



**I N C O G E N**

**VIBE**

**Visual Integrated Bioinformatics Environment**

Whitepaper

**Enter the Visual Age of Computational Genomics**

**INCOGEN, Inc.**  
104 George Perry  
Williamsburg, VA 23185

[www.incogen.com](http://www.incogen.com)  
Phone: 757-221-0550  
[info@incogen.com](mailto:info@incogen.com)



## Introduction

The Visual Integrated Bioinformatics Environment (VIBE) is a cross-platform, visual programming interface that provides user-friendly access to many common bioinformatics algorithms on various platforms. The graphical interface provides an extensible drag-and-drop environment for the creation of analysis pipelines and visualization of results. The VIBE Software Development Kit (SDK) provides extensibility through an integrated Application Programming Interface (API) that allows users to incorporate in-house and third-party tools into a single environment.

VIBE provides a powerful, integrated package for workflow creation and management in a high-throughput environment that allows user-level integration of custom software packages into the workflow definition process. The drag-and-drop analysis pipeline construction makes VIBE an intuitive and highly visual tool to automatically generate, edit, manipulate, and execute pipelines. It also provides users with the capability to save and recall pipelines for later execution through the interface or for merging with other workflows. The system allows real-time monitoring of pipeline execution statistics and results for active and completed analyses.

VIBE's standard toolkit provides users with an interactive sequence analysis workbench and state-of-the-art visualization and querying tools for efficient data mining. The intuitive operation and visual interface allow researchers with limited bioinformatics background to take full advantage of powerful analysis resources available for life science research.

The software can be deployed using one of two architectures: client/server (Workgroup Edition) or client-only (Desktop Edition). The multi-tiered design of the software allows the processing components to be isolated from the business logic and storage format of the data. The Workgroup Edition can be scaled and tuned within the middle layer to significantly improve performance with few or no changes to the clients while the Desktop Edition provides simple setup and configuration.

### *What can VIBE do for you?*

- provide users with an **extensible** framework for bioinformatics analysis
- **increase efficiency** for bioinformaticists by providing an alternative to manually creating analysis scripts and data management for a **high-throughput** environment
- provide researchers with **intuitive, user-friendly** access to data analysis and mining.

## Providing a Bioinformatics Analysis Framework

The wide array of analysis tools and platforms available for genomic data analysis provides an unprecedented potential for discovery. The heterogeneity of users, tools, and data creates complex challenges such as managing high-throughput data analysis infrastructures, multi-user and multi-platform environments, providing sufficient computational resources, and performing these tasks on time and within budget constraints.

The ratio of bioinformaticists to biologists in biotechnology companies and laboratories ranges from one bioinformaticist per five biologists to one per 20 or more. Given that a major fraction of bioinformaticists' time and responsibilities is devoted to supporting the analysis needs of biologists, it is clear that individual, custom support creates a significant bottleneck in the discovery process. This bottleneck greatly impairs the ability to fully exploit the vast amounts of data that are produced and available for genomic discovery. The Visual Integrated Bioinformatics Environment provides two aspects to significantly reduce the analysis bottleneck.

### **Aspect 1: Benefits of VIBE for Bioinformaticist Users**

The first aspect consists of providing bioinformaticists with a highly efficient method to perform tasks associated with constructing high-throughput analysis and data mining pipelines as efficiently as possible. Expert users working in a high-throughput environment will benefit from dramatically decreased programming time. In addition, with the explosion of the amount and heterogeneity of genomic data, bioinformaticists are often overwhelmed (and overworked!) by the tasks of providing individual support to biologists, such as performing specific analyses or developing custom analysis pipelines. By providing biologists with the ability to construct their own pipelines (see Aspect 2 below) bioinformaticists are able to devote more time to developing novel approaches and to answering important biological questions, rather than performing highly repetitive and time-consuming tasks. The Director of Discovery at a pharmaceutical company that employs VIBE in their daily activities attests that the platform enables one bioinformaticist to support 100 of their researchers effectively.

