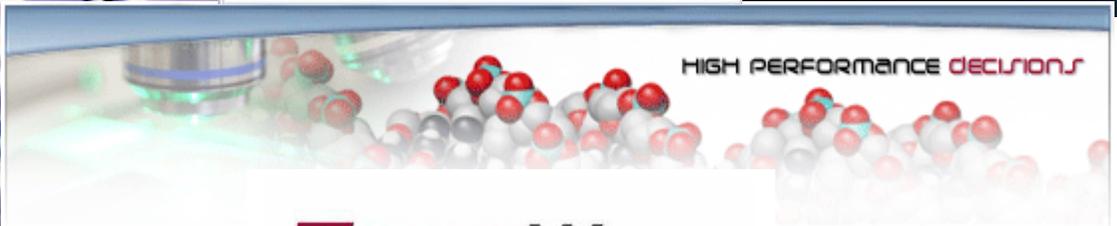
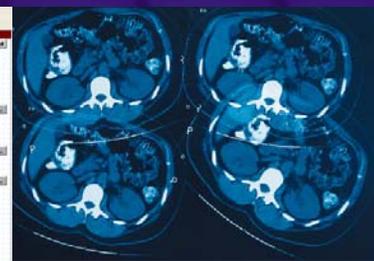
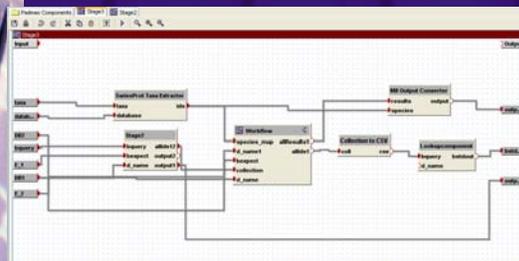
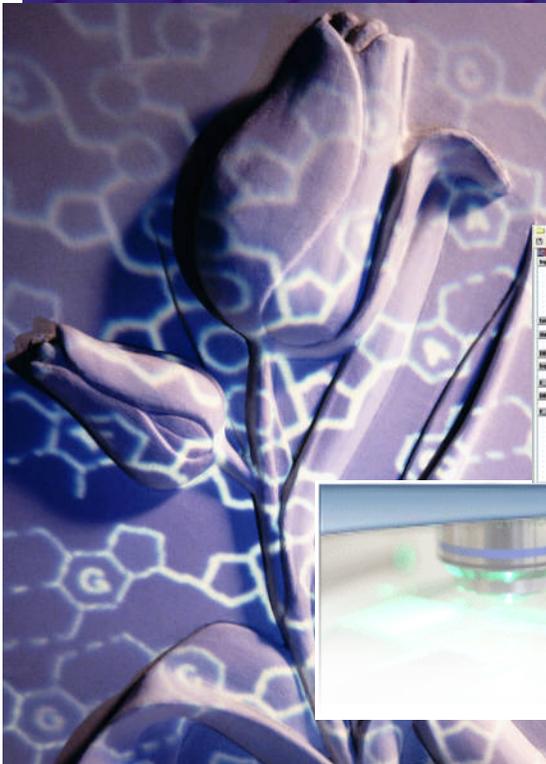




TurboWorx High Performance Decision Making for Life Sciences



TURBOWORX

TurboWorx Solutions for the Life Sciences

Accelerated Discovery Leads to Better Decision Making and New Capabilities

TurboWorx Solutions Value

- Improve application performance and automation, allowing for more sophisticated and complex discovery.
- Help accelerate and manage the use of compute and storage resources across the organization, optimizing the use of existing hardware and reducing infrastructure operating costs.
- Offer an affordable, flexible solution that supports and integrates with a wide range of packaged and custom applications.

The Discovery Process and Computing Challenges

The discovery process depends on pipelining genomic, proteomic and other types of data through diverse applications. Flowing data through applications requires researchers to spend valuable time cutting and pasting, or writing single use scripts. In addition, IT resources are constantly juggled to meet growing demands to process more information and accommodate changing business priorities.

Gene-sequence alignment and pattern discovery are examples of high speed, complex problems that compare characters in sequences or search for matching phrases and terms. Molecular dynamics, protein folding, and metabolic systems modeling are computationally intense and often built around complex algorithms borrowed from physical chemistry or quantum mechanics. In addition, a large majority of problems in the life sciences are not only computationally intense, but also involve moving large amounts of data from diverse sources and locations through a complex series of calculations.

Automated Workflows in Diverse Clusters and Grids

TurboWorx Enterprise™ provides life science companies with an efficient and flexible way to automate and accelerate applications to achieve discovery faster. TurboWorx Builder™ links together the complexities of bioinformatics applications to construct pipelined workflows. Without writing time-consuming, single-use scripts, applications are “wrapped” to become interchangeable components that make it easy for scientists to build, manage, and change workflows as research requirements evolve.

Because problems in life science research usually involves passing large data sets from diverse sources through complex sequences of calculations, any viable solution must include parallel execution and workflow management. TurboWorx's SmartGrid™ allows parallelism and iterative processes to be designed into the workflow.

