PipeWorks™

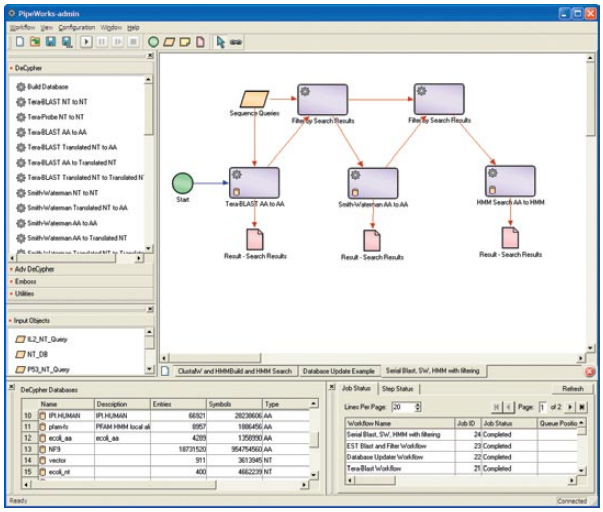
Genomics Workflow Software

Design and Process Genomic Workflows without Scripting

PipeWorks™ provides biologists and genomicists with a visual environment for developing multi-step workflows that are rapidly processed on the TimeLogic biocomputing systems. With drag-and-drop simplicity, PipeWorks lets you stitch together sequence comparisons, data filters, and output results—without Perl scripting. The workflows you create can easily be saved, shared and modified by other PipeWorks users.

By incorporating TimeLogic's accelerated sequence comparison methods (Tera-BLAST™, Tera-Probe™, Smith-Waterman, Framesearch, HMM analysis, Profilesearch and GeneDetective™) into your workflows, you can process huge data sets without a computer cluster. In order to minimize errors from data transfer, manual analysis and scripting, PipeWorks features automatic format translators, reusable input objects (*e.g.*, a gene family being studied by your lab), and easy creation of custom databases.

Intutive Visual Interface Streamlines Workflow Design



The PipeWorks Client interface (above) can be installed on any Windows computer in your lab (or at home), and provides easy access to search and analysis tools, stored input objects and databases installed on the CodeQuest or DeCypher system running the PipeWorks Server. Just drag the appropriate search tools to the workflow space, connect them with filters and analysis tools, then launch the workflow. Visual indicators display workflow status and projected completion time for each step.

PipeWorks' data and workflow sharing enable more effective research collaboration, and the visual workflow streamlines the design of analyses for your entire team. In comparison to command line bioinformatics applications, PipeWorks is much easier to learn and use on a daily basis.

Advanced Pipeline Development for Bioinformaticists

Build once, deploy to many, maximize productivity

- Design and deploy workflows that can be updated and executed without your involvement, freeing you from repetitive tasks.
- Researchers can add new input sequences, experiment with search parameters and databases, schedule their own workflows, and retrieve & analyze their own results without bioinformatics support.

Extend workflows with custom scripts or 3rd party applications

- Workflows may incorporate EMBOSS tools like Pepinfo for hydrophobicity plots and residue histograms.
- Additional EMBOSS tools include CpGplot, DotMatcher, CompSeq, Prophecy, ShowSeq and OddComp (www.ebi.ac.uk/emboss).
- To extend workflows with additional tools—such as open source programs for next-generation sequence assembly—PipeWorks' Script Executor Tool enables you to incorporate Perl or Python scripts into any workflow.
- To share custom tools with all PipeWorks users, use the Generic Wrapper to register new tools within the PipeWorks environment.
- The PipeWorks API helps you define XML inputs, outputs and data translators that are required to extend workflows with your own code.

System Requirements & Details

- PipeWorks 1.0 (Cat. No. 75150)—including the server and 5 supported clients requires a CodeQuest workstation or DeCypher server running Windows. CodeQuest Z1 and Z2 include 1 or 2 SeqCruncher PCIe accelerators, respectively (Cat. Nos. 75008 & 75016).
- The PipeWorks Client application may be installed on any laboratory computer, enabling you to conveniently build, launch and monitor your workflows remotely. Additional 5-named-user license contracts may be purchased separately (Cat. No. 75160).
- The PipeWorks Client requires a computer running Windows XP or Server2003, a minimum of 512 MB RAM, 10 GB storage and a network connection to the TimeLogic system running PipeWorks Server.
- PipeWorks is free of the limitations of other commercial pipelining systems, and offers easier implemention than open source toolkits, which must be tied to clustered computing environments.

