SNP filtering

Peptide prediction (FrameDP)

145,827 ESTs (Sanger chromatograms)

1,948,579 ESTs (454 sff files)

unique peptides

Clustering (BLASTClust)

Domain search (HMMER)

In silico SNP detection

222,671 unigene elements (OakContigV1:69,154 contigs and 153,517 singletons)

Peptide prediction (FrameDP)

Clustering (BLASTClust)

40,944 Sanger unigene elements

40,944 Sanger unigene elements

116,826 MIRA unigene elements

40,944 Sanger unigene elements

Hybrid assembly (MIRA)

Hybrid assembly (MIRA)

Hybrid assembly (CAP3)

578,192 ESTs

Oak Contig Browser

SNP filtering

Coverage calculation

Search for candidate genes

20,826 candidate SNPs for genetic analysis

Quercus petraea and Q. robur ESTs

SURF database

EMBL 125,886 ESTs

Quality filtering (QualityTrimer)

125,925 ESTs

trace2dbEST

PartiGene

Clustering (CLOBB)

Assembly (phrap)

40,944 Sanger unigene elements

Homologous sequence search (Reciprocal Blast)

40,944 Sanger unigene elements

116,826 MIRA unigene elements

189,268 debris ESTs

OakContigV1

Peptide prediction (FrameDP)

Mapping (MIRA)

CDS detection

Homologous sequence search (Reciprocal Blast)

Homologous sequence search (Reciprocal Blast)

Peptide prediction (FrameDP)

Hybrid assembly (CAP3)

15,835 tentative contigs

Selection of 224 tentative contigs with full length cDNA

1,798 frame corrected nucleotide sequences

468 predicted peptides

125,886 ESTs

NCBI (SRA)

NG6

SURF database

125,886 ESTs

Base call (phred)

Library specific vector/adaptor detection (cross_match)

Low complexity sequence detection (RepeatMasker)

Un/Re contamination detection (cross_match and Blast)

Other contaminants detection (cross_match)

Chloroplast sequence detection (cross_match)

‘Doubtful’, ‘poor’ and ‘not valid’ sequence detection

Base call (phred)

Sequence extraction (ufile)

Adaptor/primer detection (cross_match)

E. coli contamination detection (cross_match)

Chimera detection

II

125,925 ESTs

PartiGene

Clustering (CLOBB)

Assembly (phrap)

40,944 Sanger unigene elements

Homologous sequence search (Reciprocal Blast)

125,925 ESTs

1,578,192 ESTs

125,886 ESTs

NG6

15,835 tentative contigs

Selection of 224 tentative contigs with full length cDNA

125,886 ESTs

NCBI (SRA)

NG6

125,925 ESTs

MG6

125,886 ESTs

Quality filtering (QualityTrimer)

125,925 ESTs

MG6

125,925 ESTs

MG6

125,925 ESTs