

“Bioinformatics challenges in the post-genomic era”

Dr. Edgar Emmanuel Vallejo Clemente

Review by *Roberto Hoyos Morales*

February the 17th, 2005

1. Main Ideas

Molecular biology has produced a vast amount of data. We now possess the raw information, but this data needs tools and applications to be processed, in order for us to understand what is happening. That's the challenge for bioinformatics.

Bioinformatics can serve to disciplines like biology and medicine, to detect and bring therapy to diseases.

There are several strategies being followed to find patterns in the genome sequence, which, of course, are limited, but have brought new information to our notions of genome and biology.

2. Results and Conclusions by the speaker

In year 2005 we got the final version of the human genome sequence. An important discovery is that above 95% of all the information that took years to process is a little bit more that useless, for only 5% of the information actually code proteins (the *exons*).

Among the strategies followed to find patterns in the genome are the following: detection by what the speaker called “homology” (which is compare to already detected sequences) and “aligning” (which is comparing two sequences and find how similar or different they are). The interesting thing to note is that you can only use two and only two sequences in aligning, because more than 2 will create a very complex scenario, which would not be solved in practical times. Other techniques are *filogenetic profile* (search for the occurrences of a pair of proteins; Rosetta Stone (to find proteins *inside* other sequences); gene neighboring (is an adjacency relation; how close they are).

Right now, they are creating models of the interactions between proteins, but only for bacteria and simple organisms like viruses.

3. Discussion

Watson and Crick could not imagine that 50 years after their discovery, genome will be finally ‘cracked’ (and I *really* mean they couldn’t imagine it, because they thought that the secrets of DNA will be close at hand). It is just recently that we got massive amounts of data, but now the challenge is to study that information in every possible detail and extract understanding from it.

Right now we understand a bit more, and our beliefs are radically changed, but the more we advance, the bigger the cliff we have to climb. As the philosopher Socrates once stated: “I know nothing except the fact of my ignorance”.

How much could we have accomplished without the use of computers? Given the fact that we need years to map the human DNA, and about weeks to months to map simple bacteria genomes, will these discoveries be without computers? I want to highlight this point, because being born in a world “with” computers I often imagine how life would be without internet, palms and computers.

Even now, human creativity will have to take full advantage of computers (regarded as *brutes* that follow programmed algorithms).

4. Conclusions

The most important thing in any research activity is not only to obtain the data to use (and to obtain it correctly), but to analyze the information given. I think like projects like SETI or Stanford’s protein folding project, which need huge computing power to apply search algorithms to discover patterns.

We need to create the tools required to undertake such a task. Fortunately there are organizations like the *GeneBank* which provide information to the general public, and we, as computer science engineers, can help to achieve this: by becoming more involved in the applications needed to this end, and perhaps, to be their programmers. There’s too much to do in the programming community.

The most tangible evidence of success in bioinformatics will translate in results for medicine and biology. Diseases such as AIDS and cancer could be controlled or cured more accurately and therapies could be ‘customized’.