

Figure S1. Comparison of the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas using 100 tanglegrams of Set50, additive trees, with N (number of runs) = 10^5 runs vs. $N = 10^4$. Parameter and test combinations: (a) Geodesic distances (GD), $n = 5$; (b) PACo, $n = 5$; (c) GD, $n = 10$; (d) PACo, $n = 10$; (e) GD, $n = 20$; (f) PACo, $n = 20$. Filled red points, $p = 1\%$; blue circles, $p = 5\%$. The line represents $y = x$.

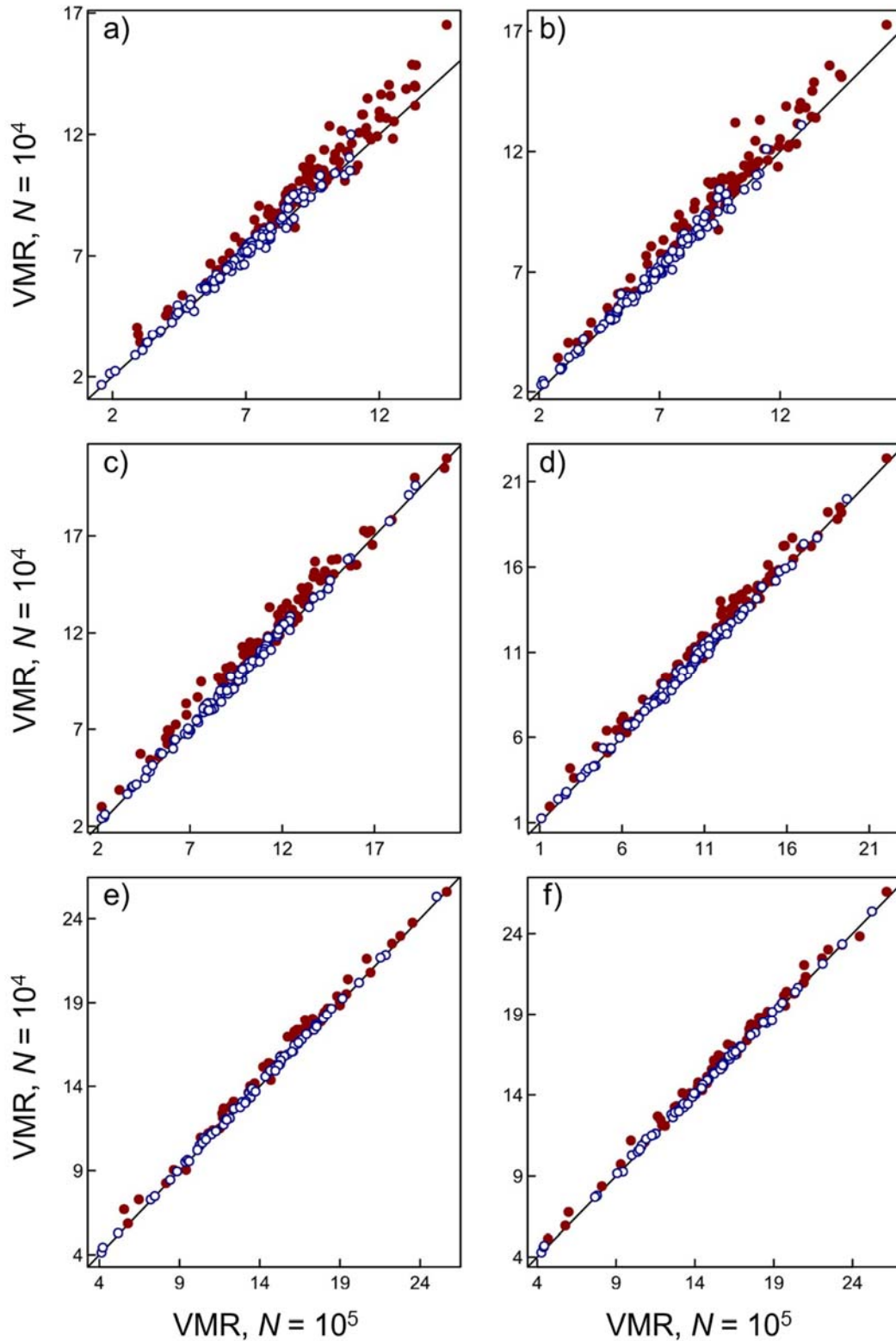


Figure S2. Comparison of the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas using 100 tanglegrams of Set100, additive trees, with N (number of runs) = 10^5 runs vs. $N = 10^4$. Parameter and test combinations: (a) Geodesic distances (GD), $n = 5$; (b) PACo, $n = 5$; (c) GD, $n = 10$; (d) PACo, $n = 10$; (e) GD, $n = 20$; (f) PACo, $n = 20$. Filled red points, $p = 1\%$; blue circles, $p = 5\%$. The line represents $y = x$.

