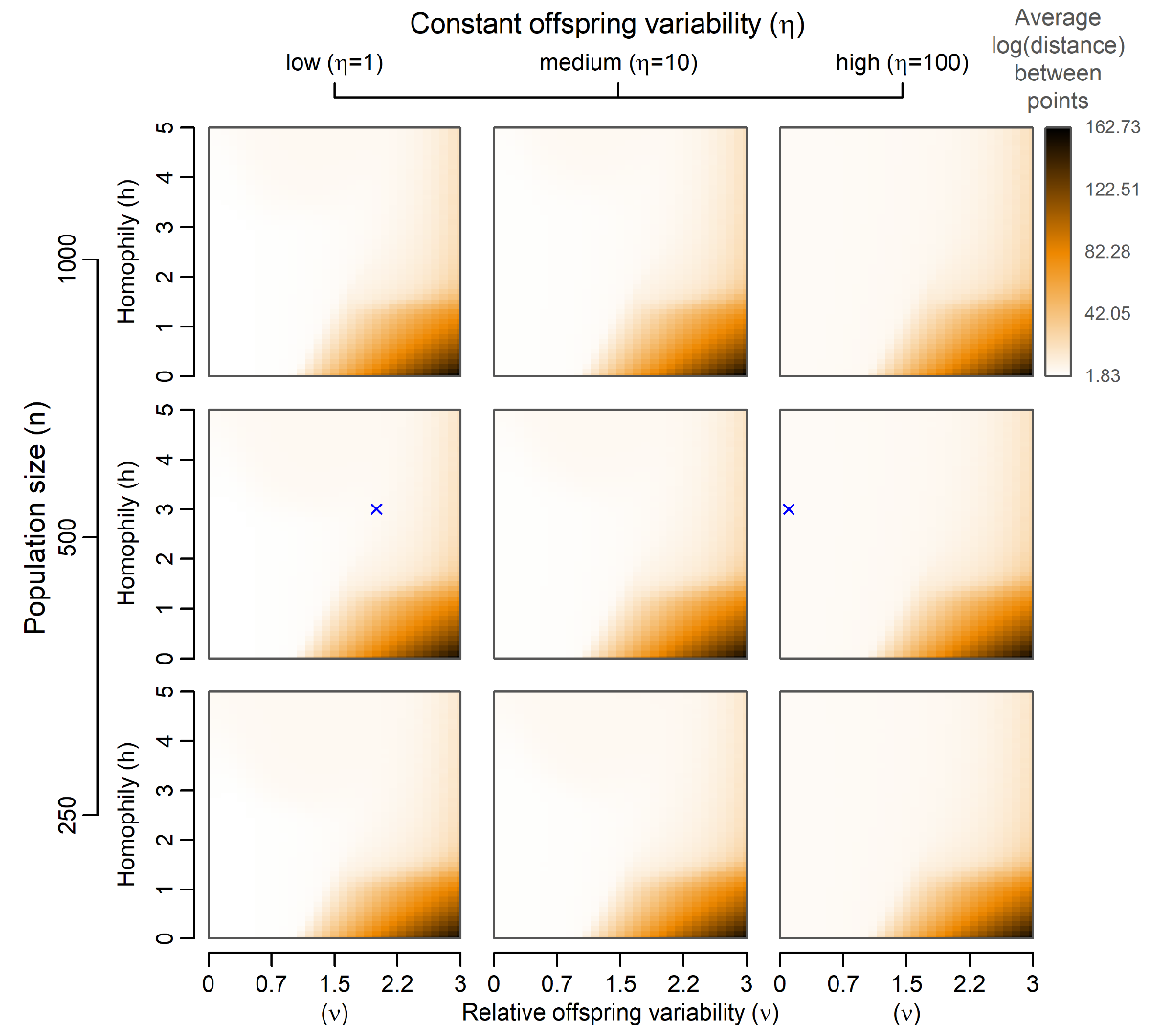
Results after this slight modification for are included below.

