### Supplemental Information Text 1: Panel population structure

The genetic structure of populations can inform our understanding of evolutionary history, but creates problems when attempting to determine whether, and which, functional relationships exist between allelic variants and phenotypic variation because history can be confounded with function. Daphnia of the group studied here have been considered by ecologists to consist of two parental species, D. pulex and D. pulicaria, that exist in different habitats but interbreed to produce hybrids that exist in a third habitat. Field surveys have historically found the hybrids to be obligate parthenogens, thus limiting gene flow between the parental species. This history sets up the expectation of strong genetic structure in the panel because we collected clones from all three habitats.

After using GATK to genotype the panel, we converted the matrix to 0/1/2 (representing the number of reference alleles possessed by each clone) for all loci with two alleles (n = 5722). We next used principal components analysis (PCA) implemented in the FactoMineR package (Husson et al. 2013) to decompose the normalized (following Price et al. 2006) genotype matrix. Contrary to expectations, there was very little structure in the panel (Figure SI\_T1\_1): the first three components each account for just over 4% of total variance. For comparison, we generated a 32 x 5000 matrix of random values on the interval (0, 2), then used the same procedure to decompose the random matrix (Figure SI\_T1\_2). The first three components of the random decomposition each account for just over 3% each. That is, the genotype data has just over 1% more structure than completely random data in the first three axes; the sum of the absolute values of observed minus random marginal variances across all 32 axes (or total deviation from random) was 8.9%.

We next considered the biplot of PCs 1 and 2 (Figure SI\_T1\_3). Three general vectors are visible, but two factors reinforce the interpretation of low panel structure. First, the clones along these three general vectors do not correspond to three Daphnia ecotypes: the upper-right quadrat includes pulex (C30), pulicaria (C2), and hybrids (C11); the upper-left, pulex (C21 and C22) and hybrid (C33); and lower-left pulex (C27, C29) and hybrids (e.g., C3). Second, these two axes are a very small portion of the total variance (2.2% departure from random, or 9.3% total).

We also analyzed the genotype matrix using STRUCTURE (Pritchard et al. 2000) to determine if model-based analysis provided a different result, given the expectation of population structure. We first removed markers < 200bp distant to reduce LD, then used 10000 burnin replicates and 20000 MCMC reps thereafter to estimate parameters; repeated runs at different K values showed these estimates to be stable. With analyses of 1 ≤ K ≤ 12 we found little population structure: the assignment of clones to origin from the K populations were evenly split as 1/K.

Because low minor allele frequency (MAF) can lead to ascertainment bias, we first reduced the matrix to loci with a minor allele frequency > 0.2, and two populations were recovered (posterior Pr(D|K=2) = 0.96). However, these assignments did not match the clone ecotypes: pulicaria, pulex, and hybrids clustered into one population and hybrids and pulex in another. Next, rather than reducing the matrix by enforcing a MAF cutoff, we set lambda, the Dirichlet distribution parameter for random allele draws, to 0.5 following the recommendations of Pritchard et al. (2012) and estimated alpha at 1 ≤ K ≤ 5. Under these parameters the posterior Pr(D|K=4) = 0.98, but the population assignments remained ~1/K (Table 1), indicating that although there are four origin populations there is effectively very little structure to the panel. Finally, we performed an analysis in which we estimated lambda given K=1, then set lambda at the estimate (9.7) for 1 ≤ K ≤ 5. Again, there was very little population structure detected, with K=1 possessing the highest –log(likelihood) (= -123155) followed by K=2 (= -123161). Although the K=1 solution is not overwhelmingly supported compared to K=2, as with the high-MAF analysis, the K=2 assignments did not match ecotypes.

Rather than observing population structure along a pulex-pulicaria axis with hybrids in-between, as expected from classical ecological work, we found only slight structure that did not correspond with the three ecotypes. We discuss in the main text (Discussion) other recent breeding and molecular work that supports the presence of gene flow among Daphnia pulex group ecotypes. In addition, we discuss the caveats of these results: this is a relatively small panel from a small geographic area, and extending the results to the Daphnia pulex group in general requires caution (i.e., should be put on hold until larger sample size is acquired).

This suggests that, at least with the clones from this panel, there has been recent gene flow among the three ecotypes. Furthermore, the lack of structure enables genome-wide association analyses with reduced concern for structure-function confounding.



**Figure SI Text 1.1. Results of genotype matrix decomposition for the *Daphnia* panel.** The blue line and left y-axis give the cumulative proportion of variance explained, while the red points and right y-axis provide the marginal variance of each eigenvector.



**Figure SI Test 1.2. Results of genotype matrix decomposition for a random matrix of approximately the same size (32x5000) as the *Daphnia* genotype matrix.** The blue line and left y-axis give the cumulative proportion of variance explained, while the red points and right y-axis provide the marginal variance of each eigenvector.



**Figure SI Text 1.3. Biplot of the first two components of PCA decomposition of the Daphnia genotype matrix.**



Table SI Text 1.1. Summary of assignments from STRUCTURE analysis for K=4, lambda=0.5.