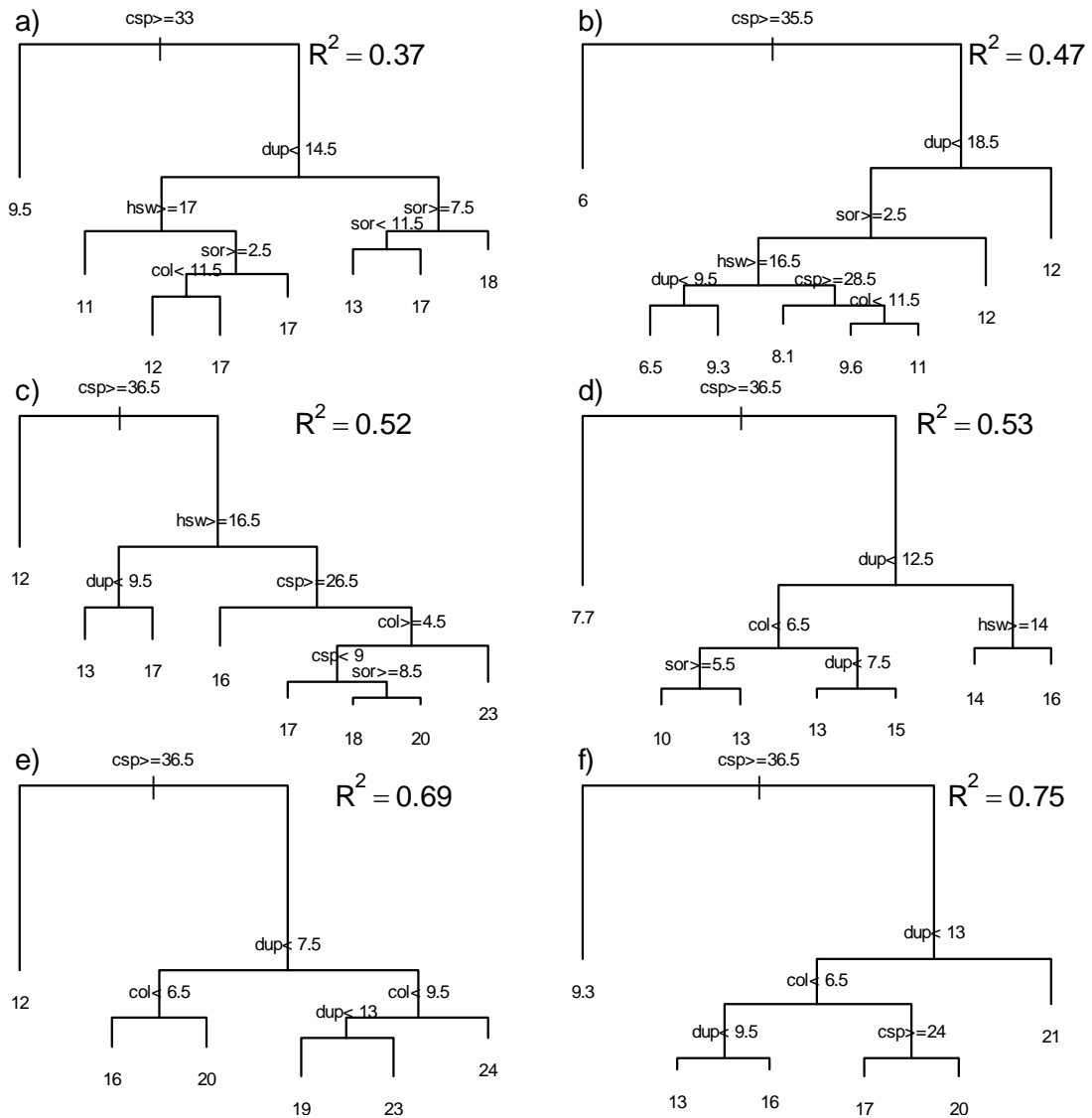


Figure S3. Regression trees relating the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas with the number of evolutionary events in simulated tanglegrams of Set50, additive trees, applying PACo as global-fit method. Expressions on nodes indicate the split criterion that shows the threshold value of the evolutionary event that was used to create the split. Branches to the left include cases that fulfill the criterion. Numbers at terminals represent the average VMR of the group. Parameter combinations were as follows: (a) $n = 5$, $p = 1\%$; (b) $n = 5$, $p = 5\%$; (c) $n = 10$, $p = 1\%$; (d) $n = 10$, $p = 5\%$; (e) $n = 20$, $p = 1\%$; (f) $n = 20$, $p = 5\%$. Event abbreviations: csp = cospeciation, sor = sorting, dup = duplication, hsw = complete host-switching, col = colonization of new host without speciation.

