1   **S1 Supplemental Methods. Generating virtual community structure by using the**

2   **ComGen package**

3   ComGen, a generator of a virtual community structure, is a collection of codes written in Perl. It

4   generates a dataset of the presence and absence of multiple species at multiple sites. The output

5   of ComGen is a two-dimensional matrix, the rows of which represent sites and the columns of

6   which represent species. The elements of the matrix are either 1 (present) or 0 (absent) (Fig A).

7

	Species 1	Species 2	Species 3	Species 4	⋯
Site 1	0	1	0	0	
Site 2	1	1	0	0	
Site 3	0	1	0	0	
Site 4	1	1	0	1	
Site 5	0	1	1	0	
Site 6	0	0	0	0	
Site 7	1	1	0	0	
Site 8	1	0	0	1	
Site 9	0	0	1	0	
Site 10	0	0	0	1	
⋯					

8   Fig A

9

10   The parameters controlling the pattern of the number of present sites among species, the

11   species richness pattern among sites, and the environmental preferences of individual species

12   can be manipulated. Before generating the dataset using ComGen, parameter values are set in

13   sequence as follows, because these parameters are interrelated.

14   First, the number of sites is set. Sites are assumed to be arranged uniformly along a

15   hypothetical environmental gradient, such as temperature.

16   Second, the total number of species and number of presence sites (N_p) of individual species

17   are set. By default, the plot of N_p of each species against the rank order of N_p of each species

18   (i.e., a rank-abundance curve) is assumed to follow a log-linear pattern (Fig B). The variation of

19   N_p among species is controlled by manipulating the slope of the log-linear relationship. If

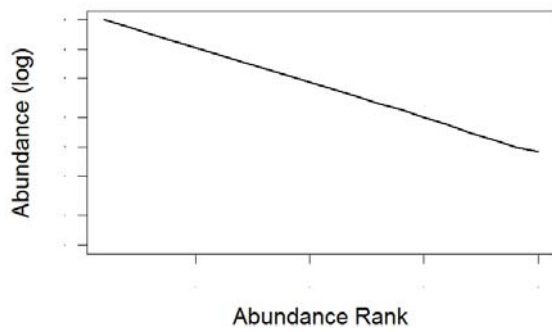20   necessary, the N_p of individual species can be chosen as desired.

21

Abundance (log) / Abundance Rank   Fig B

22

23   Third, species richness (i.e., the number of species) at each site is determined. The species

24   richness pattern among sites is assumed to follow a log-linear pattern (Fig C). The mean

25   richness is adjusted so that the sum of the species richness over all the sites is equal to the sum

26   of the N_p values of all species. The variation of species richness among sites is controlled by

27   manipulating the slope of the log-linear relationship. Then the determined species richness is

28   assigned to individual sites randomly, irrespective of the environmental conditions of the sites.
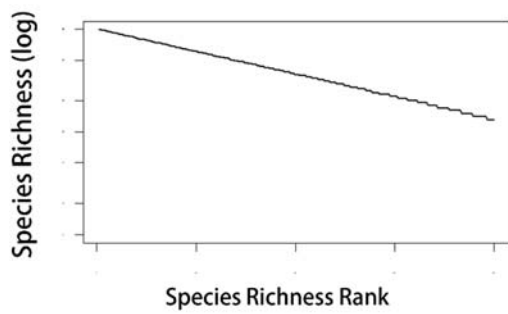
29

Fig C

30

31

32   Fourth, habitat preference (i.e., the range of suitable environments) is determined for each

33   species. The centers of the habitats of all species are located randomly over the entire

34   environmental gradient (Fig D). The width of the habitat preference of each species depends on

35   the number of presence sites of the species. More abundant species have wider habitat
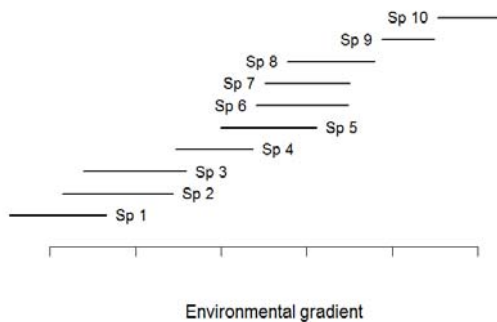
36   preferences.

Fig D

37

38

39   Once the parameters are set, ComGen generates a presence/absence dataset. It first

40   generates a tentative distribution of each species by randomly assigning the predetermined

41   number of presence sites, considering the number of species present at each site, but without

42   considering the environmental preference of the species. Next, we adjust the random tentative

3

43   distribution to reflect the environmental preference of each species. We randomly select two

44   sites, and a pair of species each of which is present in one of the two sites are swapped, if the

45   swapping makes the distribution more consistent with the environmental preferences. This

46   process is repeated until there is no more swapping to improve consistency with the

47   environmental preferences. The richness pattern and rank-abundance curve are kept unchanged

48   through the swapping process. Absence sites of each species are defined as the sites where the

49   presence of the species is not assigned.

50   The distribution of each species is not necessarily confined to its potential habitat after the

51   adjustment. There may be no solution satisfying all requirements if the values of parameters are

52   incompatible. Thus, the generated dataset was checked for consistency with the environmental

53   preference of each species by examining the ratio of presence data outside habitats and graphic

54   expression of the distribution of individual species before conducting a conservation area

55   selection analysis. There was no case in which the environmental preference of a species was

56   violated in the parameter range we used in the analysis.

57

58

59