

Resource Summary Report

Generated by SPARC SAWG on Aug 9, 2026

Yandell Lab Portal

RRID:SCR_000807

Type: Tool

Proper Citation

Yandell Lab Portal (RRID:SCR_000807)

Resource Information

URL: <http://www.yandell-lab.org/software/index.html>

Proper Citation: Yandell Lab Portal (RRID:SCR_000807)

Description: Sequenced genomes contain a treasure trove of information about how genes function and evolve. Getting at this information, however, is challenging and requires novel approaches that combine computer science and experimental molecular biology. My lab works at the intersection of both domains, and research in our group can be summarized as follows: generate hypotheses concerning gene function and evolution by computational means, and then test these hypotheses at the bench. This is easier said than done, as serious barriers still exist to using sequenced genomes and their annotations as starting points for experimental work. Some of these barriers lie in the computational domain, others in the experimental. Though challenging, overcoming these barriers offers exciting training opportunities in both computer science and molecular genetics, especially for those seeking a future at the intersection of both fields. Ongoing projects in the lab are centered on genome annotation and comparative genomics; exploring the relationships between sequence variation and human disease; and high-throughput biological image analysis. Current software tools available: VAAST (the Variant Annotation, Analysis & Search Tool) is a probabilistic search tool for identifying damaged genes and their disease-causing variants in personal genome sequences. VAAST builds upon existing amino acid substitution (AAS) and aggregative approaches to variant prioritization, combining elements of both into a single unified likelihood-framework that allows users to identify damaged genes and deleterious variants with greater accuracy, and in an easy-to-use fashion. VAAST can score both coding and non-coding variants, evaluating the cumulative impact of both types of variants simultaneously. VAAST can identify rare variants causing rare genetic diseases, and it can also use both rare and common variants to identify genes responsible for common diseases. VAAST thus has a much greater scope of use than any existing methodology. MAKER 2 (updated 01-16-2012) MAKER is a portable and easily configurable genome annotation pipeline. It's purpose is to allow smaller eukaryotic and prokaryotic genome projects to independently annotate their genomes and to create genome databases. MAKER identifies repeats, aligns ESTs and proteins to a genome, produces ab-initio gene predictions and

automatically synthesizes these data into gene annotations having evidence-based quality values. MAKER is also easily trainable: outputs of preliminary runs can be used to automatically retrain its gene prediction algorithm, producing higher quality gene-models on subsequent runs. MAKER's inputs are minimal and its outputs can be directly loaded into a GMOD database. They can also be viewed in the Apollo genome browser; this feature of MAKER provides an easy means to annotate, view and edit individual contigs and BACs without the overhead of a database. MAKER should prove especially useful for emerging model organism projects with minimal bioinformatics expertise and computer resources. RepeatRunner RepeatRunner is a CGL-based program that integrates RepeatMasker with BLASTX to provide a comprehensive means of identifying repetitive elements. Because RepeatMasker identifies repeats by means of similarity to a nucleotide library of known repeats, it often fails to identify highly divergent repeats and divergent portions of repeats, especially near repeat edges. To remedy this problem, RepeatRunner uses BLASTX to search a database of repeat encoded proteins (reverse transcriptases, gag, env, etc...). Because protein homologies can be detected across larger phylogenetic distances than nucleotide similarities, this BLASTX search allows RepeatRunner to identify divergent protein coding portions of retro-elements and retro-viruses not detected by RepeatMasker. RepeatRunner merges its BLASTX and RepeatMasker results to produce a single, comprehensive XML-based output. It also masks the input sequence appropriately. In practice RepeatRunner has been shown to greatly improve the efficacy of repeat identification. RepeatRunner can also be used in conjunction with PILER-DF - a program designed to identify novel repeats - and RepeatMasker to produce a comprehensive system for repeat identification, characterization, and masking in the newly sequenced genomes. CGL CGL is a software library designed to facilitate the use of genome annotations as substrates for computation and experimentation; we call it CGL, an acronym for Comparative Genomics Library, and pronounce it Seagull. The purpose of CGL is to provide an informatics infrastructure for a laboratory, department, or research institute engaged in the large-scale analysis of genomes and their annotations.

Resource Type: data or information resource, software resource, software application, data analysis software, laboratory portal, portal, organization portal, data processing software

Defining Citation: [PMID:21700766](#), [PMID:21700266](#), [PMID:21325948](#), [PMID:21347285](#)

Keywords: software, gene, genome annotation, human

Resource Name: Yandell Lab Portal

Resource ID: SCR_000807

Alternate IDs: nlx_144364

Record Creation Time: 20220129T080203+0000

Record Last Update: 20260808T185731+0000

Ratings and Alerts

No rating or validation information has been found for Yandell Lab Portal.

No alerts have been found for Yandell Lab Portal.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We have not found any literature mentions for this resource.