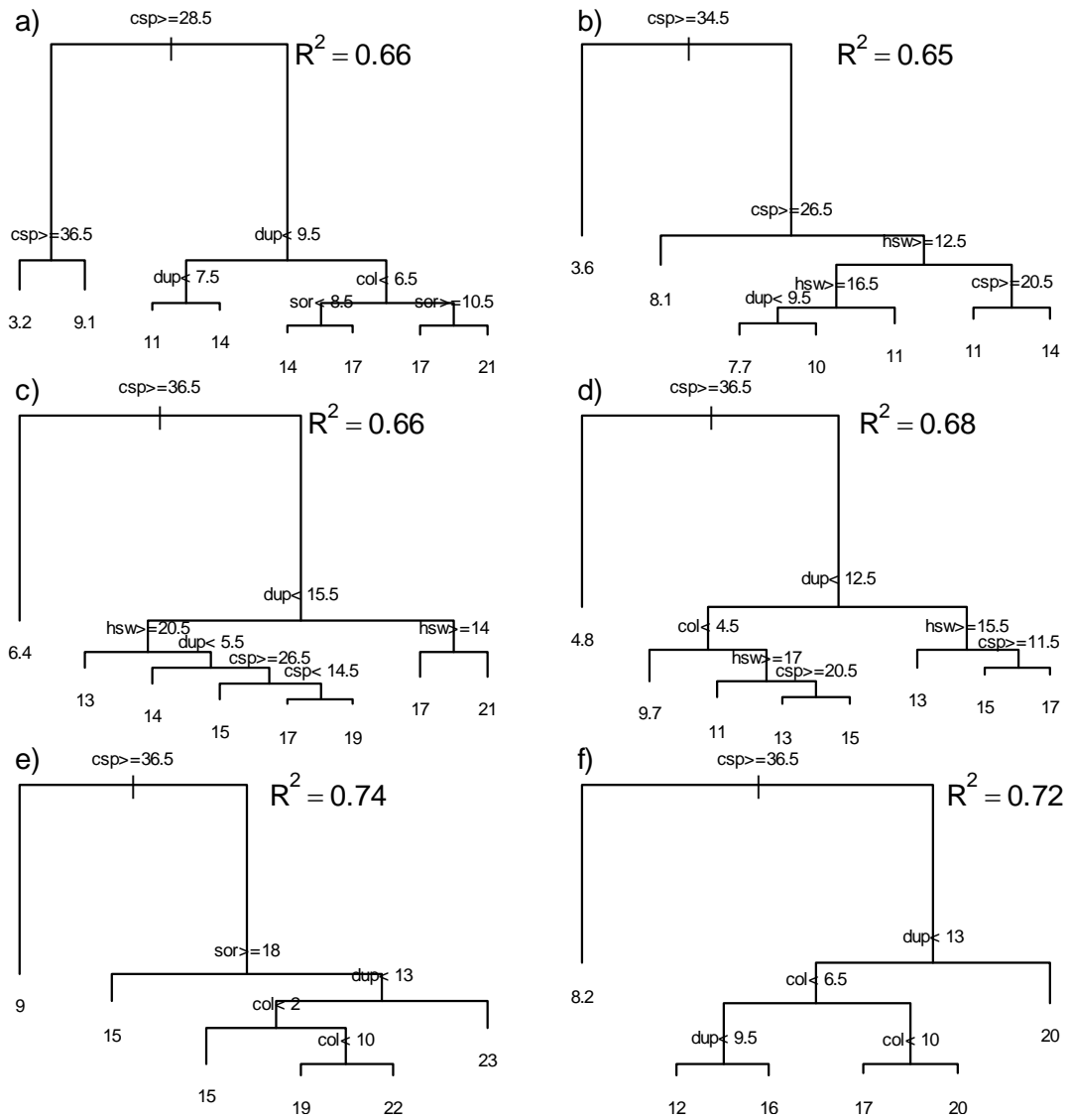


Figure S4. Regression trees relating the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas with the number of evolutionary events in simulated tanglegrams of Set50, ultrametric trees, applying geodesic distances as global-fit method. Expressions on nodes indicate the split criterion that shows the threshold value of the evolutionary event that was used to create the split. Branches to the left include cases that fulfill the criterion. Numbers at terminals represent the average VMR of the group. Parameter combinations were as follows: (a) $n = 5$, $p = 1\%$; (b) $n = 5$, $p = 5\%$; (c) $n = 10$, $p = 1\%$; (d) $n = 10$, $p = 5\%$; (e) $n = 20$, $p = 1\%$; (f) $n = 20$, $p = 5\%$. Event abbreviations: csp = cospeciation, sor = sorting, dup = duplication, hsw = complete host-switching, col = colonization of new host without speciation.

