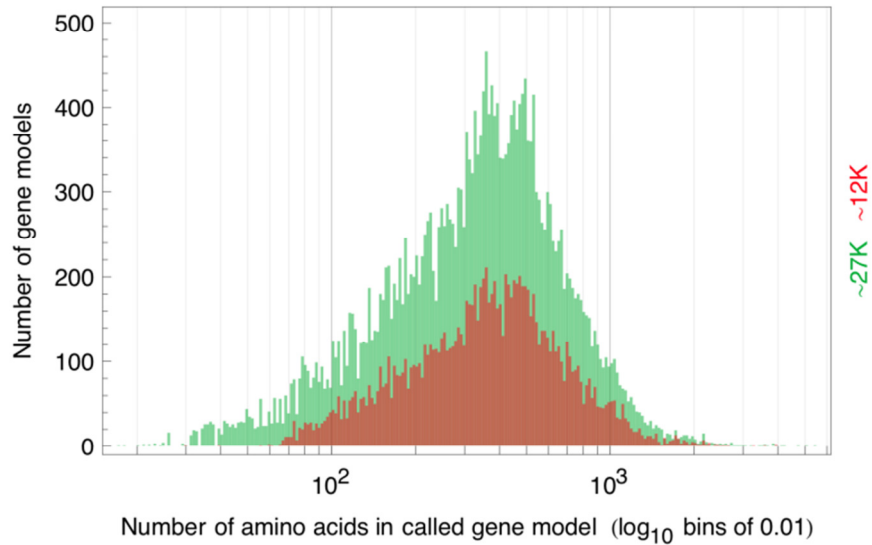


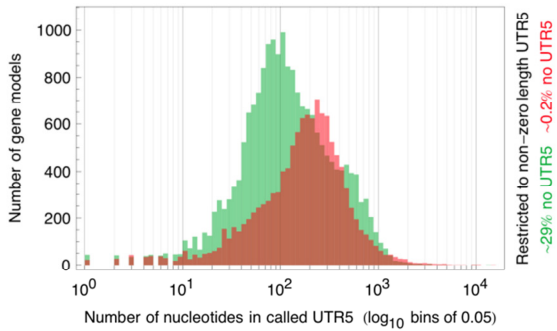
Amino acid usage: (left→right is most→least common)

*Arabidopsis* TAIR10 PCG models (representative per locus): L S E V G K A R D I T P N F Q Y M H C W  
 | | | | | | | | | | | | | | | | | | | |  
*Q. lobata* AUGUSTUS draft final models (complete only): L S A E G V K D I R T P N F Q Y M H C W

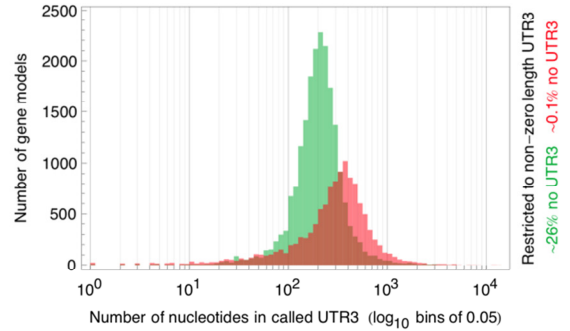
*Arabidopsis* TAIR10 protein coding gene models (TAIR10's representative model per locus) *Q. lobata* AUGUSTUS draft final models (complete only)



*Arabidopsis* TAIR10 protein coding gene models (TAIR10's representative model per locus) *Q. lobata* AUGUSTUS draft final models (complete only)



*Arabidopsis* TAIR10 protein coding gene models (TAIR10's representative model per locus) *Q. lobata* AUGUSTUS draft final models (complete only)



### Additional file 6: Comparisons of oak and *Arabidopsis* amino acid usage and length of gene models and UTRs

(a) Relative frequency of amino acid usage in *Arabidopsis* TAIR10 gene models compared to complete *Quercus* gene models ordered left to right from most to least common; and distributions of the length of (b) gene models in amino acids, (c) 5'-UTRs in nucleotides, and (d) 3'-UTRs in nucleotides for *Arabidopsis* TAIR10 (green) versus complete *Quercus* models (red).