

Figure S2. Confirmation of SNPs in W22 3'-anchored sequence reads versus B73 genomic sequence and in homopolymer runs

A TC298173: Histone 3 (*O. sativa*)

CAT1	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT2	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT3	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT4	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT5	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT6	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT7	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT8	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT9	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT10	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT11	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT12	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
CAT13	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
ACT1	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
ACT2	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
ACT3	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
ACT4	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
consensus		
W22	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG
B73	CGGCCCCAGTGGCGCGGAGGGAT- GGAGGAGGCAATTTGTAACCTTATTATGCTGTTTGTTCGTTTTAGACGGCGTGGGAAGAA-CTAGTGTAAAAACGGCTTTA	AAGTCTATTTTGTG

* * * *

B NCBI UniGene 2567165: heavy-metal-associated (*O. sativa*)

CAT1	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
CAT2	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
CAT3	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
CAT4	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
CAT5	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
CAT6	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
CAT7	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
ACT1	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
consensus	
W22	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA
B73	CGGTTTCATGT-AGATGTAGTGTGCGTTTTGATTTTGTGAAAAAATAAA-CCAGTTTGTAAACGGTAGACCACCTTGTGCAAACATGATCAGAAAA

*

C NCBI UniGene 514900: At1g74950 (*A. thaliana*)

CAT1	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- CT
CAT2	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- CT
CAT3	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- C
CAT4	CGG- AGGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- C
CAT5	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- CT
CAT6	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- C-CT
ACT1	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- - CTCT
ACT2	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- CTCT
ACT3	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- CTCT
consensus	
W22	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- CTCT
B73	CGG- GGCGATTTGATTGGCTGGGTTCTGGCTTC- TTTGGGCT- GCTGGATTGGGGTTGGTGTGGAGATCGGAGACGGAGGATTGTAGAAACGATTCTTGT- CTCT

* * * ^

Figure S2. A, A CAP3 assembly of 17 *Histone 3* reads from the two, multiplexed, W22, 454 sub-libraries (ACT, CAT keys). The 5 SNPs (highlighted) that distinguish the W22 allele were confirmed by independent W22 ESTs. B, A SNP embedded in a homopolymer tract of a putative *HMA* mRNA is detected. An adjacent 8 bp run of A's yields the correct consensus. C, A unique transcript annotated as "unknown function" includes 4 putative SNPs: 3 (* highlighted) have independent cDNA support, the 4th (^ highlighted) is most likely an error due to under-estimation of the 8mer run of T. Trimmed reads from combined sub-libraries were assembled using CAP3 (p=90). Polymorphisms were confirmed by BLASTN of the Industry Unigene cDNAs and public ESTs.