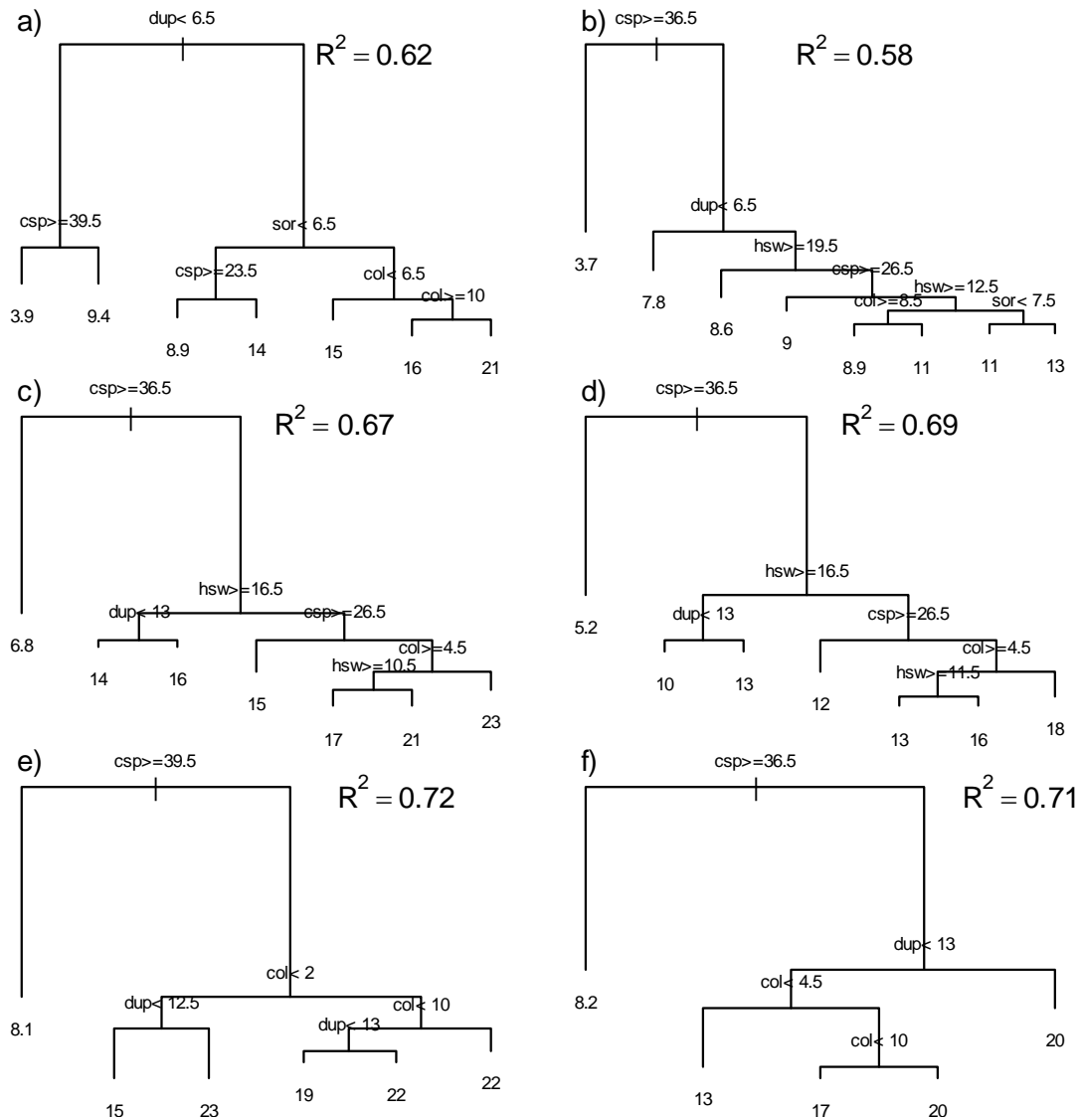


Figure S5. Regression trees relating the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas with the number of evolutionary events in simulated tanglegrams of Set50, ultrametric trees, applying PACo as global-fit method. Expressions on nodes indicate the split criterion that shows the threshold value of the evolutionary event that was used to create the split. Branches to the left include cases that fulfill the criterion. Numbers at terminals represent the average VMR of the group. Parameter combinations were as follows: (a) $n = 5$, $p = 1\%$; (b) $n = 5$, $p = 5\%$; (c) $n = 10$, $p = 1\%$; (d) $n = 10$, $p = 5\%$; (e) $n = 20$, $p = 1\%$; (f) $n = 20$, $p = 5\%$. Event abbreviations: csp = cospeciation, sor = sorting, dup = duplication, hsw = complete host-switching, col = colonization of new host without speciation.

