Dr Richard Kliman has offered a thoughtful answer to the question entered by Jordi Planas, about understanding population genetics analysis. Jordi is interested in understanding more about issues discussed in a recent review article by Fraser et al,. 2009.

Understanding a neutral model of molecular evolution

Under a neutral model of molecular evolution, the expected level of nucleotide diversity (variation within a group) is $2fN_e\mu$, where:

N_e is effective population size
μ is the "neutral" mutation rate, and
f is a correction factor for ploidy

A word about f--
f is 2 for autosomal loci in a diploid species, and somewhat smaller ($\sim$1.5) for X-linked loci in a species with XY sex determination and about equal numbers of breeding males and females, etc.

Breaking this down a bit further, fN_e is the number of generations back to the common ancestor of two randomly chosen gene copies. As N_e increases, randomly chosen individuals are expected to be more distantly related. If two individuals trace the ancestry of the given gene to an ancestor fN_e generations in the past, then the extant gene copies have been separated by 2fN generations (fN_e along each lineage). Multiplying this by the mutation rate μ (mutations/site/generation), we get the expected nucleotide diversity.

A corollary to this model (related to coalescent theory) is that the common ancestor of all gene copies in a population exists, on average, $2fN_e$ generations in the past. Thus, in an isolated population, all observed nucleotide diversity has arisen, on average, in the past $2fN_e$ generations. The variance associated with the expected "time depth" of the population's genealogy is quite large, but it's unlikely that the common ancestor would be found much more than $4fN_e$ generations in the past.

How does this relate back to the review by Fraser et al.? First, they note that usually N_e is much lower than census population size (N_c). There are a number of possible explanations, but the authors focus on two quite reasonable explanations: selective sweeps and metapopulation structure, both of which involve bottlenecks of a sort. In the case of selective sweeps, all copies of a gene descend from a single ancestor (i.e., the copy in which the highly beneficial mutation arose). Thus, while the population may never have been small, most individuals living at the time the mutation arose fail to leave descendants, so there is a *virtual* bottleneck.

In the case of a metapopulation, suitable habitats have independent probabilities of local extinction and local repopulation. Some subpopulations are able to act as *sources* of immigrants (by dispersal) into an empty habitat. If a subpopulation goes extinct, its habitat may be repopulated by founders dispersing from another habitat. Assuming these

founders are small in number, then there is a *true* bottleneck. This alone would not dramatically reduce genetic variation throughout the metapopulation, since only the new subpopulation suffers the bottleneck. However, assuming *all* subpopulations are at risk of extinction, and since variation is low within subpopulations, the continual process of replacing extinct subpopulations with individuals dispersing *from* populations with low variation will lead to low variation throughout the metapopulation.

In either case (selective sweeps or metapopulation structure), N_e is reduced, due to the transient small number of individuals who contribute genetically to subsequent generations. And if N_e is low, then nucleotide diversity should be low, since individuals will, on average, be more closely related to each than when N_e is high.
I hope this helps clarify the confusion.

Regards,
Dr. Richard Kliman