A HIGH QUALITY DRAFT OF THE GENOME SEQUENCE OF VITIS VINIFERA L.

THE FRENCH-ITALIAN PUBLIC CONSORTIUM FOR THE SEQUENCING OF THE GRAPEVINE NUCLEAR GENOME

The grapevine Vitis vinifera L. is in economic terms the principal fruit crop in the world. Its haploid genome is estimated to be about 500 Mb, organized in 19 chromosomes. The grapevine is the forth plant whose sequence has been made public, after A. thaliana, rice and poplar, and it is the second ligneous and the first fruit plant with a public sequence. Here we present an academic consortium project that completed a 12X shotgun sequence of a quasi-homozygous PN40024 lineage. Sequencing was performed using a Whole Genome Shotgun strategy: all data were generated by paired-end sequencing of cloned inserts from plasmid and fosmid libraries of different insert size, using Sanger technology on ABI3730xl sequencers. Using a 8.4X coverage an intermediary assembly of 498 Mb composed of 3830 scaffolds. Of the 317.364 available V. vinifera ESTs 96% were mapped, together with 99.5% of the 38.586 cDNA clones that were produced and sequenced by our Consortium, proving the high genomic coverage of this assembly. The longest scaffold and contig are respectively 12.7 Mb and 557 kb. Half of the assembly is represented by scaffolds longer than 1.9 Mb or by contigs longer than 64 kb. A large majority of these are anchored on linkage groups. Sequence analysis allowed to estimate to less than 7% the residual heterozygousity of the lineage used. Three different approaches revealed that approximately 41% of the grape genome is of repetitive/transposable elements (TE) origin. All classes and superfamilies of TE are represented in the grape genome. The proteome was determined by an annotation strategy reconciling proteins, cDNA alignments and ab initio predictions. A large number of duplicated and pseudogenes were found.