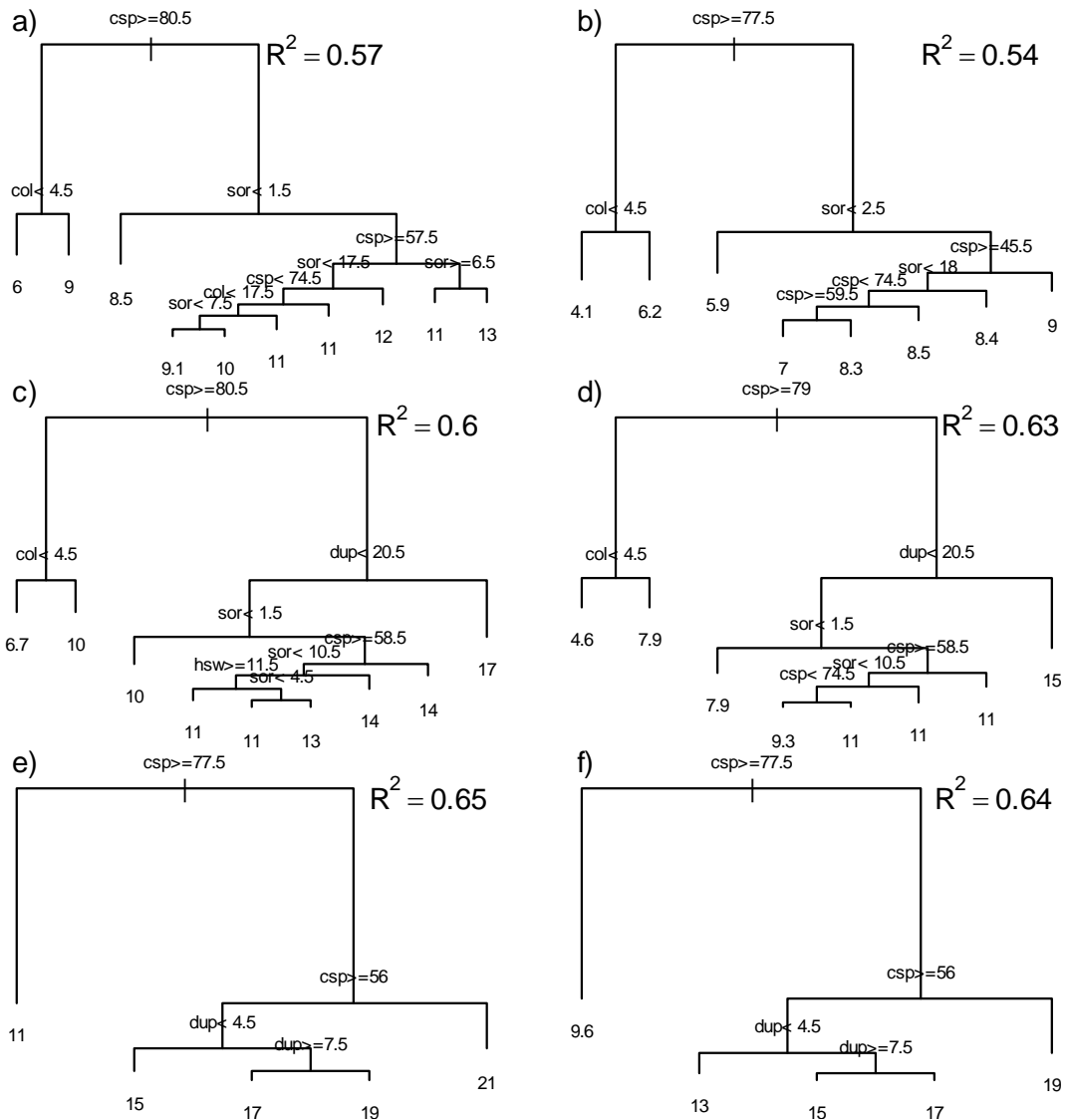


Figure S6. Regression trees relating the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas with the number of evolutionary events in simulated tanglegrams of Set100, additive trees, applying geodesic distances as global-fit method. Expressions on nodes indicate the split criterion that shows the threshold value of the evolutionary event that was used to create the split. Branches to the left include cases that fulfill the criterion. Numbers at terminals represent the average VMR of the group. Parameter combinations were as follows: (a) $n = 10$, $p = 1\%$; (b) $n = 10$, $p = 5\%$; (c) $n = 20$, $p = 1\%$; (d) $n = 20$, $p = 5\%$; (e) $n = 40$, $p = 1\%$; (f) $n = 40$, $p = 5\%$. Event abbreviations: csp = cospeciation, sor = sorting, dup = duplication, hsw = complete host-switching, col = colonization of new host without speciation.

