

Integrative Services for Genomic Analysis (ISGA)

An Intuitive Web Server for Prokaryotic Genome Annotation and Other Analyses

Chris Hemmerich, Aaron Buechlein, Ram Podicheti, Jeong-Hyeon Choi

Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana



ISGA is an intuitive web application designed for biologists to run and customize bioinformatics pipelines built using the Ergatis workflow system. ISGA addresses the gap between prokaryotic annotation services such as JCVI's Annotation Service that are easy to use but provide few customization options and workflow systems such as Ergatis which are powerful but prohibitively complex for most biologists to use.

Sequencing facilities can use a local installation of ISGA to offer data analysis services in conjunction with NGS operations. ISGA allows biologists to initiate and customize pipelines themselves rather than involving bioinformaticians. ISGA is open source software and freely available under the Apache License, Version 2.0.

We are active members of the Ergatis community and ISGA pipelines can be run directly in Ergatis and converting new Ergatis pipelines to ISGA is straightforward.

Key Features:

- Easy-to-use web interface for bioinformatics pipelines
- Free demonstration site and download for local install
- Run pipelines immediately or customize to meet your needs
- Form validation catches pipeline errors immediately
- Account system allows data privacy and future retrieval
- Integrated analysis of pipeline results
- Pipelines are built and run with Ergatis Workflow System
- Integrated tools for analyzing pipeline results

Step 1: Select and Customize Pipeline

Customize the component and parameters of a pipeline or run with defaults

Toggle components and open parameter editor

Edit program parameters (w/ help and validation)

Additional Gene Analysis

Please check the components you wish to run or use the Edit Parameters link to edit that component's parameters.

Component	Turn On/Off	Edit Parameters
NCBI BLASTX Intolerance Regions	<input type="checkbox"/>	Edit
SignalP	<input type="checkbox"/>	No Parameters
TMHMM	<input type="checkbox"/>	No Parameters
Protein Scan	<input type="checkbox"/>	Edit
TransTerm	<input type="checkbox"/>	Edit
Lipoprotein Hoof Discovery	<input type="checkbox"/>	No Parameters
Prinm plot	<input type="checkbox"/>	Edit
BLAST vs COG	<input type="checkbox"/>	Edit
Oligo Picker	<input type="checkbox"/>	Edit
Asgard	<input type="checkbox"/>	Edit

Select All | Deselect All

Cancel Save

BLASTP vs COG

Required Parameters

Scoring Matrix: BLOSUM62

e-value: 1e-5

BLAST Filter: none

Number of Matches: 150

Number of Descriptions: 150

Maximum number of GSs: 5

Optional Parameters

Cancel Reel Form Save Parameters

The CGB hosts a local instance of ISGA for our collaborators and public evaluation. ISGA is open source and available for download and local installation at:

<http://isga.cgb.indiana.edu/>

Available Pipelines:

- Prokaryotic Genome Annotation (see left)
- Prokaryotic Assembly

Pipelines In Development:

- Transcriptome Annotation (see below)
- Gene Clustering against a Phylogenetic Tree
- Functional Genomics Analysis of NGS Data

Step 2: Run Pipeline

Submission Builder

Before submitting your run, you must select the data files the submission will process. Additionally, you may need to set organism information specific to the file being submitted.

Name	Pipeline
Prokaryotic Annotation Pipeline Run 13	Prokaryotic Annotation Pipeline

Organism Information

The following parameters will be used for this run. You may set or edit these parameters by clicking the link [Provide Organism Information](#) in the right toolbox.

Detail	Value
Organism Type	gram-
Genus	Escherichia
Species	coli
Locus Prefix	ISGA

Input Data

Required	Format	File
Genome Sequence	FASTA	Celera Assembly Pipeline Run 25.fasta

Prokaryotic Annotation Pipeline Run Status

This page was refreshed at Jul 7, 2010 21:32. It will refresh next at Jul 7, 2010 21:37.

Name	Prokaryotic Annotation Pipeline Run 12
Account	Chris Hemmerich
ID	53699812
Status	Running (Show Detailed Status)
Started At	Jul 07, 2010 21:31 EDT
Description	

Input Files

Input	File
Genome Sequence (FASTA)	Celera Assembly Pipeline Run 25.fasta

Results

Contents	Format	Created On	View / Download
Output ?	Table ?	Not Available Yet	
Pipeline Summary ?	GenBank ?	Not Available Yet	
Pipeline Summary ?	GFF ?	Not Available Yet	
Additional Gene Analysis ?			
BLAST Search Result ?	BLAST Raw Results ?	Not Available Yet	
OligoPicker Result ?	OligoPicker Raw Results ?	Not Available Yet	
Asgard Result ?	Asgard Raw Results ?	Not Available Yet	

Raw Output

As per our [data retention policy](#), your raw output may be deleted on or after 5 Oct 2010. If this data is important to you, download these files before that date.