****

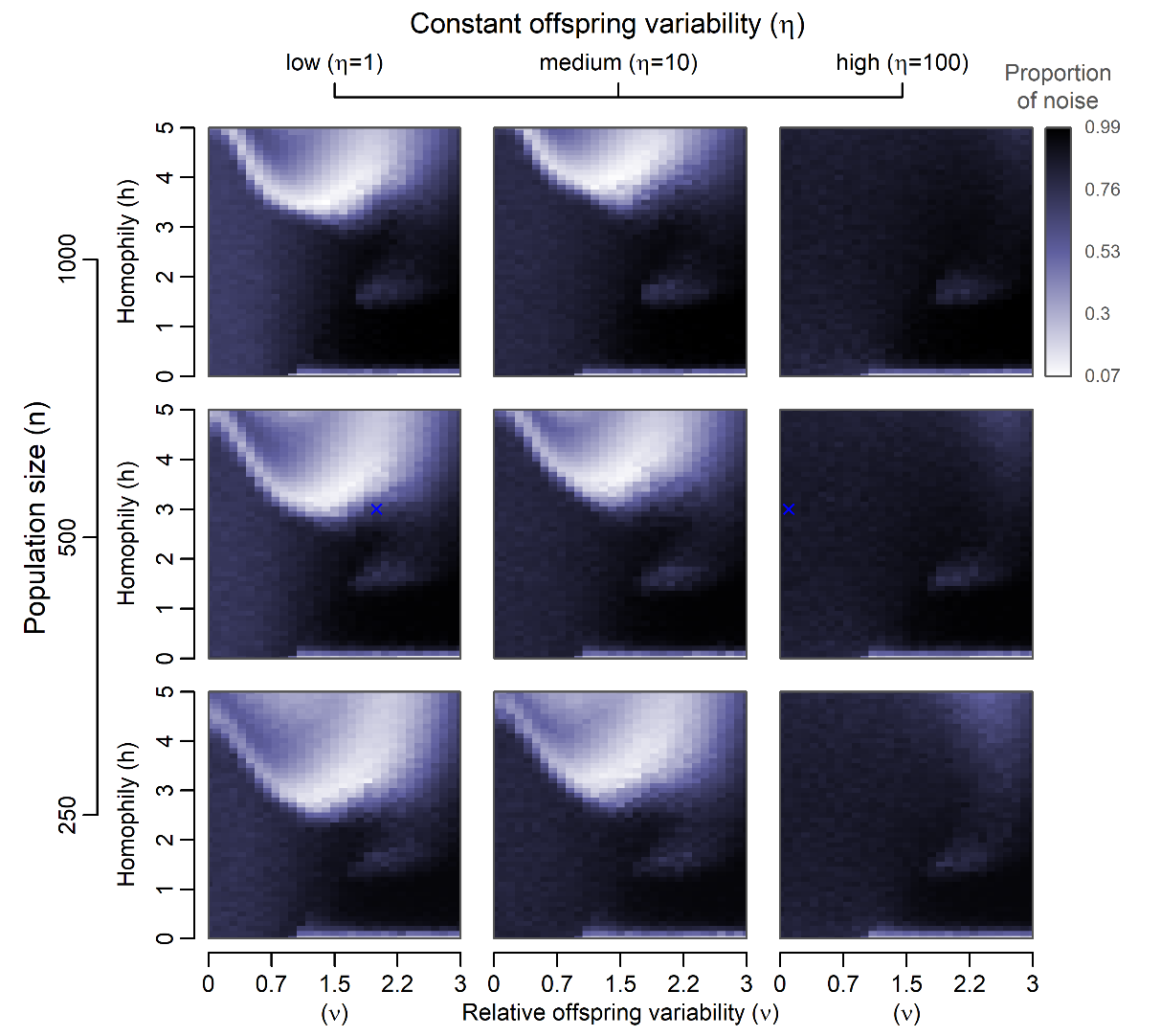
**Figure S20. A graphical summary of the tendency to form subcultures** **after 200 model generations in a system with additional normal noise**. The points in the 10-dimensional culture space were normally distributed across all dimensions at the beginning of each simulation run and 100 simulation runs were executed for each parameter combination. Blue crosses indicate examples available below.



