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 4 de 7

region (Grossman et al. 2010). Second, they applied their method to full sequence genome variation, from the recently published phase1 data from the 1,000genome project, and thus, they were able to explore the complete allele spectrum of different populations to identify putative selected variants. Third, they took advantage of the latest publications of the Encyclopedia of DNA Element (ENCODE) Consortium (Dunham et al. 2012), as well as of the extended genome-wide association (GWA) studies database to appraise the functional relevance of the variation detected as a target of positive selection.

The ENCODE Consortium has provided a very rich catalog of functional annotated elements, identified by a variety of different high-throughput next generation sequencing technologies, as for example the are the characterization of novel non-coding elements, enhancers, patterns of methylation, expression, chromatin states, of the complete human genome of different tissues and cell lines.

This project by Grossman et al (2013) represents a decisive shift in the research of evolutionary adaptation since it is the first one that provides a rich list of positively selected candidate variants as well as a description of their possible functional predicted impact, as described in figure 11, which will be the base of generating new hypotheses about adaptation and new research lines in human evolutionary genetics.

3.2.2 Comparative genome-wide scans

As explained in the previous section, statistical tests using divergence data compare rates of potentially functional and non-functional replacements between species, and try to identify the proportion of functional substitutions that have been fixed due to positive selection. The classical debate among evolutionary biologists about what proportion of differences between different species have been fixed because they were beneficial and because

they enabled the adaptation of that species to their environment is beginning to find an answer.

Due to the nature of the statistical tests used (dN/dS ratio-based test and/or MK-based test, see previous section for more details) most of the genomic surveys have been applied to protein-coding sequences. This is due to the fact that both are based on calculating the ratio of non-synonymous to synonymous substitution rates, assuming synonymous substitutions as neutral sites since they do not account for functional changes and, thus, do not contribute to phenotypic variation. By some modification of the tests, regulatory elements are also beginning to be assessed (see next section).

Initial measures of the rate and proportion of adaptive substitutions revealed important differences between different organisms. For example, the *Drosophila melanogaster* genetic reference panel, based on full genomic sequence data from two different populations, showed that an average of 25.2% of fixed substitutions are adaptive (Mackay et al. 2012). Furthermore, a polymorphism and divergence study in a set of around 100 genes revealed that 57% substitutions in mice were fixed due to positive selection (Halligan et al. 2010). Recent published exome sequenced data on a chimpanzee population of 10 individuals characterizing genome-wide patterns of segregating polymorphisms, found non fixed-adaptive substitutions in the autosomes, but observed that 30% were so in the fast evolving X chromosomes (Hvilsom et al. 2011).

There is compelling evidence in all studies that the proportion of adaptive mutations is proportional to the effective population size (*Ne*), which is something expected since the larger the population is, the more effectively selection acts. This can explain why adaptive evolution is low in humans (α reaches values of at most 13%) (Boyko et al. 2008) compared to other organisms.

3.2.3 Functional enrichment analysis

Aside from identifying individual new candidate loci, genome-wide analysis allow a more global question on which type of genes or functions have participated in the adaptation to new environments. For this purpose, functional enrichment analyses are very useful. They basically test statistically whether the set of loci identified by maps of positive selection are enriched in a given biological relevant processes in the collection of genes studied, which is to say, if there are more loci belonging to a class than what is expected by chance from the background list of genes analyzed.

Many databases are currently available to explore the set of genes that belong to a particular functional category, as for example the Gene Ontology (GO), the Protein Analysis Through Evolutionary Relationships (PANTHER) and/or the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

Population-based studies have revealed different categories enriched for signals of positive selection such as metabolic processes, olfactory reception, skin pigmentation, immunity, hair and sweat, etc (Kelley et al. 2006) and comparative-based studies have revealed enriched categories such as immunity and defense, or sensory perception (Marques-Bonet, Ryder, and Eichler 2009) (Kosiol et al. 2008) (see figure 9).

An important limitation of the method is that it assumes gene independence, or no gene-to-gene interaction, and it does not weight the importance of genes within each functional category.
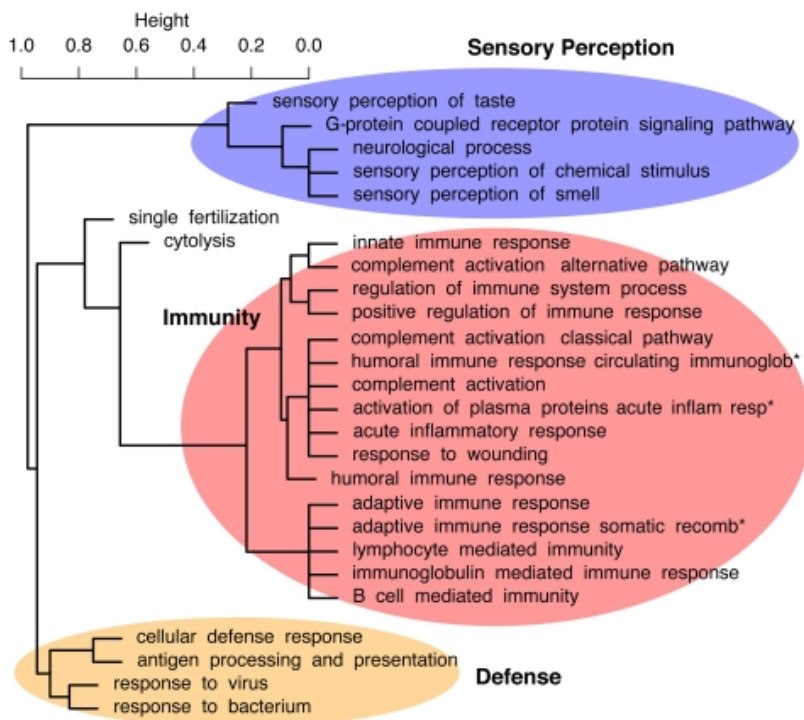
Figure 9. Enrichment functional analysis of positive selected genes (PSG).
Cluster analysis of over-represented GO categories in six different mammals
(Kosiol et al. 2008).