BaseBuild Coders, On Demand



Created by David McKillen, Basebuild, Inc.

Client Overview

The Marine Genomics project, as detailed in the BMC Genomics article, aimed to create a centralized platform for curating and analyzing Expressed Sequence Tags (ESTs) and microarray data specific to marine organisms. Recognizing the fragmented nature of marine genomic data, the project sought to provide researchers with a unified, accessible repository to facilitate comprehensive studies on marine species' transcriptomic responses to environmental stresses.

The Challenge

Marine genomic research faced significant hurdles due to the dispersed and inconsistent nature of available data. Researchers lacked a centralized system to store, process, and analyze EST and microarray data, leading to inefficiencies and potential redundancies in research efforts. The absence of standardized tools for data submission, quality control, and analysis further compounded these challenges, hindering collaborative efforts and the advancement of marine genomics.

Our Solution

To address these issues, the Marine Genomics project developed a web-based platform utilizing PHP for the user interface, Apache as the web server, and PostgreSQL for database

Case Study: BMC Genomics Data Warehouse

management. MATLAB was integrated for statistical analyses and computational processes. The platform offered tools for automated sequence trimming, quality control, BLAST searches, and facilitated direct submissions to GenBank. By supporting data from 19 species and enabling contributions from researchers worldwide, the system provided a robust infrastructure for data curation and analysis.

Results

The implementation of the Marine Genomics platform significantly enhanced the efficiency and collaboration within the marine genomics community. Researchers gained access to a centralized repository with integrated analytical tools, streamlining the process of studying marine organisms' genomic responses to environmental factors. The platform's success is evidenced by its adoption by researchers across various continents and its role in facilitating the submission of curated data to public databases like GenBank, thereby advancing the field of marine genomics.

Tech Stack

- Frontend/User Interface: PHP
- Web Server: Apache
- Database: PostgreSQL
- Statistical Analysis & Computation: MATLAB
- Bioinformatics Tools: BLAST (for sequence alignment and identification)
- Hosting Environment: Unix-based systems (specifically FreeBSD)
- Data Submission Integration: GenBank submission pipeline

Why it matters

This case demonstrates how Basebuild can support scientific and academic institutions in processing and presenting complex datasets. Our ability to bridge backend data pipelines and interactive front-end tools is key to empowering researchers and developers alike.

