3,790 Arabidopsis single copy genes - ATH_SCG

predicted Arabidopsis open reading frames (ORFs)

BLAST-P against themselves cutoff: 1e-10

selection for sequences having single strong hit against self only with expectation cutoff value 1e-20 or better

412,827 Rosaceae ESTs NCBI GenBank

BLAST-X ESTs versus ATH_SCG 1e-15 cutoff value

30,801 RosCOS EST candidates

selection of regions corresponding to Arabidopsis ORF

CAP3 assembly 80 nt overlap 90% identity

7,573 RosCOS unigene candidates

3,820 contigs

3,753 singletons

Selected set of 1,039 RosCOS contigs corresponding to 901 ATH_SCG

7,247 RosCOS unigenes

3,818 contigs

3,429 singletons correspond to 2,324 ATH_SCG

QC BLAST-X versus ATH_SCG expect cutoff 1e-10