

An integrated bioinformatic platform for SNPs detection: the test case of the pig genome

Fronza R⁽¹⁾, Fontanesi L⁽²⁾, Russo V⁽²⁾, Casadio R⁽¹⁾

⁽¹⁾Biocomputing Group, Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna, Italy

⁽²⁾Dipartimento di Protezione e Valorizzazione Agroalimentare, Sezione di Allevamenti Zootecnici, Università di Bologna, Italy

Corresponding author: Rita Casadio, Biocomputing Group, Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna.

Motivation

Motivation The sequencing of the pig genome that has grown rapidly in recent years, is expected to impact both human medicine and pork production. As a matter of fact, pig is an important biomedical model and, for its similarity in several aspects to humans, is more relevant than mouse to human health research priorities such as obesity and diabetes. On the other hand, pig constitutes one of the most important meat protein sources for human consumptions worldwide. Identification of polymorphisms at the whole genome level provides the tools to connect genetic variations to complex phenotypic traits and to understand the biological aspects underlying the pig productions and the use of this species as human model. Single nucleotide polymorphisms (SNPs) are now the markers of choice for high throughput genotyping due to the possibility to design and automate SNP assays, for the stability of these markers, and their density in the genome. In silico SNPs discovery in expressed genes makes use of sequence assembling and aligning, starting from large sets of expressed sequence tags (ESTs). Sequence databases contain more than three millions of porcine ESTs than can be mined and selected to this purpose.

Methods

Methods: We developed an integrated platform for detecting pig SNPs in user-specified fashions using specifically designed query tools that can select genes involved or predicted to have a role in different processes. The pipeline runs in batch mode and includes BLAST package to cluster ESTs and to align identified SNPs with partial porcine genomic sequences, PHRAP to build contigs, CROSS_MATCH and RepeatMasker to remove vectors and repeats. In the pipeline all the components are connected by shell and perl scripts.

Results

Results: The pipeline was first tested on a set of 50 human obesity related genes yielding 177 porcine contigs. Our results highlight 226 biallelic SNPs. The EST contigs were mapped on porcine genomic sequences obtaining sequence information for the design of specific genotyping assays. Genotyping of a sub-set of the putative/computed SNPs is under way. We therefore propose our platform as a suitable tool to select SNPs within a large number of porcine candidate genes chosen according to specific requirements of the end-user. Furthermore, the platform, integrating also other systems biology tool will offer an invaluable help for large genome screening also for any complete genome.

Contact : casadio@biocomp.unibo.it