## Aspect 2: Benefits of VIBE for Biologist Users

The second aspect to reduce the analysis bottleneck consists of providing researchers who have limited computational experience with a user-friendly and intuitive sequence analysis workbench. VIBE provides an interface for biologists to perform their own analyses. This functionality not only serves to increase the accessibility of powerful computational resources to a more general audience, but also provides biologists with a common language for communicating and exchanging sequence analysis research among their peers. Users can make notes on individual modules or on a pipeline as a whole, and pipelines can be emailed directly from VIBE as an attachment to a colleague, who can then open the pipeline in their install of VIBE. Functionalities such as these clearly indicate that VIBE was designed with the biologist end-user in mind.

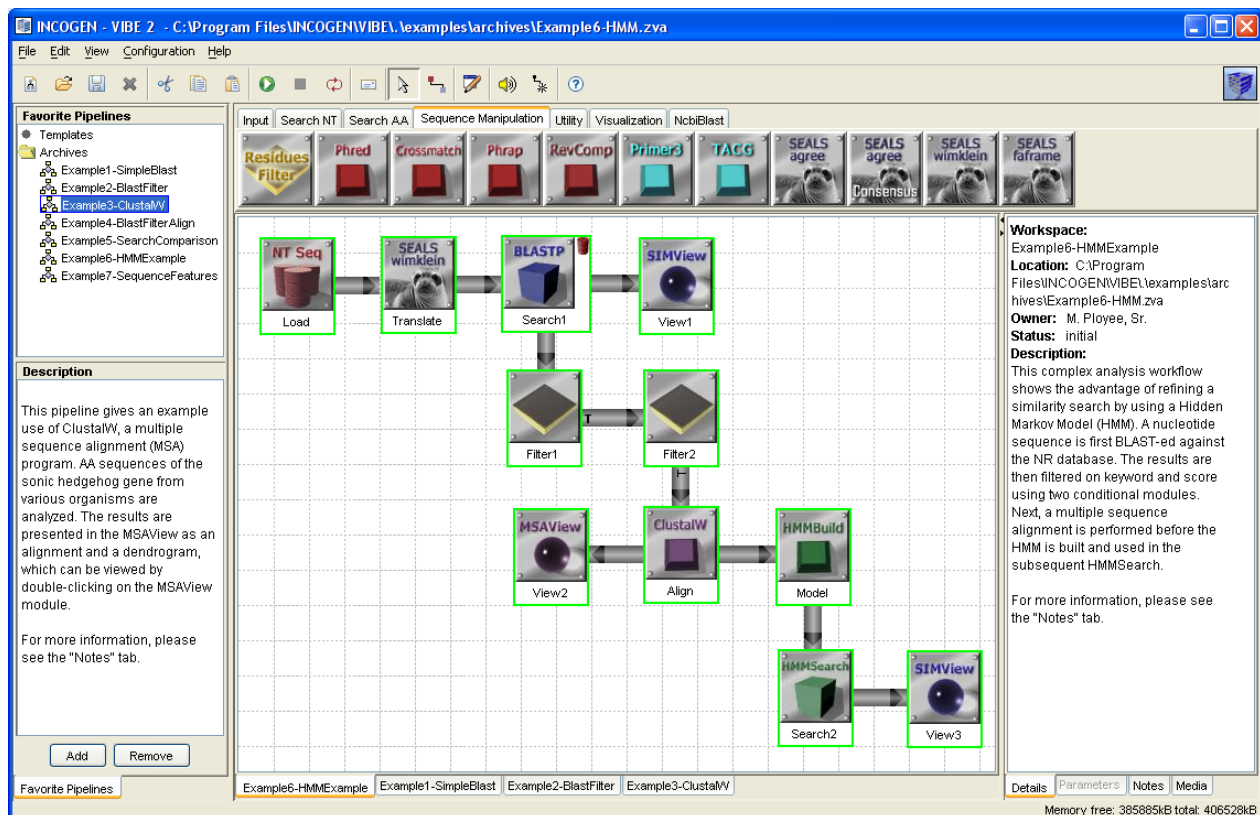
### *What about your existing resources?*

- VIBE **increases utilization** of IT resources
- **Existing** scripts and tools can be **integrated** into VIBE

In addition to reducing the obstacles in the discovery process, VIBE allows organizations to leverage their existing resources and processes. The investment in IT infrastructure will show higher return through increased utilization once VIBE Workgroup is deployed. VIBE Workgroup allows many researchers to share the same server resources to perform analyses without the need for user-level command-line access or individual accounts on compute servers. Additionally, the investment in coordinating processes via Perl scripts or other “glue” mechanisms can be enveloped in a module in VIBE using the SDK, allowing the existing, in-house techniques to become graphically accessible to a wide range of researchers and to be used in larger scale workflows as well.