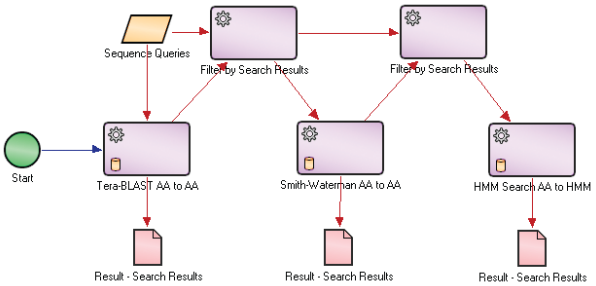
TimeLogic®
biocomputing solutions

PipeWorks™

Sample Workflows

Connect Sequential Searches for More Powerful Annotation

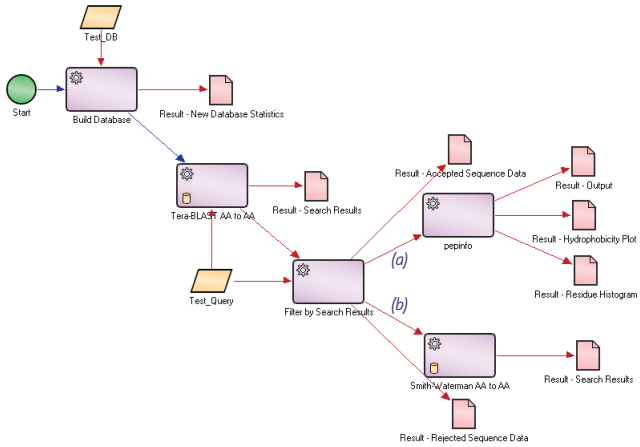
PipeWorks includes filters that enable you to route result data based upon percent identity, e-value, bitscore, gap percentage or score count. In the example below, input sequences are initially processed with a Tera-BLAST comparison. Sequences that generate no hits—or score below your filter threshold—are compared to Swiss-Prot using Smith-Waterman (S-W). Sequences that are not confidently identified by BLAST or S-W are routed to an HMM search that links them to a specific protein family or superfamily.



At each search step, you can view multiple results (extracted sequences, alignments, benchmark timing and log files) to help optimize your workflows. You can also merge multiple *extracted sequence results* to serve as the input for additional searches.

Search, Filter and Analyze

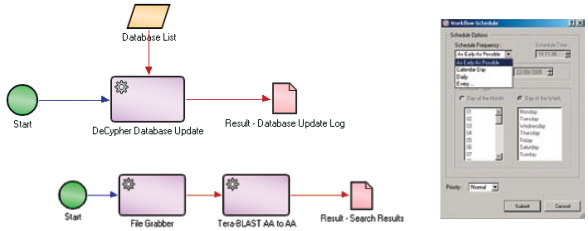
Use Tera-BLASTP to compare new protein sequences to a custom protein database, then filter the results by e-value so that high-confidence hits are processed by the EMBOSS tool Pepinfo *(a)*. Pepinfo's hydrophobicity plots and residue histograms will be displayed for each hit.



Results that score below the filter threshold *(b)* are routed to a Smith-Waterman search, which is optimal when comparing sequences with lower shared identity.

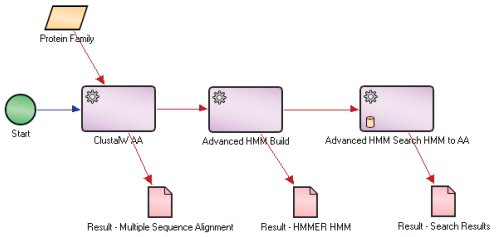
Automate and Schedule Repetitive Tasks

PipeWorks' Database Updater workflow downloads database sources from one or more FTP sites, then formats them for local searching. Workflows can also employ the File Grabber to retrieve a remote set of query sequences. All workflows can be scheduled by date or frequency.



Build Hidden Markov Models to Expand Protein Families

Run ClustalW to generate a multiple sequence alignment from a protein family, then build a hidden Markov model (HMM) for searching UniProt. This workflow helps you identify evolutionarily distant family members.



The Power to Design and Process Genomics Workflows is on your Desktop

The unique combination of PipeWorks and TimeLogic's SeqCruncher solutions enables you to analyze next-gen sequencing output, process high-throughput EST and protein annotation workflows, and run HMM-heavy metagenomics projects. Your entire lab can design and process workflows on a resource-friendly desktop system to advance their genomics projects or drug discovery efforts.

TimeLogic®
biocomputing solutions