

# Alien: A tool for handling sequence alignments

Dhiego Souto, Lucas Bleicher

*Universidade Federal de Minas Gerais*

Many softwares developed for bioinformatics studies use multiple sequence alignments as an input. To build a multiple alignment is not a trivial task. To generate its global optimum it is required an infeasible time for a computer to calculate in larger cases. So, there are many softwares that build multiple alignments with heuristics and meta-heuristic techniques, which can run in a feasible time although its response is not the optimum. The more accurate is the alignment, the better is the extraction of information from these softwares. Alien implements the following pair alignment algorithms: local alignment algorithm (Smith & Waterman, 1981), also implemented by BLAST; global alignment algorithm (Needleman & Wunsch, 1970). When dealing with pair alignments, is possible to reach optimum in feasible time. Alien also implements a profile hidden Markov model (Eddy, 1994), the Viterbi algorithm (Viterbi, 1967), specific Blosom matrices from any protein multiple sequence alignments and two evaluation methods based on sum of pairs and hmm probabilities. Running the Viterbi algorithm through the profile hmm it is possible to generate a multiple sequence alignment. Alien multiple alignments for short sequences showed good results comparing to Hmmer. Recently ARCA, (Architecture Alignment) was implemented, which is an alignment of proteins architecture. A protein may possess one or more domains, each configuration of these domains is denominated as an architecture of the protein. ARCA searches proteins databases for protein sequences which share the same architecture and align them with the aligners HMMER and MUSCLE. The domains are already aligned by HMMER, and available on Pfam database, but on Pfam these alignments carry only the amino acids related to each domains. So ARCA's job is to gather all the family alignments available on Pfam database, the deleted sequences by subtracting the domains from the entire protein sequence. After the gathering of all the fragments, ARCA aligns the domains and the deleted sequences forming one multiple alignment. The generated architecture alignments showed that many of the deleted sequences between two domains from Pfam can be classified as domains by HMMER, showing new possible architectures.