## VIBE – Detailed Description

The following sections provide a detailed description of the operation of the program, the analysis tools, and a description of the architecture of the program.

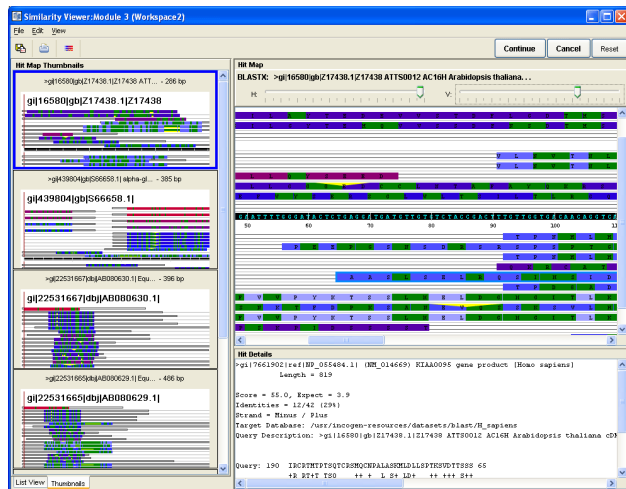


**Figure 1:** VIBE screenshot - the icons represent modules that can be dragged onto the main canvas to construct pipelines. The panel on the right-hand side is used to change parameters for the analysis modules, access detailed help about each module (shown), or make notebook entries about the pipeline or individual analysis modules. The panel on the left-hand side provides quick access to favorite VIBE pipelines or other local programs.

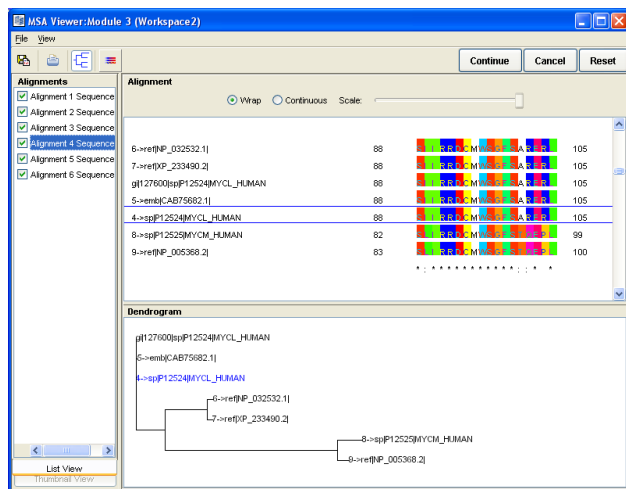


## Description of Operation

VIBE provides a graphical drag-and-drop interface to create analysis pipelines (Figure 1) from a wide selection of tools and algorithms. The program also provides state-of-the-art visualization tools for analysis results. Analysis modules are grouped by type and presented to users as icons on a toolbar. Users can choose among sets of functionality representing: data input, similarity searches, sequence alignment, utilities such as email notification agents and data filters, model building and searching, and visualization tools. (Please see the next section for a listing of all tools in the standard sequence analysis toolkit.) Each analysis module contains a set of default parameters and may be executed with the default settings. The parameters can also be easily adjusted through a tabular interface. The program provides detailed descriptions in hypertext format for all modules.



**Figure 2a:** An example screenshot of SimView, the similarity search viewer. The viewer presents a color-coded, intuitive representation of similarity search results and provides users with interactive features such as zooming, mouse-over results viewing and active links to GenBank or other local or remote repositories to obtain further details about a particular hit.



**Figure 2b:** An example screenshot of MSAView, the multiple sequence alignment viewer. The viewer provides color-coded, scalable alignment views as well as a linked dendrogram representation of the alignment. All viewers provide printing and exporting capabilities, both as text reports and as images.

