Genomic and Functional Approaches to Genetic Adaptation

Elena Carnero Montoro

Thesis Director

Dra. ELENA BOSCH

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

UNIVERSITAT POMPEU FABRA

Fitxer PDF de la tesi dividit en 7 parts

Part 2 de 7

3. Evidence of positive selection

Over the last decades, many examples of candidate genes have been revealed by applying statistical tests to variability data in order to identify targets of positive selection. Most neutrality tests were devised in the 1980s, but they could not be widely applied since the last decade, when large quantities of sequence data have been produced for humans and closely related species.

This section will be dedicated to a description of the discovery of genes under positive selection by both candidate-gene and genome-wide studies.

3.1 Candidate gene studies

Traditionally, the identification of genes targeted by positive selection has been driven by a priori hypothesis of the influence of a particular gene on adaptive traits.

Hypothesis-driven studies have provided us with some illustrative examples on how natural selection has shaped the genomic variability in the regions where selected alleles are located, and on the identification of candidate selected loci responsible for adaptation processes. The earliest discovered selected variant related to malaria infection in Africa represent textbook examples of candidate gene studies, such as mutations in the β-globin gene *HBB* (Pagnier 1984), *G6PD* (Tishkoff et al. 2001), the Duffy gene *FY* (Hamblin and Di Rienzo 2000) or in the CD40 ligand gene *TNFSF5* (Sabeti et al. 2002).

Although this type of study has yielded notable success in enlightening some nice examples of recent and past human adaptation, it suffers from three main limitations:

i) An a priori hypothesis about which genes could have been the subject of selection, plus, prior understanding of genotype-phenotype relationships is required. The fact is that often, the functional target of selection is unknown. Some other times, when there is genetic evidence that supports the role of a variant in adaptation, an explanation of how the particular variant confers an advantage to its carriers is unclear. Therefore, the biological meaning of the process of adaptation is missing.

ii) In the cases where there is previous knowledge of which genes could have been involved in adaptation, the favorable variant might not be located in the coding or flanking regulatory sequences, but far away from the gene location, although it still influences its function. This scenario would make it impossible to detect the adaptive variant without a prior evidence of the regulatory elements of the given gene.

iii) It could be difficult to interpret patterns of genetic diversity as some confounding factors mainly due to drift and demography, might have caused a similar pattern. The significance of candidate genes is based on a comparison to the expectation under a neutral model, rather than on empirical evidence.

3.2 Genome-wide studies of positive selection

In the last decade, different technological advances such as high-throughput SNP genotyping or next generation sequencing have been applied to the field of population genomics. Besides producing huge amounts of genome-wide data cost-efficiently, these technologies have facilitated the characterization of genetic variability in thousands of individuals and allowed obtaining complete genomes on a number of different species. More

particularly, the recent availability of large catalogs of genetic variability in different human populations such as those provided by the HapMap or the 1000Genomes projects (Altshuler et al. 2010) (Abecasis et al. 2012) provided an unique opportunity to investigate the impact of natural selection on our genome.

Consequently, genome-wide scans of positive selection became possible and popular, and many large-scale studies have been published on different species and populations, searching for the footprints of different molecular signatures of positive selection both at specific lineage level and at population level. This genomic revolution has represented a great opportunity to overcome the limitations of candidate genes studies.

3.2.1 Population-based genome wide scans

Since Akey et al (2002) work, more than 20 genome-wide scans have been performed to detect recent or ongoing positive selection in humans, with an increase over time of the number of individuals, populations and markers included in the analysis.

The combination of results from all these studies have revealed in one or more studies more than 5000 regions of the genome with some type of signature of positive selection, encompassing more than 400Mb in the genome, and containing more than 4,000 UCSC RefSeq genes. Undoubtedly, this multiplication of candidate loci regarding the ones discovered by gene-by-gene studies, hold a promising chance to guide us to a richer understanding of where positive selection has shaped patterns of genetic variability (Akey 2009) and which traits have been important for the adaptation of each population.

It is important to state that most of the aforementioned studies have based their strategy for identifying targets on the outlier approach,