Integrating SSHAARP Maps into the Updated Allele Frequencies Net Database (AFND v2)

Steven J. Mack[1], Katrina A. Eaton[1], Livia Tran[1], Erik Rozemuller[2], Wietse Mulder[2], Derek Middleton[3,4], Faviel González[3,5]

1: Department of Pediatrics, UCSF, Oakland, CA, USA; 2: TxMiller Foundation, Zeist, Netherlands; 3: Allele Frequencies Net Database;

4: Institute of Infection and Global Health, University of Liverpool, Liverpool, UK; 5: Center for Biomedical Research, Autonomous University of Coahuila, Torreón, Mexico.

OVERVIEW

Allele Frequencies Net Database (AFND; www.allelefrequencies.net)

- Houses global allele and haplotype frequency data for the HLA, KIR, MIC and cytokine genes for over 1,800 population studies[1], including >156,000 HLA allele and haplotype frequencies for 1324 population studies. AFND has included global maps identifying population locations and the local frequencies of individual alleles.

Searching Shared HLA Amino Acid Residue Prevalence (SSHAARP) R package (CRAN.R-project.org/package=SSHAARP)

- SSHAARP's PALM() function generates high-contrast color global frequency heat maps illustrating the frequency distributions of HLA alleles, haplotypes and amino-acid motifs (Fig 1).

AFNDtoPALM (Fig.2):

- A new SSHAARP function that generates SSHAARP-style frequency heatmaps from AFND allele frequency data.

METHODS

SSHAARP API (Fig.3)

- The API is built in Python 3 using FastAPI.
- rpy2 allows SSHAARP functions to be called from a Python environment.
- SQLAlchemy connects to the MSSQL database, with communications mediated by the pyodbc (python-based open database connectivity) driver.
- SQLAlchemy downloads AFND frequency data to a pandas dataframe, which is passed to AFNDtoPALM() for map generation.

AFND Map Generation (Fig 4)

- A python-based web service integrates FastAPI, rpy2, mssql-python, pandas and SQLAlchemy to extract and format AFND data for AFNDtoPALM().
- AFNDtoPALM() converts AFND data to *solberg_dataset*[2] format and builds a SSHAARP heatmap from AFND frequencies for an HLA allele.

AFND MAP Storage (Fig. 5)

- The SSHAARP API communicates with AFND's Python environment, querying the AFND-Maps repository for extant maps, or calls AFNDtoPALM() to generate new maps, which are stored in the AFND-Maps repository and returned to the user.

RESULTS

- The AFNDtoPALM() function was added to version 2.1.0 of the SSHAARP Package.
- The AFND-SSHAARP maps in Figure 2 were generated using 'toy' datasets of 31, 64, and 151 AFND populations.
- AFND-PALM maps have been generated for the 545 common, 513 intermediate, and 1997 well-documented (CIWD) HLA alleles[4] (github.com/nmdp-bioinformatics/CIWD).

CONCLUSIONS

We integrated python and R systems with AFND's database to generate SSHAARP frequency heat maps for HLA alleles in the CIWD 3.0.0 catalogue.

AFND houses data for alleles at a variety of resolutions. While initial frequency heat maps will be generated at two-field resolution (Figure 2). AFND users will be able to specify 1- to 4-field resolution map generation.

The greater number of AFND populations (~1800), relative to the Solberg Dataset (~500), can result in more granular, frequency-information-dense maps than those that are available via the GFBs. These can maps be generated for any HLA allele or multi-locus haplotype in AFND.

LITERATURE CITED

1. Gonzalez-Galarza, F. F., et al. 2024. "Allele Frequency Net Database." Methods in Molecular Biology 2809: 19–36.
2. Solberg, O. D., et al. 2008. Balancing Selection and Heterogeneity Across the Classical Human Leukocyte Antigen Loci: A Meta-Analytic Review of 497 Population Studies. Human Immunology 69, no. 7: 443–464.
3. Tran L., et al. HLAtools, Searching Shared HLA Amino Acid Residue Prevalence, and the Global Frequency Browsers: New Computational Resources for Working With HLA Data and Visualizing Global Patterns of HLA Variation. Int J Immunogenet. 2025 Dec;52(6):358-370.
4. Hurley, C. K., et al. 2020. Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0 HLA. 2020 Jun;95(6):516-531.