The icons can be dragged onto the workspace and connected to other analysis modules to generate a pipeline flow for data analysis. VIBE provides design-time checking to assist users in creating valid workflows and to reduce the probability of a runtime error or conflict. Users can also ask VIBE for a list of modules which can come next in a workflow based on the modules currently placed on the workspace, select the modules of interest, and have VIBE automatically place and connect the selections.

Pipelines can be saved as XML on the client computer or on any network-accessible machine. A pipeline can be saved before execution as a *template* (that is, with no data associated with it) and used later with other input data sets or it can be saved after execution as an *archive* to capture all associated data and results. The user can re-open the saved pipeline at any later point and view the saved results or conduct further analysis. Multiple workspaces allow users to design new pipelines while continuing to monitor the progress of active pipelines.

Through the simple graphical interface, users may employ tools such as alert modules and data filter modules to diverge data flow to customize each pipeline to perform the exact analysis required.

State-of-the-art, interactive visualization tools are available for each analysis module to efficiently present the user with the most important results of each analysis. Figure 2a shows an example of the similarity search result viewer. Figure 2b presents an example of the multiple sequence alignment (MSA) viewer.

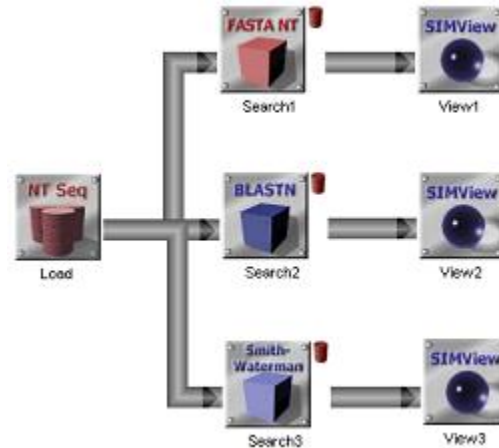
VIBE is an intuitive and elegant platform for data pipelining. To further reduce the startup time for learning a new software tool, VIBE includes video tutorials on its core functionality that are accessible directly through the interface. Users create workflows as the videos are playing, following along and learning the standard interaction patterns with the software as well as tips and tricks for getting more from the system.



## Analysis Tools Provided with VIBE

The standard VIBE distribution includes the Sequence Analysis Toolkit, which provides a rich set of sequence analysis tools. It consists of a set of algorithmic, utility and visualization programs. This package includes:

- Data source selectors for defining and importing query sequences, chromatograms, and Hidden Markov Models
- Base calling (Phred\*) and fragment assembly (Phrap\*)
- Similarity searches:
  - Smith-Waterman
  - BLASTN, BLASTX, BLASTP
  - BLATN, BLATP\*
  - FASTA, FASTX, FASTY
  - TBLASTN, TBLASTX
  - TFASTX, TFASTY
- Multiple sequence alignment:
  - ClustalW
  - SEALS Agree (consensus sequence building)
- Sequence manipulation utilities
  - Reverse Complement
  - Crossmatch\*
  - Residue filtering
- Restriction enzyme analysis (TACG)
- Primer Design (primer3)
- ORF detection and frame translation (SEALS Wimklein, SEALS Faframe, FAView).
- Hidden Markov Models:
  - HMM Build
  - HMM Scan (search protein query against model database)
  - HMM Search (search model against protein database)



- Protein analysis
  - Wisconsin Package/GCG PeptideSort\*
  - Wisconsin Package/GCG PepPlot\*
  - Wisconsin Package/GCG TransMem\*
- Alert agents (Email agents for data mining, offline status reports, or warning messages)
- Conditional filters to remove, redirect, or diverge data flow dynamically based on analysis results
- Visualization tools to provide dynamic interaction, flexibility and efficient visualization of the most significant analysis results. Viewers also allow users to manually select which data items will continue in the workflow, as well as to print or export the results. Tools such as the following are available:
  - SimView (displays results of similarity searches)
  - MSASView (sequence alignment results)
  - FAView (sequence manipulation with primer and restriction enzyme mapping)
  - TraceView (browse chromatograms)
  - PhrapEdit (allows viewing and editing of assemblies from Phrap)
  - ImageView (displays images, such as from the Wisconsin Package's PepPlot)
  - TextView (displays text reports from any analysis module)

*\*Note: The list above specifies many of the tools with which VIBE is capable of interacting; some of the tools require a separate license from their respective owners.*





## VIBE Enterprise Architecture

The design of VIBE incorporates Java 2 Enterprise Edition (J2EE) object-oriented architecture standards. These standards yield a robust and flexible multi-tiered system (Figure 3). The modular and layered design allows VIBE to be configured in either *Desktop* (single-user) or *Workgroup* (client/server) configurations.

