The mosaic structure of HLA genes

Pedro Cano and Katie Rozar
HLA Laboratory, Moffitt Cancer Center, Tampa, FL, USA

Aim

Although single-point (SP) mutations are observed more often in the HLA laboratory than gene conversion (GC) events, the polymorphism of established alleles is mostly accounted for by GC. According to the neutral theory of molecular evolution, the fate of alleles in a population is either allele fixation (frequency reaches 100%), or loss (frequency falls to 0%). This occurs via random transmission of alleles from generation to generation, disregarding natural selection. High allele turnover is evident in any population, with new alleles emerging from diverse mutation mechanisms and existing alleles disappearing by genetic drift. Low frequency alleles, or those that have emerged most recently, are the most likely to disappear. The HLA system violates the neutral allele turnover prediction because it evolves in a *balancing selection pattern* – since HLA polymorphism provides survival advantages, genetic diversity is maintained.

Most emergent alleles within a population result from SP mutations, which appear and disappear. New alleles resulting from GC events can also disappear, but are more likely to persist – why? It is hypothesized that new alleles attributed to GC events have a selective advantage in their ability to present peptides since DNA segments already have key elements in building antigen-presenting pockets.

Methods

The laboratory HLA typing database (SQL Server) was queried for all HLA alleles identified since 2017, as well as those denoted as new (e.g., HLA-A*NEW) according to laboratory protocol. Allele assignment notations for each new allele were categorized as SP or GC based on the type of mutation described.

Results

Out of 492,318 total alleles identified, 68 (0.01%) were considered new, or not classified by the IMGT nomenclature version available at the time of analysis (Table 1). Of the 68 new alleles, 64 (94.1%) resulted from SP mutations, while 4 (5.9%) resulted from GC events.

Table 1. Characterization of Alleles Identified through Query of Laboratory HLA Typing Database

Metric	Value
Total alleles identified	492,318
New alleles identified	68
New alleles due to SP mutations	64
New alleles due to GC mutations	4

Note. ‘New alleles’ refers to the number of new alleles identified based on the IMGT nomenclature version
available at the time of analysis. Abbreviations: single-point (SP); gene conversion (GC).

Conclusions

Although HLA gene polymorphism results mainly from GC events, most mutations observed in the laboratory are SP mutations. This discrepancy provides empirical evidence of the evolutionary advantage of GC events. Further, HLA polymorphism may be characterized by the proportion of SP to GC mutations, which may be calculated by building phylogenetic trees of HLA alleles. This proportion is difficult to calculate with precision because different GC events can account for the same polymorphisms. Nonetheless, the proportion favors GC events, as most SP mutations are observed in rare alleles that hardly contribute to HLA diversity. Additionally, GC units share DNA segments that account for antigen-antibody epitopes, explaining their selective advantage.

Importantly, the role of GC events provides the basis for a mosaic structure of HLA genes, whereby alleles are composed of DNA segments that have been exchanged and recombined over time.