

Background:

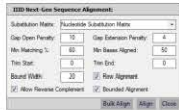
Bioinformaticians have gained extensive knowledge and accumulated diverse sets of skills in the use and understanding of a wide range of genome alignment and analysis tools. Many of these tools have been developed by computer scientists with minimal domain expertise, or by experts in the bioinformatics field lacking strong graphical user interface design skills. This project has been funded by a world class institute and aims to combine ideas from experts in the appropriate fields to develop a solution which combines cutting-edge genomic based algorithms with an enhanced visual user experience.

Many viruses, including the clinically relevant RNA viruses HIV (human immunodeficiency virus) and HCV (hepatitis C virus), exist in large populations and display high genetic heterogeneity within and between infected hosts. VGAS's next generation alignment tool can generate accurate alignments more suited for samples with minimal intra host variations. VGAS ties together the alignment, visualization and analysis of the sequence data combined with epitope mapping to facilitate the assessment of intra-patient viral genetic diversity. This is essential for understanding the evolutionary dynamics of viruses, for designing effective vaccines, and for the success of antiviral therapy.

Methods:

VGAS has followed a three pronged approach to ensure you can use its powerful visualization features on sequences that can be aligned using various approaches.

- 1) Aligned sequences can be imported into VGAS. This allows any aligner to be used.
- 2) Sequences can be aligned within VGAS by VGAS passing the data to a selected range of sequence aligners.
- 3) Sequences can be aligned within VGAS using the built in sequence alignment tools. These have been extensively tested with FLX sequence data, from viral genomes and the HLA regions of the human genome.



The built in alignment tool follows three steps to align all reads to a reference.

- 1) Reads are roughly aligned by finding the offset of the read to the reference genome by breaking the read into k-mers which are matched to the reference.
- 2) Reads are pairwise aligned to the reference sequence using a bounded Gotoh aligner.
- 3) A single alignment is formed by combining the pairwise alignments, using the once a gap, always a gap policy.

Results / Conclusions:

Based on this expertise, a visual genome analysis studio (IIID VGAS) was specifically developed with the ambition of setting a new benchmark in genomic alignment visualization. Furthermore, VGAS offers an expanding suite of post-alignment tools presenting all data visually allowing for further analysis and the interpretation of results. To date, the set of integrated tools include various scalable viewers (e.g. reference genome and epitope mapping) and high performance sequence aligners (supporting FASTA, Sanger, and Next-Gen sequencing). References can be imported directly from various publicly available repositories (Genbank, Ensembl, IMGT) or loaded from the VGAS core database, the design structure is the result of over 5 years collaboration between scientists and software engineers.

VGAS running at the Institute



Displayed over three 42 inch television screens with over 50,000 full length HIV-1 sequences loaded. The application runs smoothly allowing the user to perform and visual such features as epitope mapping against patient sequences and HLA.

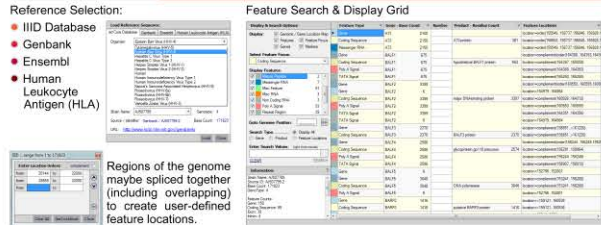


Screenshots Reference Genome Viewer - Epstein Barr Virus

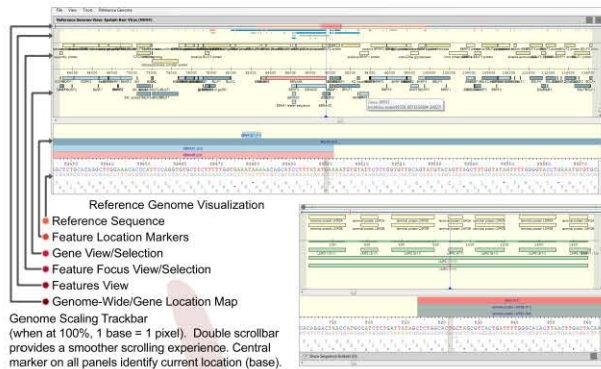
Reference Selection:

- IIID Database
- Genbank
- Ensembl
- Human Leukocyte Antigen (HLA)

Feature Search & Display Grid



Regions of the genome maybe spliced together (including overlapping) to create user-defined feature locations.



Reference Genome Visualization

- Reference Sequence
- Feature Location Markers
- Gene View/Selection
- Feature Focus View/Selection
- Features View
- Genome-Wide/Gene Location Map

Genome Scaling Trackbar (when at 100%, 1 base = 1 pixel). Double scrollbar provides a smoother scrolling experience. Central marker on all panels identify current location (base).

Sequence, Analysis, Alignment & Visualization - HIV-1

Sequence Loading & Alignment Options:

- Load reference aligned FASTA files
- Load multiple sequence Aligned FASTA files
- Align unaligned FASTA files
- Align Next-Gen sequences
- Align Sanger (ABI) files (see images right)