At run-time, workflow tasks are executed in parallel dividing problems into a large number of individual calculations, regardless of operating system or platform. The TurboWorx Hub distributes dynamically all workload tasks among all of the available machines in systems. TurboWorx Enterprise supports fault tolerance for the TurboWorx Hub and nodes, ensuring the integrity of the entire system and no lost work. Workflows can be executed from the TurboWorx Web interface or Builder application.

TurboWorx Workflow Computing for Life Sciences

TurboWorx products include a set of proprietary software technologies for:

- Enterprise-wide management of distributed computational resources, applications, and data to deliver maximum performance and throughput for high-performance applications.
- Creation, management, and execution of compute- and data-intensive application workflows in distributed computing environments including clusters and grids.
- Construction of pipelined workflows integrating diverse collections of complex applications.
- Automated high-performance execution of workflows in highly heterogeneous environments including corporate networks and wide-area computing grids.
- Comprehensive management of compute clusters, other computational resources, and data.

TurboWorx Bioinformatics Workflow Example

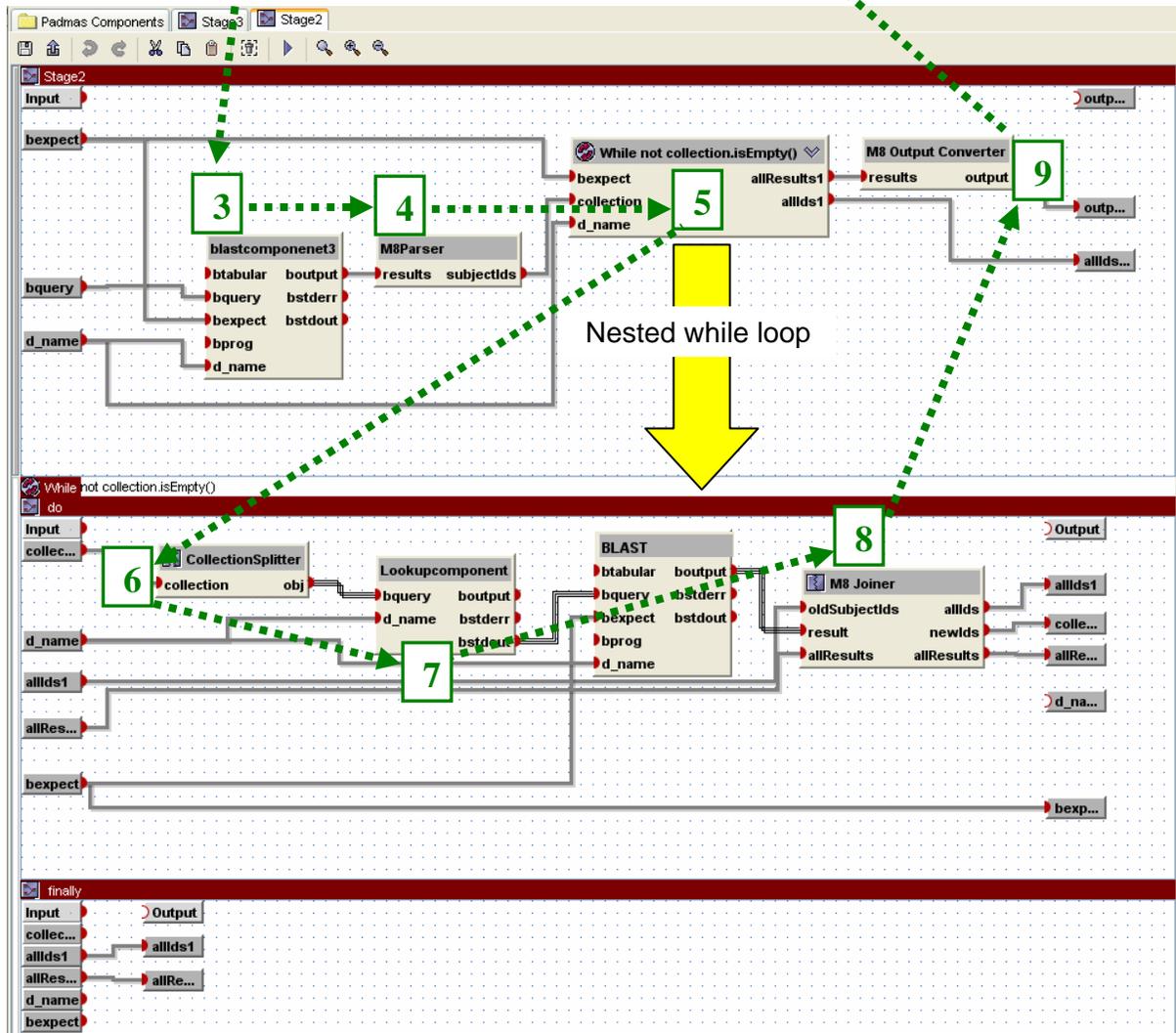
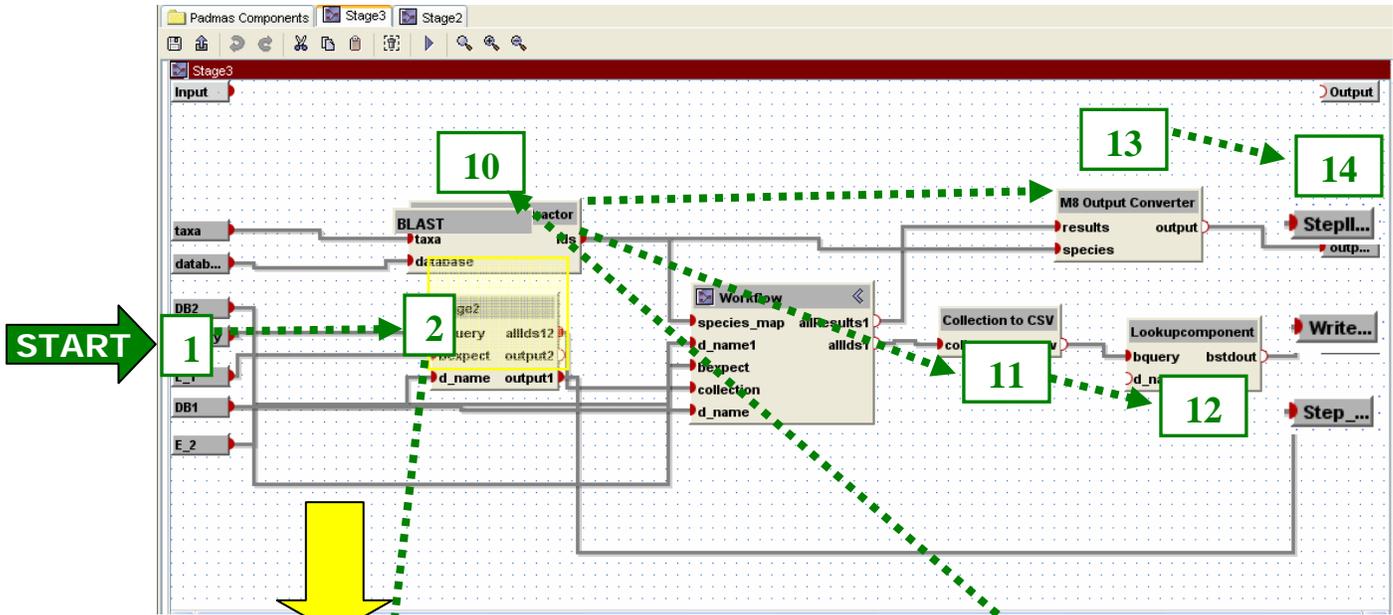
The workflow shown on the next page is an example of a typical use case scenario that consists of an iterative sequence comparison and BLAST sequencing of genes from Arabidopsis.