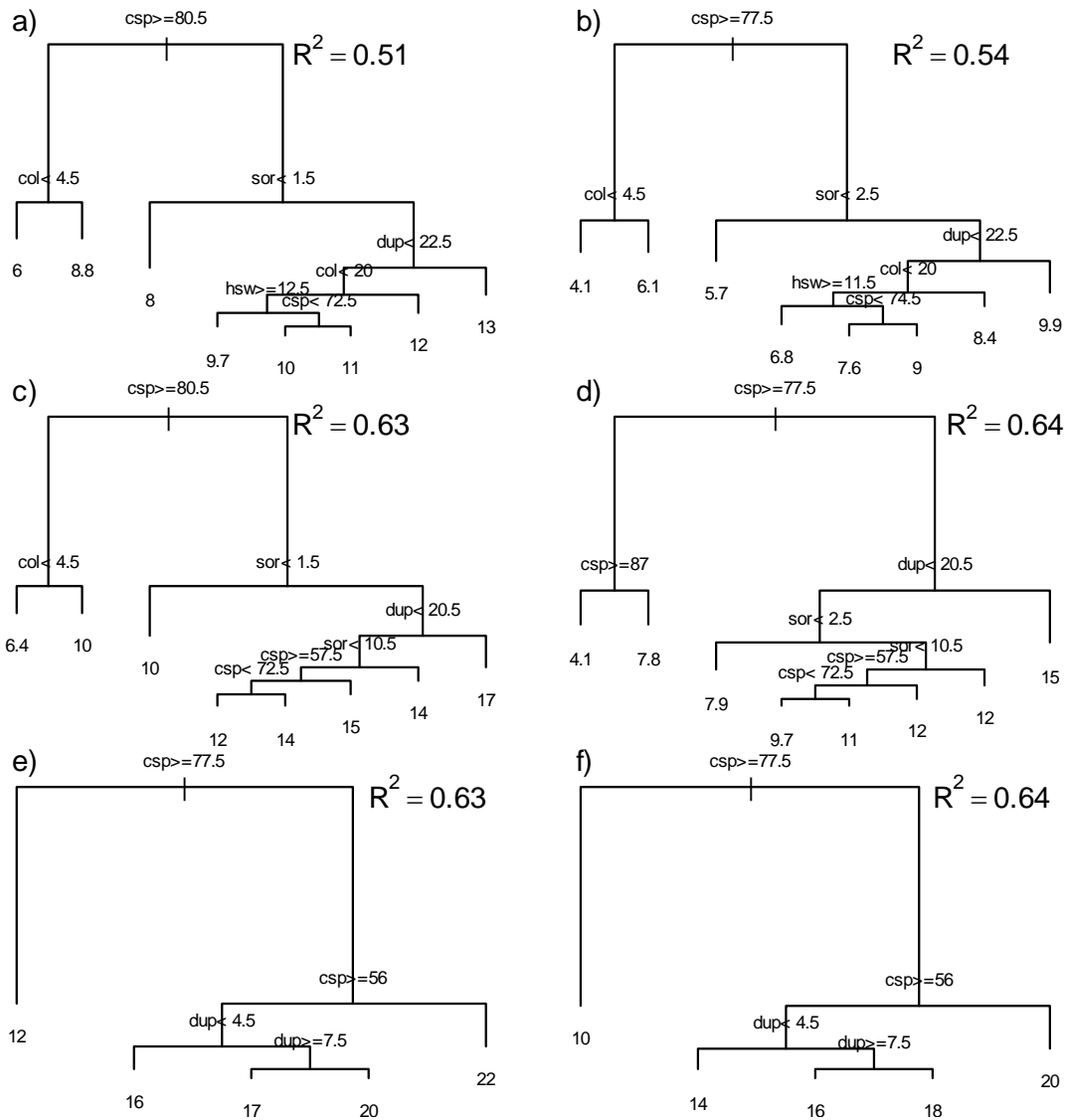


Figure S7. Regression trees relating the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas with the number of evolutionary events in simulated tanglegrams of Set100, additive trees, applying PACo as global-fit method. Expressions on nodes indicate the split criterion that shows the threshold value of the evolutionary event that was used to create the split. Branches to the left include cases that fulfill the criterion. Numbers at terminals represent the average VMR of the group. Parameter combinations were as follows: (a) $n = 10$, $p = 1\%$; (b) $n = 10$, $p = 5\%$; (c) $n = 20$, $p = 1\%$; (d) $n = 20$, $p = 5\%$; (e) $n = 40$, $p = 1\%$; (f) $n = 40$, $p = 5\%$. Event abbreviations: csp = cospeciation, sor = sorting, dup = duplication, hsw = complete host-switching, col = colonization of new host without speciation.