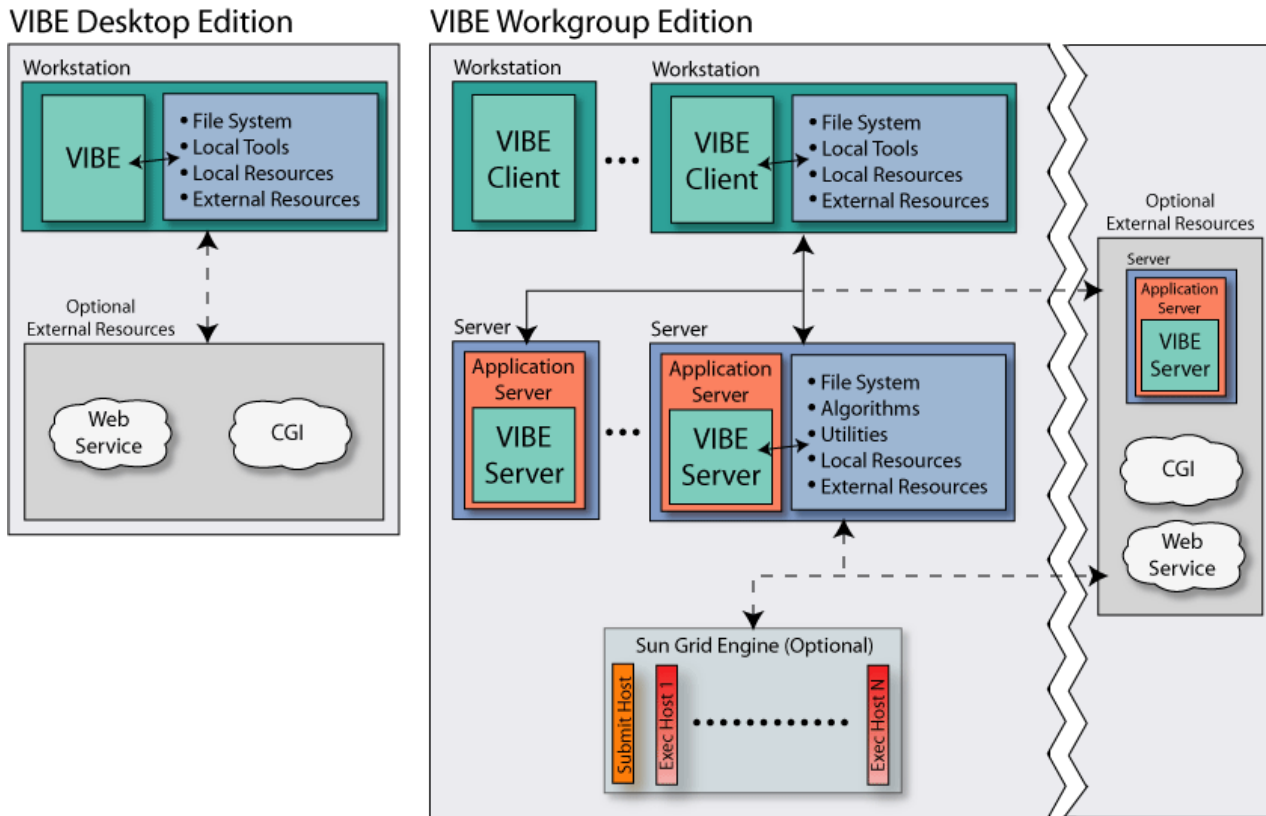


Figure 3: Diagram of the VIBE multi-tiered system architecture.