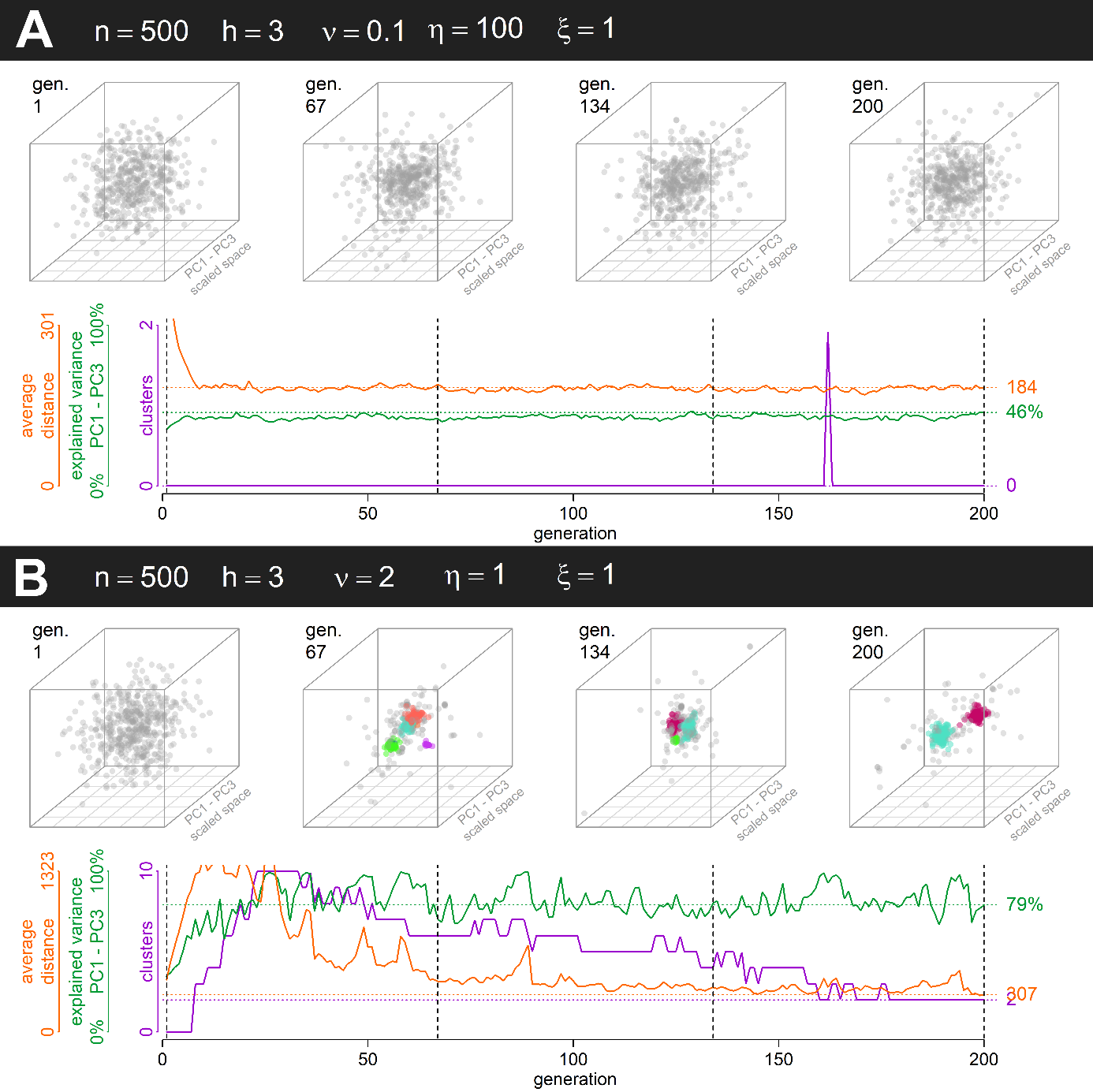
**Figure S21. A graphical summary of expected reduction in the number of effective dimensions after 200 model generations in a system an additional normal noise.** The same set of simulation runs as in Figure S20 was used to generate the image. PC1–PC3 stand for the first three principal components. Blue crosses indicate examples available below.



**Figure S22. A graphical summary of the expected average distance between agent positions in a culture space after 200 model generations in a system with additional random noise**. The same set of simulation runs as in Figure S20 was used to generate the image. Blue crosses indicate examples available below.



**Figure S23. A graphical summary of the proportion of noise (points that are not a part of any cluster)** after 200 model generations. The same set of simulation runs as in Figure 20 was used to generate the image. Blue crosses indicate examples available below.



**Figure S24. Two simulation runs in a system with additional random noise, one with a strong influence of Galton-Pearson inheritance (A) and one with a strong influence of PVDI (B).** For detailed description see the caption of Figure 3 in the main article. Simulation run parameters were chosen to match those of Figure 3 Equivalents of Figures S5–S14 are not included for brevity, but can be easily obtained using the code at <https://doi.org/10.17605/osf.io/pvyhe> (See Supplementary animations S24A and S24B.)