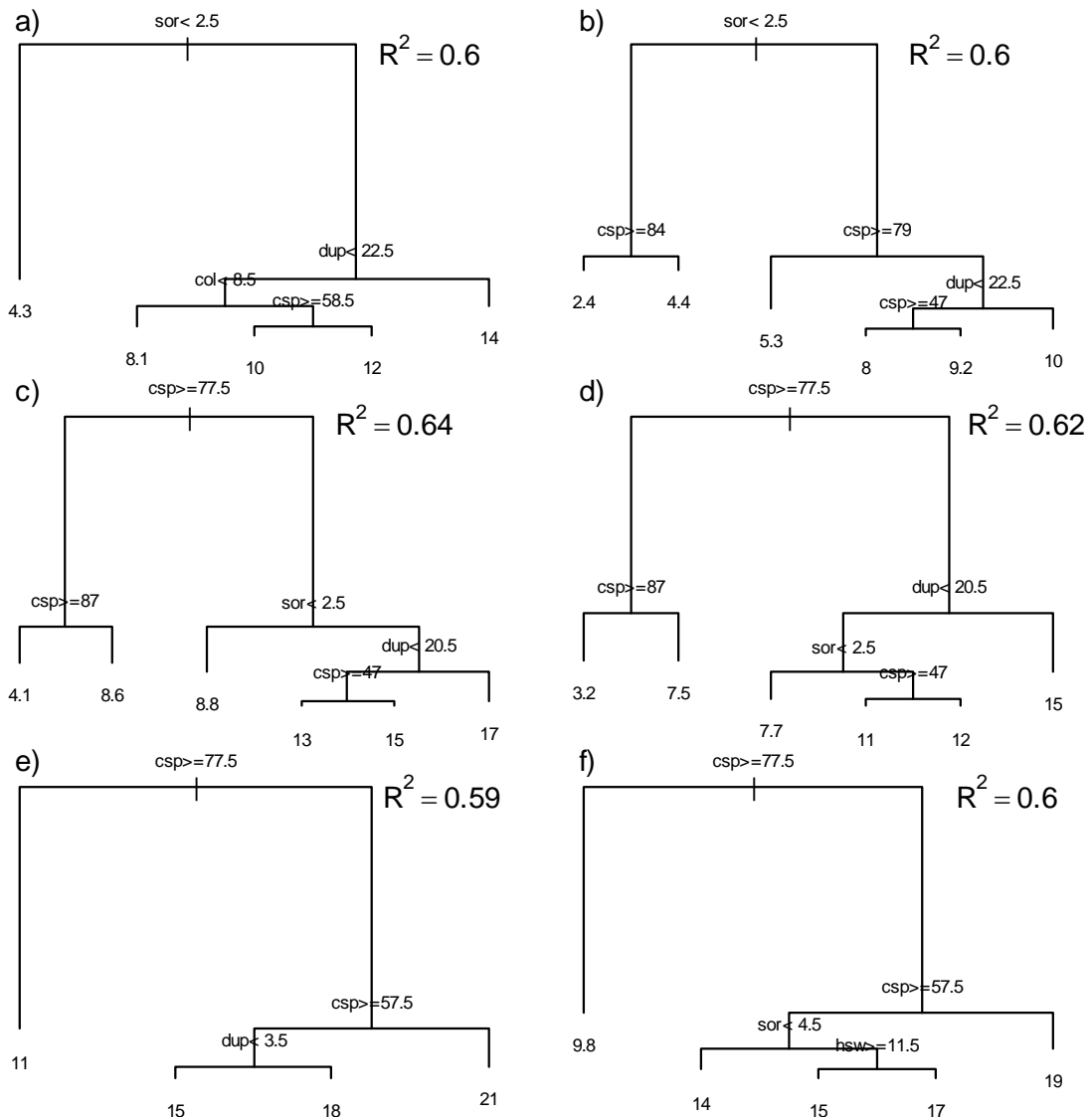


Figure S8. Regression trees relating the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas with the number of evolutionary events in simulated tanglegrams of Set100, ultrametric trees, applying geodesic distances as global-fit method. Expressions on nodes indicate the split criterion that shows the threshold value of the evolutionary event that was used to create the split. Branches to the left include cases that fulfill the criterion. Numbers at terminals represent the average VMR of the group. Parameter combinations were as follows: (a) $n = 10$, $p = 1\%$; (b) $n = 10$, $p = 5\%$; (c) $n = 20$, $p = 1\%$; (d) $n = 20$, $p = 5\%$; (e) $n = 40$, $p = 1\%$; (f) $n = 40$, $p = 5\%$. Event abbreviations: csp = cospeciation, sor = sorting, dup = duplication, hsw = complete host-switching, col = colonization of new host without speciation.

