File S1

Supporting Materials and Methods

Identifying shared and unique incompatibilities with PhyloQTL

PhyloQTL jointly analyzes multiple crosses on a common genetic map to place QTL on a phylogenetic tree. This analysis relies on the assumption that each trait of interest is under the control of a single diallelic QTL which has the same effect in all crosses. The genotype of each individual is first recoded to follow a dichotomous partition. For example – in the case of the three *M. musculus* subspecies *M. musculus musculus* (subsequently referred to as *musculus*), *M. m. castaneus* (*castaneus*), and *M. m. domesticus* (*domesticus*) – there are three potential genotypic partitions: *musculus***|***castaneus*, *domesticus*; *castaneus***|***musculus*, *domesticus*; *domesticus***|***castaneus*, *musculus*. The subspecies in these partitions are paired (delineated with vertical bar) by their shared genotype. Standard interval mapping is applied using each genotypic partition, with an indicator variable designating the cross from which the individual derives as an additive covariate. The measure of support for each partition is calculated using an approximate Bayes procedure and the partition with the maximum LOD score is inferred to be the true partition of that QTL. To identify significant QTL, we used 10,000 permutations for each trait and identified peaks that reached the 5% significance level. The permutations were stratified by cross and the maximum LOD score across both the genome and the partition were taken from each permutation in building the significance thresholds. Using this method, QTL from our crosses are identified as shared when the inferred partition matches the shape of the subspecies tree.

Constructing gene trees from whole genome sequences

In order to construct the gene trees used to estimate the subspecies tree, we began with the consensus sequences of CAST/EiJ (*castaneus*), WSB/EiJ (*domesticus*), and PWK/PhJ (*musculus*) from Keane et al. (2011) mapped to an alignment of the mouse and rat genomes. In order to break these sequences into loci that correspond to a coherent localized evolutionary history, Keane et al. (2011) used the principle of minimum description length (Ané and Sanderson 2005). This technique partitions the genome into units of consistent topological history based on the compressibility of the sequence information. We took these loci and analyzed them with MrBayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), running four Markov chains for 2,000,000 generations in two simultaneous runs and discarding the initial 25% of trees as burn-in. Prior distributions for topology and branch lengths were left at their default settings. While the minimum description length principle creates topologically consistent loci based on sequence entropy, the analysis by MrBayes returns a distribution of gene trees for each locus which includes trees of varying branch lengths and topology. From the posterior distribution for each locus, the 50% majority rule consensus tree was taken as a representative phylogeny, yielding 43,255 gene trees. This collection of consensus gene trees was the input for summary methods designed to estimate a species tree from this type of data (main text).

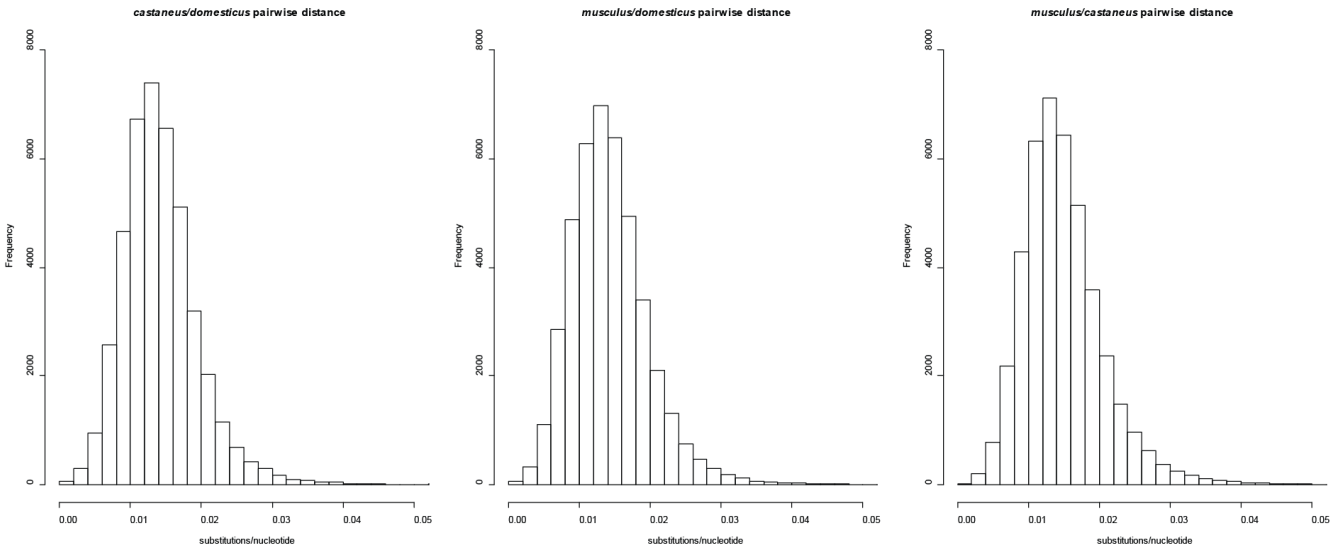

Figure S1 Distribution of pairwise branch lengths for each subspecies pair across loci. Each count is a pairwise distance from one of the 43,255 gene trees, each a consensus tree from a posterior distribution, from the MrBayes analysis. The similarity between these distributions highlights the near simultaneity of divergence between house mouse subspecies and the limited levels of gene flow.