Cluster	Download
Process Gene Prediction	Not Available Yet
TFBS Search	Not Available Yet
ORF Prediction	Not Available Yet
Additional Gene Analysis	Not Available Yet
Protein Domain Search	Not Available Yet
Sequence Similarity Search	Not Available Yet
RNA Prediction	Not Available Yet
Alternate Start Site Analysis	Not Available Yet

Upload new input file or selecting existing file in ISGA
Provide any run-specific parameters (e.g. Organism species)

Monitor run progress and download results

Step 3: Analyze Results

Showing 10 kbp from gi_49175990, positions 1 to 10,001

Instructions: [Bookmark this](#) | [Upload your own data](#) | [Show banner](#) | [Share these tracks](#) | [Link to image](#) | [Help](#) | [Print](#)

Overview

Load results of annotation into integrated GBrowse genome browser

ISGA_2

Locus ISGA_2
Position 8168 - 8971 on plus strand
Length 804

Gene Symbol
Gene Symbol Source

Description 2,5-diketo-D-gluconic acid reductase B
TIGR_role 157

Gene Product SPI30863
Name Source [C] COG0667 Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)

Top COG Hit [C] COG0667 Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)

Enzyme Nomenclature 1.1.1.234

Gene Ontology
Gene Name: Prokaryotic_Annotation_Pipeline_Run_1.gene.11743763.1
mRNA Name: Prokaryotic_Annotation_Pipeline_Run_1.transcript.11743760.1
CDS Name: Prokaryotic_Annotation_Pipeline_Run_1.CDS.11743759.1

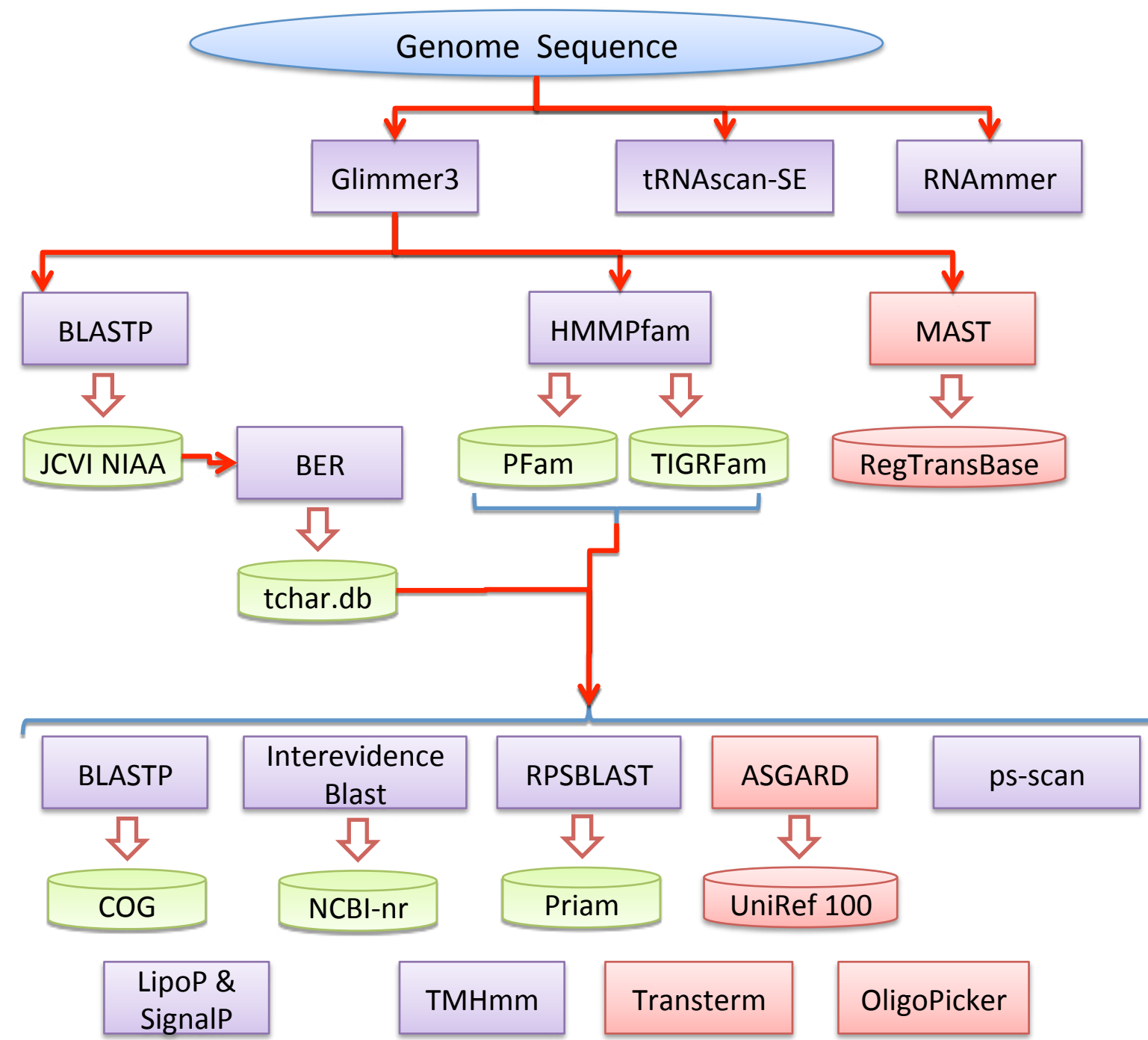
Gene Sequence
Includes bases upstream and bases downstream