DQB1*02:01 DPB1*01:01

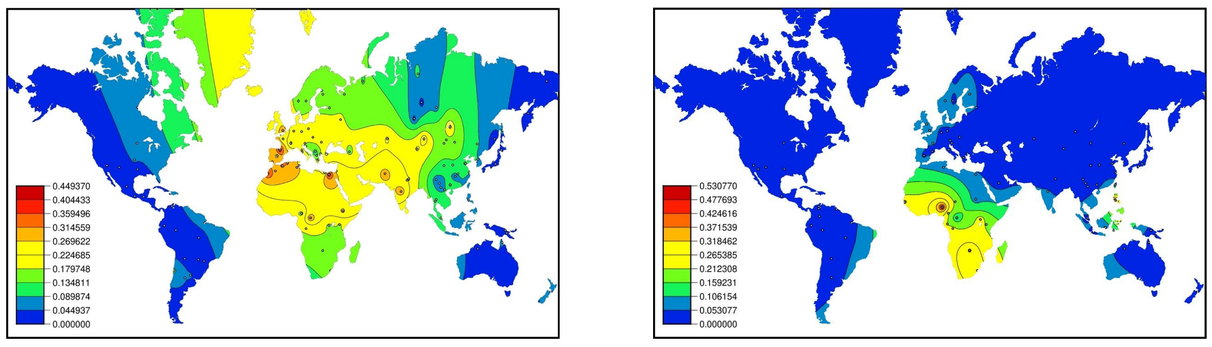
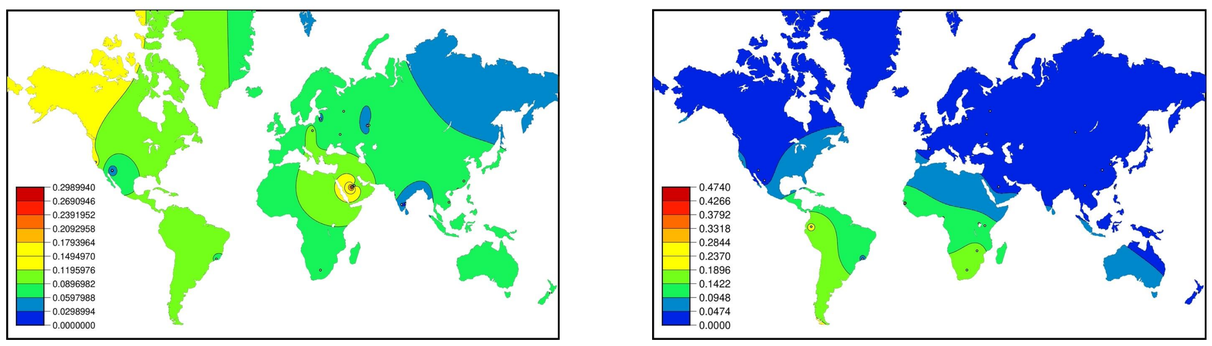


Figure 1: Global Frequency Maps describing HLA-DQB1*02:01 and DPB1*01:01 frequencies in indigenous populations, generated by **SSHAARP** from solberg_dataset[2].

DQB1*02:01 DPB1*01:01



DQB1*02:02

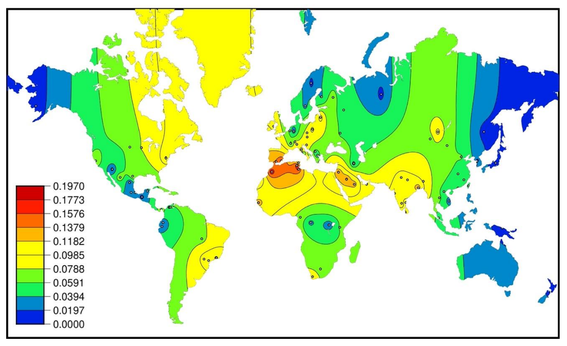


Figure 2: Global Frequency Maps describing DQB1*02:01, DQB1*02:02 and DPB1*01:01 frequencies in all available populations, generated by SSHAARP's **AFNDtoPALM()** function from AFND data.