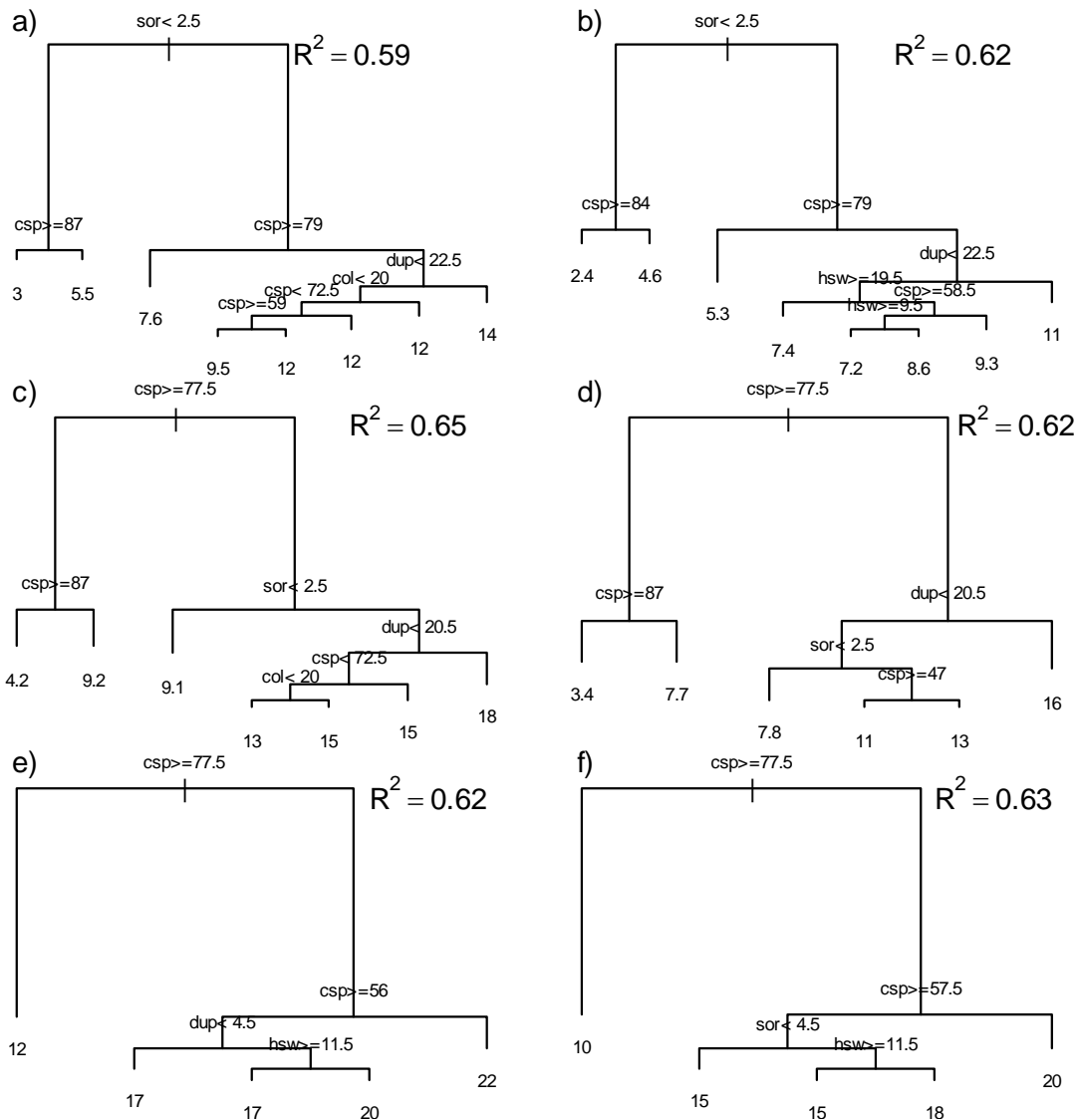


Figure S9. Regression trees relating the variance to mean ratio (VMR) of the frequency distribution of host-symbiont associations produced by Random TaPas with the number of evolutionary events in simulated tanglegrams of Set100, ultrametric trees, applying PACo as global-fit method. Expressions on nodes indicate the split criterion that shows the threshold value of the evolutionary event that was used to create the split. Branches to the left include cases that fulfill the criterion. Numbers at terminals represent the average VMR of the group. Parameter combinations were as follows: (a) $n = 10$, $p = 1\%$; (b) $n = 10$, $p = 5\%$; (c) $n = 20$, $p = 1\%$; (d) $n = 20$, $p = 5\%$; (e) $n = 40$, $p = 1\%$; (f) $n = 40$, $p = 5\%$. Event abbreviations: csp = cospeciation, sor = sorting, dup = duplication, hsw = complete host-switching, col = colonization of new host without speciation.