

alignment segments (32). The BLAST algorithm has been implemented in several programs. BLASTP searches a protein sequence database using a protein-sequence as the query; TBLASTN searches a nucleotide-sequence database conceptually translated in six frames and uses a protein sequence query; BLASTX searches a protein sequence database with a nucleotide sequence query translated in six frames; finally, TBLASTX performs the searches after translating in six frames both the query and the database. In terms of both statistical robustness and speed, BLAST clearly outperforms all other existing methods for database search. Methods producing gapped alignments are often considered to be more sensitive than BLAST (e.g., ref. 33), even though they are much less selective. The lower sensitivity of BLAST may be overcome by additional analysis of the alignments in the twilight zone, using methods for motif search. Combining BLAST with such methods results in a powerful strategy for analyzing subtle relationships between protein sequences and delineating protein superfamilies (34).

A typical BLAST command line is **blastp nr aa.seq -filter seg >aa.br** . Here, **nr** is the nonredundant protein-sequence database, **aa.seq** is an amino acid-sequence file in the FASTA format, the “**-filter seg**” option indicates that low-complexity regions in the query sequence are masked using the SEG program, and **aa.br** is the output file.\*

The BLA program is used in order to perform the BLAST search in a batchwise mode, and to combine it with the search for amino acid patterns that may be conserved in the query sequences and their homologs (35). The typical command line is **bla aa.lib nr prosite.pat blast=blastp -130%**. Here, **aa.lib** is a FASTA library file containing several amino acid sequences, and **prosite.pat** is a file containing the patterns from the PROSITE library (36); **blast=blastp** indicates which program of the BLAST family is to be used; **-130%** indicates that a pattern is considered to be conserved if it is detected in 30% of the alignments produced by BLAST for the given query sequence (by default 50%).

### 2.2.3. Database Searches with Plant Virus Capsid Protein Sequences

Analysis of the amino acid sequences of the plant virus capsid proteins using the BLA program confirms that the sequence conservation in these proteins is very limited (6). Statistically significant sequence similarity was observed mostly with proteins from viruses that belong to the same group, and only in some cases with proteins from related groups of viruses. In a few exceptional cases did sequences of capsid proteins show detectable similarity to other viral

\*Here and below, commands for the UNIX operational system are shown. Programs and algorithms described in this chapter are supported at many local sites and are also accessible by e-mail and World Wide Web, where the commands and their repertoire may be different.