The DQB1*02:01 map was generated using 35 AFND populations, the DQB1*02:02 map using 151 populations, and the DPB1 map using 64 populations.

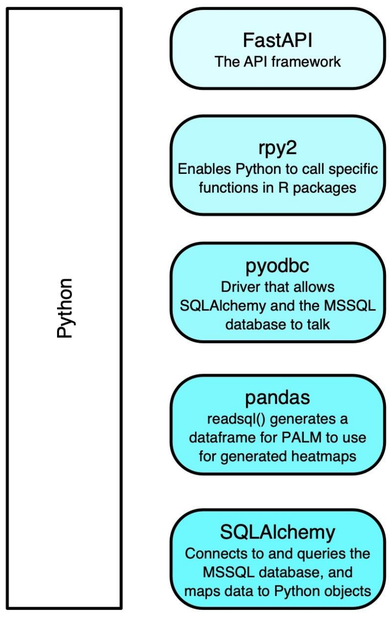


Figure 3. SHAARP API Components

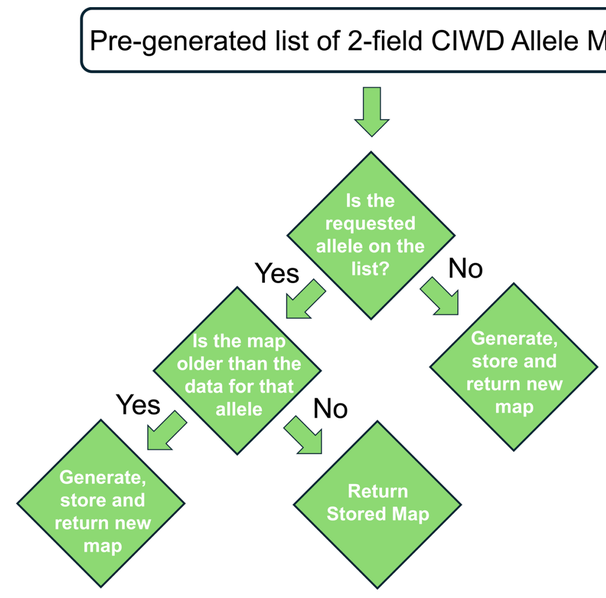


Figure 4. SSHAARP API Mapmaking Decision Tree

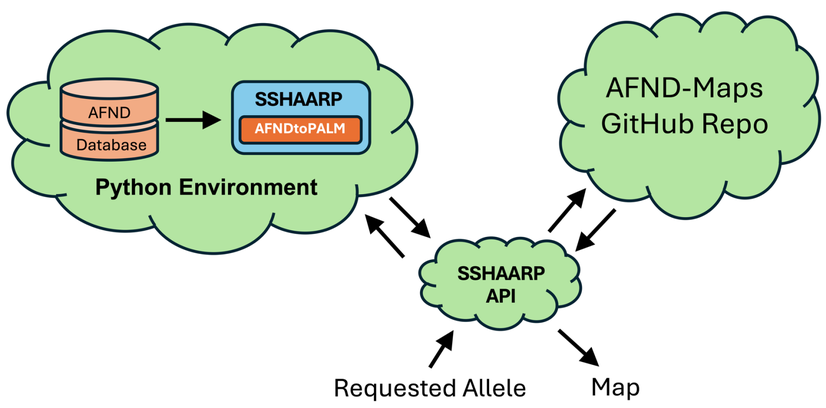


Figure 5. Cloud-based SSHAARP API Resource Interactions