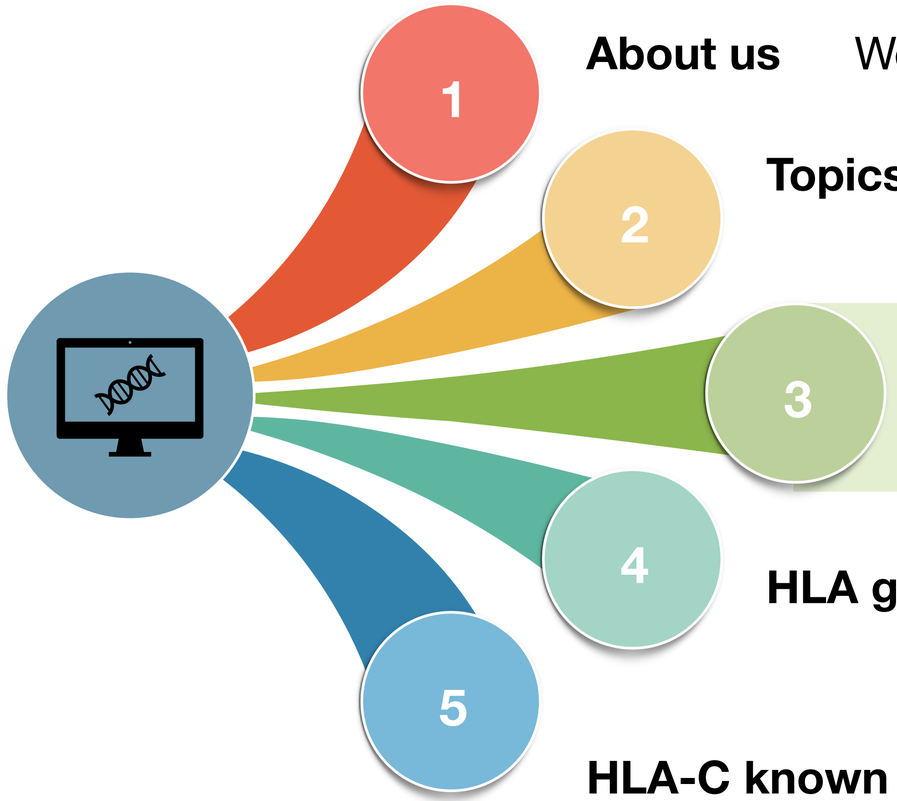
Dissection of HLA-C gene region to investigate its association with complex traits

Locatelli E.[1], Moron Dalla Tor L.[1], Treccani M.[1], Veschetti L.[1], De Tomi E.[1], Stefani C.[1], Tamburin S.[1], Zipeto D.[1], Trabetti E.[1], Patuzzo C.[1], Malerba G.[1]

[1] Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy

Background



About us We are a bioinformatic group of the University of Verona, in Italy.

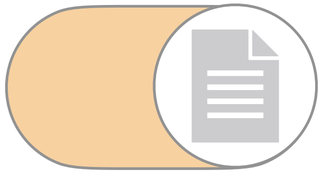
Topics covered Bacterial genomes, RNA-seq analysis, microbiome analysis, and association studies.

Project Focus on HLA-C gene and its variability.
Classification of the alleles.

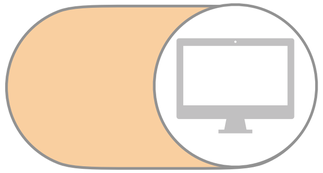
HLA gene The genomic region of HLA gene cluster (6p21.3) contains several highly polymorphic genes involved in the immune response.

HLA-C known alleles More than 3500 existing alleles.

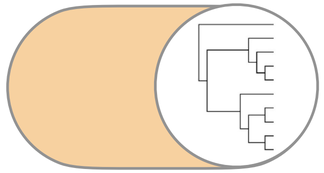
Methods



Data Collection:
We investigated more than 3500 known alleles of HLA-C gene (from IPD-IMGT/HLA database) that are grouped into 14 serogroups (e.g., C*01:02:01:01).



Data analysis:
All the sequences have been aligned against the human genome reference sequence (both versions; hg19 and hg38), highlighting more than 1500 SNPs.



Phylogenetic trees:
A clustering approach was used in order to understand how the alleles are evolutionarily connected.
Neighbor Joining algorithm to build phylogenetic trees.

Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy

Results

Alleles of the same serogroup share strong identity in their sequence. Interestingly, as general rule, we observed that the main tree (figure 1) presents two branches, containing each a similar structure (relative distance between serogroups).

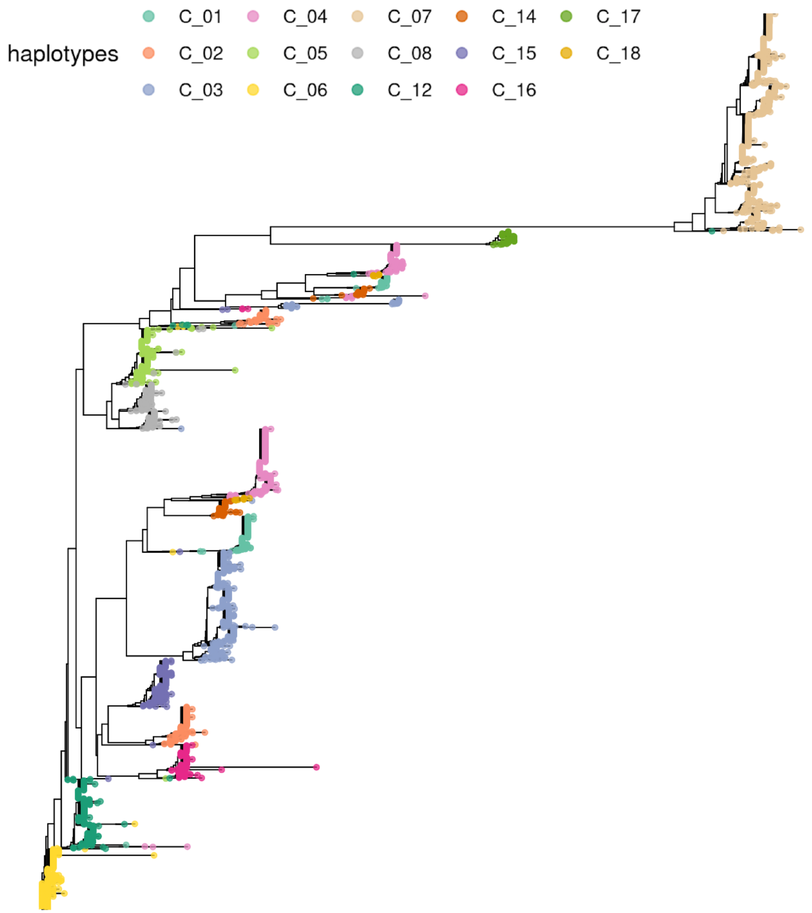


Figure 1. The entire set of HLA-C alleles

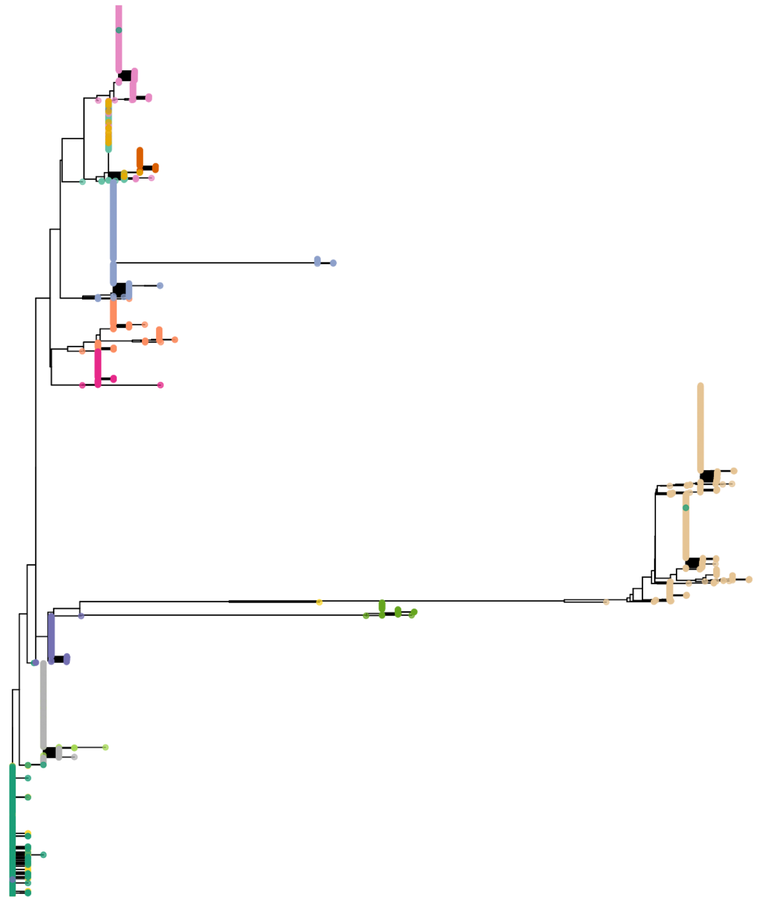


Figure 2. Alleles subsequences (from 1 to 2000 bp)