INCOGEN has leveraged the multi-tiered enterprise architecture to provide many design advantages, such as:

- reduced overall system cost without loss of performance or flexibility through the implementation of a thin client layer
- isolation of both the data presentation and algorithmic processing from the business logic and storage format, allowing flexibility in the middle layers of the system without modifications to the user interface or server component
- ability to distribute the server and/or algorithms to increase system performance
- independent analysis on the application server layer without client intervention



## VIBE Software Development Kit (SDK)

The central feature of VIBE is the ability to integrate any third-party or in-house tool, regardless of vendor, into one easy-to-use environment. The VIBE SDK exposes an integration Application Programming Interface (iAPI) to the system via a few succinct Java classes and their methods, as well as through extensive documentation and guidelines for using the SDK. Frequently, modules can be integrated into the environment by adding only an XML file to describe how to use the tool it represents, regardless of the language in which the tool was written. The SDK provides mechanisms for adding tools that are executed locally on the client's machine, that are executed remotely through one or more VIBE servers, and that are accessible via a web-enabled interface such as SOAP or CGI. It also provides the ability to add visualization tools or process utilities for execution within the VIBE client interface itself. VIBE SDK adds value to your current environment by providing four critical factors:

- **distributedability:** ability to deploy a new service into any number of VIBE servers, and a new module into any number of VIBE clients
- **flexibility:** ability to integrate any tool that can be accessed programmatically
- **simplicity:** succinct and effective iAPI minimizes overhead in integration of new tools
- **efficiency:** platform that supports the “write once, integrate everywhere” paradigm

## VIBE vs. Traditional Methods of Sequence Analysis

Types of sequence analysis can be separated into high-throughput applications, typically performed through developing Perl codes; and web-based, low-throughput applications that offer users access to one tool at a time. While web-based sequence analysis portals offer a starting point for combining multiple tools and provide users with a degree of automation, issues such as throughput and response time limitations, data security (for off-site services), and limited browser-based viewers may still render them as unacceptable solutions for many users and companies.

The table on the next page offers a comparison of sequence analyses performed through VIBE vs. traditional methods (web-based interface and scripts).

Capability	Web	Scripts	VIBE
Required programming expertise	None	Extensive	None
Data throughput	Light, bandwidth typically a bottleneck	Light-high	Light-high
Access to previous data	None	Scripts	Graphical, XML, Text
Access to previous analysis	Cut-and-paste or email	Custom	XML (templates and archives), email, cut-and-paste
Visualization	Static / image based	Scripts	Dynamic / interactive
Integration with existing infrastructure	None	Manual / custom	Can integrate with existing software and OTS packages.
User friendly	Yes	No	Yes
Application maintenance and upkeep	None	Extensive	Minimal, with optional technical and onsite support
Data management	None	Manual	Automatic and configurable. Import and export utilities
Development time/cost	Small	Extensive	Small

**Table 1:** Comparison of VIBE to traditional methods of sequence analysis



## Summary

VIBE combines the intuitive, point-and-click interface of web-based tools, the high-throughput capacities of traditional scripted analysis pipelines, and powerful built-in data mining functionalities. It provides a powerful, extensible framework that serves to reduce analysis bottlenecks.

The cost of analysis is determined in part by the cost incurred from generation of the code to perform the analysis. At a cost below the average annual cost of a bioinformaticist, VIBE is an excellent and cost-effective investment because it dramatically improves productivity for all bioinformaticists by reducing programming time. Furthermore, VIBE allows biologists to interactively perform their own customized analyses, thereby reducing the number of custom analysis pipelines that need to be manually designed by bioinformaticists. This allows bioinformaticists to concentrate on more cost-effective and higher-return tasks such as developing new approaches to address biologically significant questions rather than providing repetitive and low-return services such as manually creating analysis pipelines.

Given the complexity, heterogeneity, and vast amounts of life science data, bioinformaticists and biologists must be given tools to maximize their efficiency and ability to fully exploit the potential of the data.

VIBE provides a state-of-the-art development and visualization environment to provide researchers with a cost-effective, integrated solution for knowledge discovery.

## Additional VIBE Information

Visit <http://www.incogen.com/vibe> to view a short multimedia overview of the software and browse additional documentation. Look for this image:

## Contact Information

**INCOGEN, Inc.**  
104 George Perry  
Williamsburg, VA 23188

[www.incogen.com](http://www.incogen.com)  
Phone: 757-221-0550  
[info@incogen.com](mailto:info@incogen.com)