



**Supplementary information, Figure S4** Assessment of gene prediction in Cq\_real\_v1.0.

(A) Cumulative frequency of the probabilistic confidence score of predicted genes in Cq\_real\_v1.0 given by GLEAN. (B) Pie charts showing the percentage of gene sets that have conserved protein domains present or absent in the protein sequence. Conserved domains were identified by blast against the Pfam database using a E-value cutoff of 1E-4. (C) Distribution of RPKM values of all the annotated genes. The mRNA-seq data from 5 different types of quinoa tissue were used.