The object of this workflow is to:

- Input a peptide sequence and e-value cutoff
- Compare the sequences to the Arabidopsis thaliana peptide collection
- Compare the hits to Swiss-Prot plant sequences
- Retrieve the results and output them to a display table and a FASTA file.

The workflow shown on the next page was created with TurboWorx Builder to meet the above objectives. The individual steps are described below and match the green numbers the next page.

- Step 1.** The database is queried.
- Step 2.** The output of this nested component is piped into the parser Java component.
- Step 3.** The BLAST component takes a peptide sequence and e-value cut-off and compares it to Arabidopsis database. The output is the resulting BLAST hits that are piped into the M8 Parser component inputs.
- Step 4.** The M8 Parser component parses an NCBI Blastall -m 8 result, extracting all subjectids. The output consists of all sequences with subjectids.
- Step 5.** Nested While Loop component is a workflow consisting of three subcomponent that split the input sequences, BLAST it, and then join the results into a single output.
- Step 6.** The collection splitter splits a java.util.Collection into individual elements. Output consists of individual elements of the collection that are pipelined into the Lookupcomponent.
- Step 7.** The Lookupcomponent takes individual element and looks it up for the given database and then outputs the data to the M8joiner component.
- Step 8.** The m8joiner component combines all of the results into a single output and sends it to the Formatconvert component.
- Step 9.** The formatconvert component displays the results in table with columns: query seq id, hit seq id, e-value and score and converts the results to table format
- Step 10.** Swissprot Taxa Extractor component takes a database name and specific taxa information to extract the seq ids for that particular taxa. The output consists of a map whose keys are the valid sequence ids and values and are the corresponding species names.
- Step 11.** Collection to CSV component takes a collection of all subject ids and outputs a comma-separated string. The output is a comma-separated string.
- Step 12.** The 2nd Lookup component takes a comma separated string and the database name and sends as output a FASTA file containing all swissprot sequences, which belong to the selected taxa and matched the specified sequence of earlier BLASTs.
- Step 13.** The 2nd M8 Output Converter component takes all the results and retrieve those for a given taxa.



The bottom workflow is a dependent sub-component of the top workflow.