Each of the six Stiff Stalk genomes, including B73, were annotated for gene models using an identical pipeline with inbred-specific transcript evidence thereby eliminating false annotations from transcripts from other inbreds. First, each genome assembly was masked with RepeatMasker (v4.1.0) (<http://www.repeatmasker.org/>) using the curated maize repeat library maizeTE02052020 (https://github.com/oushujun/MTEC). RNA-seq libraries from five core tissues were cleaned using Cutadapt (v2.9) (Martin 2011), aligned to their respective genome assembly using HISAT2 (v2.2.0) (Kim et al. 2019), and assembled using Stringtie (v2.1.1) (Kovaka et al. 2019). Augustus (v3.3.3) (Stanke et al. 2008) was used to generate gene predictions on the masked assemblies using the maize5 training parameter set and the RNA-Seq alignments as hints. Gene predictions were refined using PASA2 (v2.4.1) (Haas et al. 2005) in two rounds of annotation comparison using the RNA-Seq transcript assemblies as evidence to generate the working model gene set. To identify high-confidence gene models, the working gene model set was searched against the PFAM database (v32) (Finn et al. 2016) with hmmscan (HMMER, v3.2.1) (Mistry et al. 2013) to identify gene models encoding a Pfam domain as described previously (Pham et al. 2020). Gene expression abundances for the working gene models (transcripts per million; (TPM)) were generated for each RNA-seq library using Kallisto (v0.46.0) (Bray et al. 2016). High confidence gene models were identified if they had a TPM value > 0 in at least one RNA-seq library and/or had a PFAM domain match. Partial gene models or gene models with matches to transposable element-related PFAM domains were excluded from the high-confidence model set. Functional annotation was assigned to the working gene model set using search results from the predicted proteins against the Arabidopsis proteome (TAIR10; Arabidopsis.org), the PFAM database (v32) (Finn et al. 2016) and Swiss-Prot plant proteins (release 2015\_08). Results were processed in the same order (TAIR, PFAM, Swiss-Prot) and the function of the first informative hit was transitively assigned to the gene model. A full description of methods can be found in Bornowski and Michel et al. “*Genomic variation within the maize Stiff Stalk heterotic germplasm pool*” (2021) The Plant Genome, in press.

Bray, N. L., H. Pimentel, P. Melsted, and L. Pachter. 2016. “Near-Optimal Probabilistic RNA-Seq Quantification.” *Nature Biotechnology* 34 (5): 525–27.

Finn, Robert D., Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Jaina Mistry, Alex L. Mitchell, Simon C. Potter, et al. 2016. “The Pfam Protein Families Database: Towards a More Sustainable Future.” *Nucleic Acids Research* 44 (D1): D279-85.

Haas, Brian J., Jennifer R. Wortman, Catherine M. Ronning, Linda I. Hannick, Roger K. Smith Jr, Rama Maiti, Agnes P. Chan, et al. 2005. “Complete Reannotation of the Arabidopsis Genome: Methods, Tools, Protocols and the Final Release.” *BMC Biology* 3 (March): 7.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. “Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype.” *Nature Biotechnology* 37 (8): 907–15.

Kovaka, S., A. V. Zimin, G. M. Pertea, R. Razaghi, S. L. Salzberg, and M. Pertea. 2019. “Transcriptome Assembly from Long-Read RNA-Seq Alignments with StringTie2.” *Genome Biology* 20 (1): 278.

Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” *EMBnet.Journal* 17 (1): 10–12.

Mistry, Jaina, Robert D. Finn, Sean R. Eddy, Alex Bateman, and Marco Punta. 2013. “Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions.” *Nucleic Acids Research* 41 (12): e121.

Pham, Gina M., John P. Hamilton, Joshua C. Wood, Joseph T. Burke, Hainan Zhao, Brieanne Vaillancourt, Shujun Ou, Jiming Jiang, and C. Robin Buell. 2020. “Construction of a Chromosome-Scale Long-Read Reference Genome Assembly for Potato.” *GigaScience* 9 (9): giaa100.

Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler. 2008. “Using Native and Syntenically Mapped CDNA Alignments to Improve de Novo Gene Finding.” *Bioinformatics*  24 (5): 637–44.