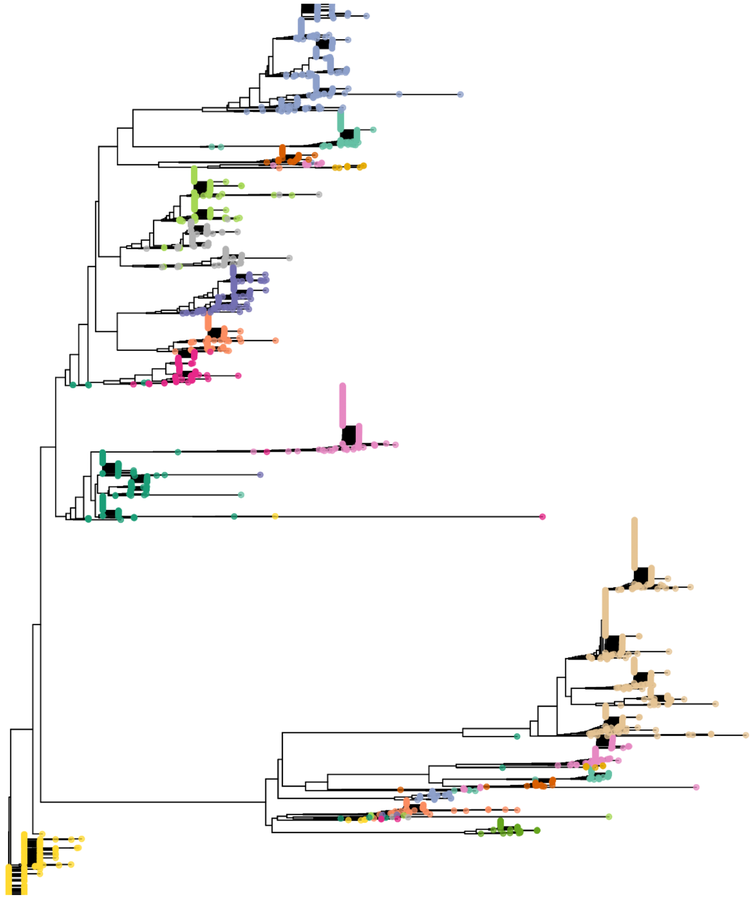


Figure 3. Terminal part of the sequences (from 2000 to 3000 bp)

Conclusions

- We used bioinformatic approaches to investigate the HLA-C gene and its many alleles, creating a phylogenetic tree to better understand the sequence similarities.

- We are studying the possible features that characterize these alleles, and which are the possible SNPs able to distinguish the several haplogroups of HLA-C gene.

- The study will go on by investigating the association of the HLA-C serogroup-SNP-based alleles with kidney related disease (INCIPE study) and Alzheimer's disease (NIAGADS database) in large cohorts of individuals.

References

Turner S, Ellexson ME, Hickman HD, Sidebottom DA, Fernández-Viña M, Confer DL, Hildebrand WH. Sequence-based typing provides a new look at HLA-C diversity. J Immunol. 1998 Aug 1;161(3):1406-13. PMID: 9686604.

Robinson J, Guethlein LA, Cereb N, Yang SY, Norman PJ, Marsh SGE, Parham P. Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. PLoS Genet. 2017 Jun 26;13(6):e1006862. doi: 10.1371/journal.pgen.1006862. PMID: 28650991; PMCID: PMC5507469.

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015 Jan;43(Database issue):D423-31. doi: 10.1093/nar/gku1161. Epub 2014 Nov 20. PMID: 25414341; PMCID: PMC4383959.

Sievers F, Higgins DG. Clustal omega. Curr Protoc Bioinformatics. 2014 Dec 12;48:3.13.1-16. doi: 10.1002/0471250953.bi0313s48. PMID: 25501942.

R: https://cran.r-project.org/