Download the Sequence

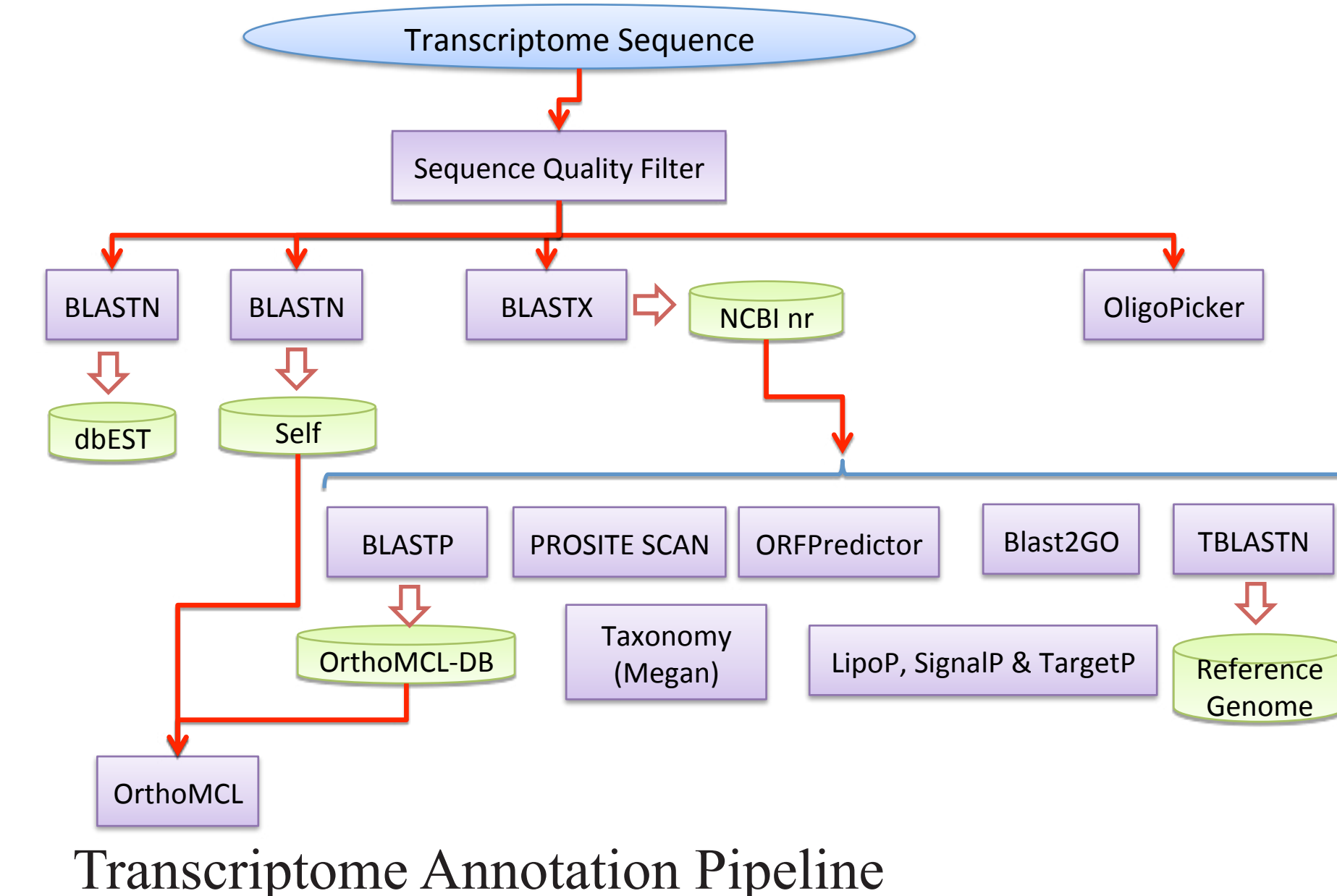
BLAST the Sequence

Download the Sequence

View feature details and BLAST sequences



Prokaryotic Genome Annotation Pipeline



Transcriptome Annotation Pipeline

ISGA Toolbox:

- View pipeline results in GBrowse
- Standalone BLAST Search
- Database query of annotation results
- Convert assembly results to Hawkeye

Manuscript:

An Ergatis-based prokaryotic genome annotation web server

Chris Hemmerich, Aaron Buechlein, Ram Podicheti, Kashi Revanna & Qunfeng Dong, (2010) *Bioinformatics*

This work was supported in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc. and in part by NIH grant 5RC2HG005806-02, "Investigation of Cloud Computing to Support Data-Parallel Health Research".