**Tool Summary: repgenHMM**

* **Tool/Model Name:**  
  The tool is named **repgenHMM**, short for "repertoire generative Hidden Markov Model."
* **Reference:**  
  It was introduced by Elhanati et al. in 2016 in *Bioinformatics* (DOI: 10.1093/bioinformatics/btw112), with code available at [bitbucket.org/yuvalel/repgenhmm](https://bitbucket.org/yuvalel/repgenhmm).
* **Category:**  
  repgenHMM belongs to the category of **V(D)J recombination simulation** and **probabilistic model inference tools**.
* **Main Purpose/Function:**  
  The tool is designed to infer the probabilistic rules governing immune receptor generation—such as gene usage, deletions, and insertions—based solely on non-productive immune sequences using a Hidden Markov Model (HMM).
* **Key Features:**  
  repgenHMM implements a probabilistic generative model that uses a modified Baum–Welch Expectation-Maximization (EM) algorithm.  
  It supports both TCR α- and β-chains and includes detailed modeling of bi-directional D gene deletions and independent VD and DJ insertions.  
  The tool can generate synthetic sequences, calculate generation probabilities, estimate repertoire entropy, and is implemented in C++ with multithreading support.
* **Input Requirements:**  
  The tool requires a FASTA or plain-text file of non-productive TCR sequences, along with reference germline V, J, and D gene segments—typically from the IMGT database.  
  It also allows optional configuration of maximum insertion/deletion lengths and alignment score thresholds.
* **Output:**  
  repgenHMM outputs learned model parameters such as P(V,J), deletion and insertion distributions.  
  It computes the generation probability for each sequence and can generate synthetic recombination events.  
  Additional outputs include entropy estimates, gene usage frequencies, and statistical correlations.
* **Validation/Evaluation:**  
  The tool has been validated using synthetic datasets with known parameters, where it successfully recovers those values.  
  In comparison to MiXCR, it more accurately reconstructs true insertion and deletion distributions.  
  It has been applied to real TCRα and TCRβ datasets, and its entropy estimations match established biological expectations.
* **Strengths:**  
  repgenHMM provides highly accurate probabilistic modeling of immune receptor generation, integrating over all plausible rearrangement paths rather than relying on a single best alignment.  
  It enables deeper biological interpretation by allowing generation probability estimation, repertoire entropy analysis, and exploration of public clones.  
  Its modeling approach is more biologically realistic than deterministic alignment-based tools.
* **Limitations:**  
  The tool is significantly slower than alignment-based tools such as MiXCR (e.g., ~200 seconds versus ~10 seconds for 50,000 sequences).  
  It does not currently support somatic hypermutation (SHM) or amino acid–level modeling.  
  It requires pre-alignment for V and J genes, and for β-chain modeling, it lacks pre-alignment for D genes, which may reduce accuracy.
* **Applications in Autoimmunity:**  
  While repgenHMM was not originally applied to autoimmune datasets, it is well-suited for modeling baseline repertoire generation in autoimmune vs. healthy individuals.  
  It can be used to detect shifts in rearrangement patterns that may indicate disease-specific immune alterations.  
  The tool also offers a foundation for comparing generative repertoires to those shaped by selection, which is valuable in immunopathology research.
* **Notable Citations and Use Cases:**  
  The tool was introduced in the 2016 publication by Elhanati et al. and is built on earlier work by Murugan et al. (2012).  
  It has been cited in numerous computational immunology studies and is frequently used by the Mora–Walczak group in immune repertoire analysis pipelines.
* **Reviewer’s Commentary:**  
  repgenHMM is considered a foundational tool in computational immune repertoire modeling.  
  It is especially well-suited for non-productive sequence analysis, where selection bias is minimized.  
  Its integration with selection models like SONIA, support for SHM, and amino acid–level generation are recommended as promising future directions.  
  The tool also has strong potential for patient-specific modeling in autoimmunity and other immune-related diseases.
* **Simulation Level / Output Granularity:**  
  repgenHMM simulates the **raw nucleotide-level V(D)J recombination process** and does not produce **AIRR-compliant repertoires** with annotations such as CDR3, isotype, or clonal lineage.  
  It does not simulate downstream processes like clonal expansion, selection, or somatic hypermutation, focusing purely on the generative stage of receptor diversity.
* **Visualization**

repgenHMM does **not include built-in visualization features**. However, its numeric outputs—such as deletion/insertion distributions, gene usage frequencies, generation probabilities, and repertoire entropy—are well-structured and can be **easily visualized using external tools** such as **R** or **Python** libraries (e.g., matplotlib, seaborn, ggplot2).