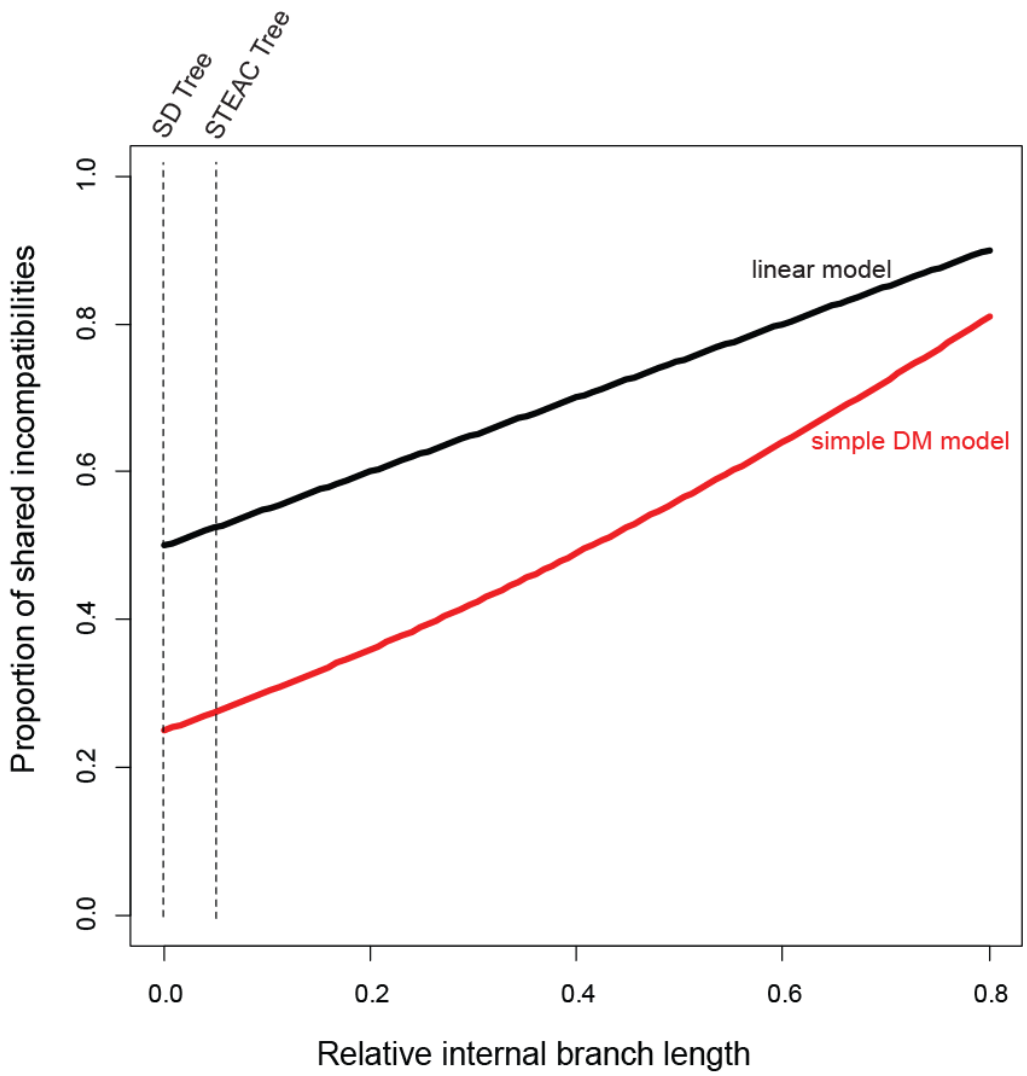

Figure S2 Proportion of shared incompatibilities as a function of relative internal branch length in a three-species tree under the linear and simple DM models. The proportion of shared incompatibilities is between the expected numbers of shared incompatibilities from the two most divergent lineages relative to the total number of incompatibilities between those two lineages. The relative internal branch length is the time between the two divergences relative to the total time since the first divergence in the tree. The estimated proportion of shared ancestral history in the house mouse complex using the SD and STEAC methods are shown as vertical lines.