The current and ongoing revolution in life sciences research has lead to fantastic achievements such as the sequencing of entire genomes of various organisms. Hidden in these vast sequences of nucleic acids are the codes that govern the behavior of the cellular machinery, and clues to how modern organisms have evolved. Sequences are ubiquitous in life sciences applications, and this issue of data engineering highlights some of the work on querying biological sequences that is ongoing in the database community.

The first article by Miranker, Briggs, Mao, Ni and Xu presents a framework, based on SQL extensions, that can be used to pose a variety of complex queries on biological sequences. Current methods for posing such queries are largely procedural, and this work highlights the benefits that declarative querying can bring to the life sciences community.

The tremendous potential of extending the SQL framework to allow querying on biological sequences has also been noticed by commercial relational database vendors. BLAST is the most common tool for querying biological sequences, and the next two articles outline how IBM and Oracle have integrated BLAST querying into their relational frameworks. Eckman and Kaufmann present the approach taken by IBM DB2 Information Integrator, and Stephens, Chen, and Thomas present the approach taken by Oracle Database $10g$.

Since approximate matching of a query sequence is perhaps the most common query in life sciences, it is natural (especially to database researchers) to look for index-based methods for evaluating this operation. Surprisingly, existing tools for evaluating biological sequence matching often don't use indices. The last three articles in this issue present various index-based methods for sequence matching.

The article by Hunt presents a novel index, called the suffix sequoia, which can be used to dramatically reduce the cost of executing a common fully-sensitive sequence search algorithm. The article by Kahveci and Singh describes how a multi-dimensional index can be used for *progressive* sequence searching, providing a better user paradigm than current blocking methods. The final article by Karakoc, Ozsoyoglu, Sahinalp, Tasan, and Zhang makes the case that distance-based indexing methods can be applied for sequence searching, but suffer from the curse of dimensionality, and in the worst case are comparable to brute-force methods. This final article also sketches how certain similarity metrics be approximated by Hamming distance, which are more amenable to indexing.

The mysteries of life are hidden in various types of data that are used in life sciences applications. The sequence data type, which is the focus of this issue, is just one of the many complex data types that are used in life sciences applications. Life science researchers around the world are working diligently on cracking the hidden codes in these vast, and rapidly growing, biological data sets. Many of the queries that these scientists want to pose require functionality that is well beyond the scope of traditional relational database engines. Database researchers and vendors have a lot to contribute to this field, and hopefully this special issue inspires more interest in the database community to explore and contribute to this exciting area.