PAPER

TersectBrowser+ : A single browser for in-depth introgression analysis of re-sequenced genome datesets

Tomasz Kurowski,[1,*] Gabrielle Baumberg,[2] Gregory Lupton,[2] Tanya Stead,[2] David Oluwasusi,[2]

[1]Centre for Soil, Agrifood and Biosciences, Cranfield University, College Road, MK43 0AL, Cranfield, UK,[2]Department, Organization, Street, Postcode, State, Country,[3]Department, Organization, Street, Postcode, State, Country,[4]Department, Organization, Street, Postcode, State, Country,

*Corresponding author. email-id.com

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year;

Abstract

The field of cultivar genomics has evolved rapidly, with growing emphasis on mapping introgressions—segments of DNA from wild plant species that have been incorporated into cultivated lines. A variety of bioinformatic tools aim to interpret this expanding complexity of plant genomic data, predict introgressions, and identify commercially relevant regions for future breeding. Demand is increasing for tailored tools to interactively explore these complex genomic queries in real time. Created in 2022, the Web-based Tersect Browser can dynamically display introgression patterns by applying flexible set theoretical expressions to sets of sequence variant data. Our 2025 development of Tersect Browser offers multiple front-end extension capabilities so that the user can view and analyze data without leaving the page. We identify previously published introgressions with our tool, illustrating its potential for future research, and enriching plant breeders' understanding of crop cultivar genomics.

Key words: introgression, browser, resequenced, SNP

Introduction

Cultivar genomics has evolved rapidly, with growing emphasis on mapping introgressions, segments of DNA from wild plant species that have been incorporated into cultivated lines. A variety of bioinformatic tools aim to interpret this expanding complexity of plant genomic data, predict introgressions, and identify commercially relevant regions for future breeding (Qin et al., 2021) **?**. Demand for tailored tools to explore these complex genomic queries in real-time is increasing.

As shown previously by Kurowski Mohareb (2020)**?**, Tersect is a lightweight, high-performing, real-time genome querying command-line utility that uses its own query language and bitmap indices to interpret and apply flexible set theoretical expressions to sets of sequence variant data for tomato genomes. In comparison to other tools such as BEDOPS, BCFtools and BDETools (Quinlan Hall, 2010**?**; Danecek et al., 2011), Tersect performs from three to over a hundred times faster than BCFTools (Kurowski Mohareb, 2020)**?**.

For biologists, command-line-based tools can be cumbersome. Web-app based tools with intuitive data visualization, real-time querying, and introgression mapping provide much better user experiences and better meet the requirements for interactive exploration of large-scale variant datasets and immediate visual feedback on genomic regions of interest.

In 2022, a web-based version of Tersect, Tersect Browser, was created (Kurowski, 2023**?**) using ***Includedataset***. Tersect Browser offered multiple front-end extension capabilities that allowed user to view and analyse data without leaving the page. These include: a dynamic introgression heatmap, tree navigation of accessions, and region-focused querying of genomic variants.

The field of tomato genomics, however, is growing with increasing emphasis on mapping introgressions and higher volumes of genomic data. As the field has built on foundational work of Aflitos et al, (2014)**?**, which sequenced 84 tomato accessions and wild relatives, exposed severe genetic bottlenecks in cultivated tomatoes and laid the groundwork for modern introgression analysis, there is increasing requirements for more powerful and more capable introgression visualization and comparative analysis tools.

Recent studies have expanded on this base with increasingly sophisticated tools and datasets. Notably, Liu et al. (2023)**?** used tomato pan-genomes built from hundreds of accessions to uncover regions of structural variation and introgression hotspots, many of which affect key traits like fruit size, shape, disease resistance, and flavour. The construction of these pan-genomes enables researchers to observe variation absent in the reference genome, improving trait mapping and introgression detection.

Here, we present Tersect Browser+, an updated introgression analysis platform with expanded front-end features extension capabilities that users can view and analyze data without leaving the page. It integrates JBrowse (Skinner et al., 2009**?**; Buels et al., 2016 **?**) capabilities to capture gene models, exon-intron structures, and positional gene information for biologists inspecting tomato introgression data sets.

JBrowse is a widely used, web-based genome browser designed for fast, interactive visualization of genomic data. Developed as a successor to earlier tools like GBrowse, JBrowse enables users to view and explore large-scale genome annotations, sequence data, and variation information through a modern, dynamic interface (Skinner et al, 2009 **?**). It supports a wide variety of standard file formats, including GFF, BAM, VCF, and FASTA, and is highly extensible through plugins, making it adaptable for a range of research applications from comparative genomics to functional annotation, including introgression web-app based analysis (Buels et al., 2016**?**). Researchers often deploy JBrowse to visualize genome assemblies, transcriptomes, and variant information. Its active development and ease of integration into web servers have made it a common choice for bioinformatics platforms.

Tersect Browser+ also includes other features such as a search feature and barcode generating feature. It allows for the investigation of other crop datasets, for example Soybean [include specific species type/name]. We identify previously published introgressions with our tool, illustrating its potential for future research, and enriching plant breeders' understanding of crop cultivar genomics.

Features

Tersect Browser+ is a web application that depends on resequenced data in VCF format and uses a reference genome in FASTA format. Utilizing a MEAN stack (MongoDB, Express, Angular, and Node) architecture, it creates a webpage displaying heatmaps and phylogenetic trees for introgressions within the loaded crop dataset. This provides an interface for users to identify differences across accessions within the same species dataset. It also includes a JBrowse interface for investigating specific variants, displaying selected genes directly alongside the generated heatmap. A search feature highlights high-impact variants on the webpage, helping to distinguish the selected reference genomes. Tersect Browser+ also offers barcode generation functionality for various cultivars and specific species in the dataset.

Gene Brwoser
We will include text here on the feature. I will look through the Technical Documentation to put this together.

Gene Model
We will include text here on the feature. I will look through the Technical Documentation to put this together.

Search Features
We will include text here on the feature. I will look through the Technical Documentation to put this together.

`path/to_image.png`

Fig. 1. Tersect Browser+ overview. The browser provides a single way to analyse variant data without leaving the page.[More description]

Barcode Generator
We will include text here on the feature. I will look through the Technical Documentation to put this together.

Funding

A question for Tom - we see this included in many papers published in this journal. What should we include?

Code availability

We will add in a link to the github reference in here and a link to the technical documentation.