PDF Annotation and Data Harmonization (PANDHA)

May 2022
Version 1.1.0

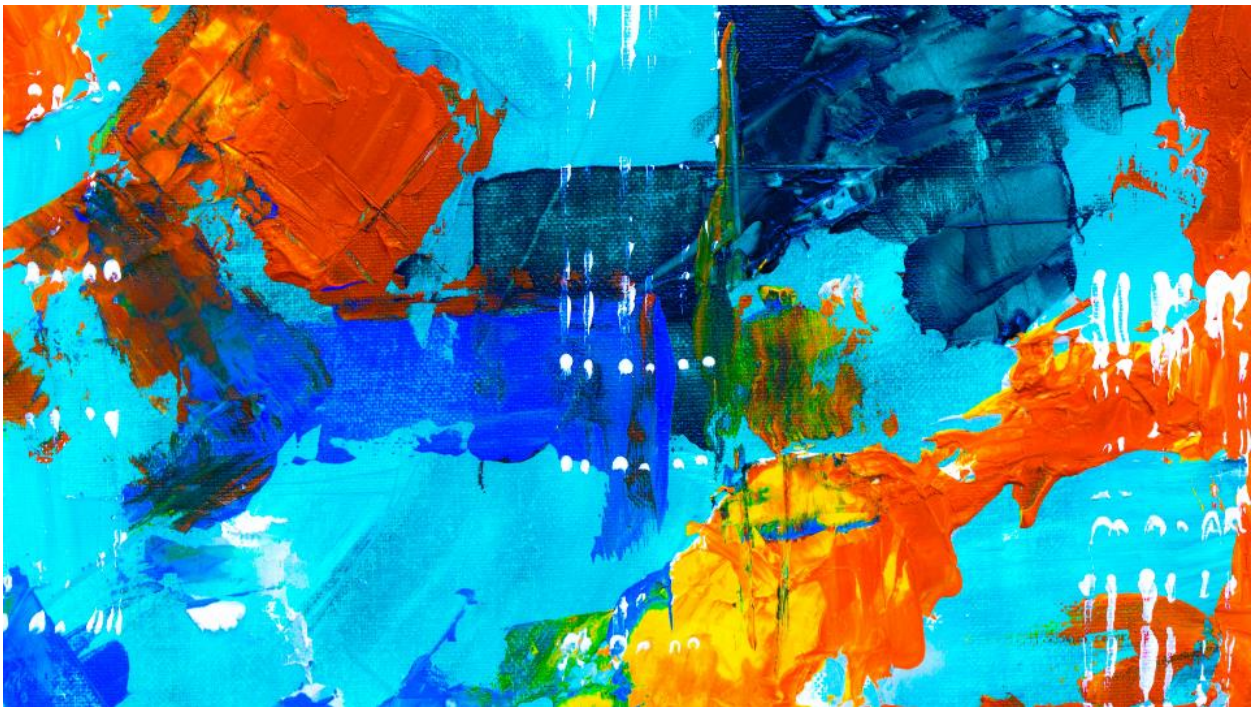


Table of Contents

Document Revision

Date	Version Number	Document Changes
01/05/2021	1.0	Initial draft
07/20/2021	1.0.1	Updated to beta PANDHA V1
09/29/2021	1.0.2	Update to beta PANDHA V1.0.2
05/25/2022	1.1.0	Update to release candidate PANDHA V1.1.0

1 Installation

1.1 Scope and Purpose

The purpose of this project is to further the research and development of tools that NCEA can use in their creation of machine-readable datasets and machine learning research. This effort consists of the following objectives:

- Research and develop software for NCEA that provides the ability to annotate scientific publications for use in machine learning algorithms. This software should be able to accept a list of tags provided by NCEA, allow the user to apply these tags to PDF documents in a web interface, and then extract out the information needed in machine-readable formats that can be used for machine learning.

1.2 Source Code Access

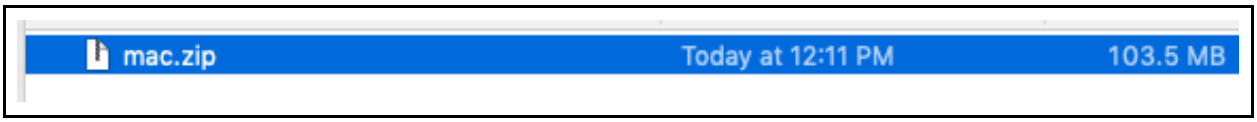
This describes the process of accessing the source code.

1. https://github.com/USEPA/pdf-data-extraction

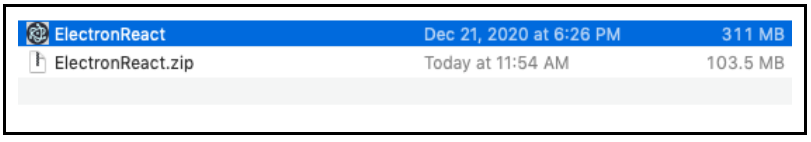
1.2 Installation MacOSX * need to work out zip issue

1. Download latest version from

2. Double click ‘mac.zip’ to unzip the file.



3. Double click ‘ElectronReact’ to launch the application.

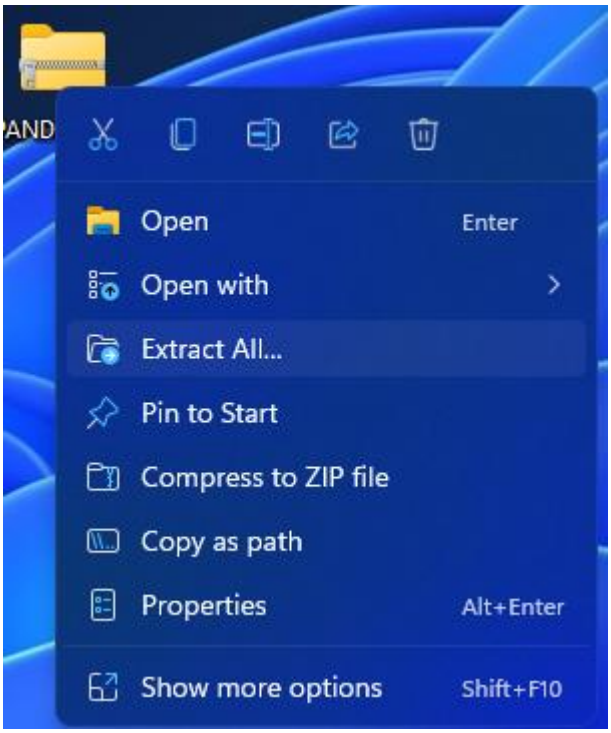


1.3 Installation Windows

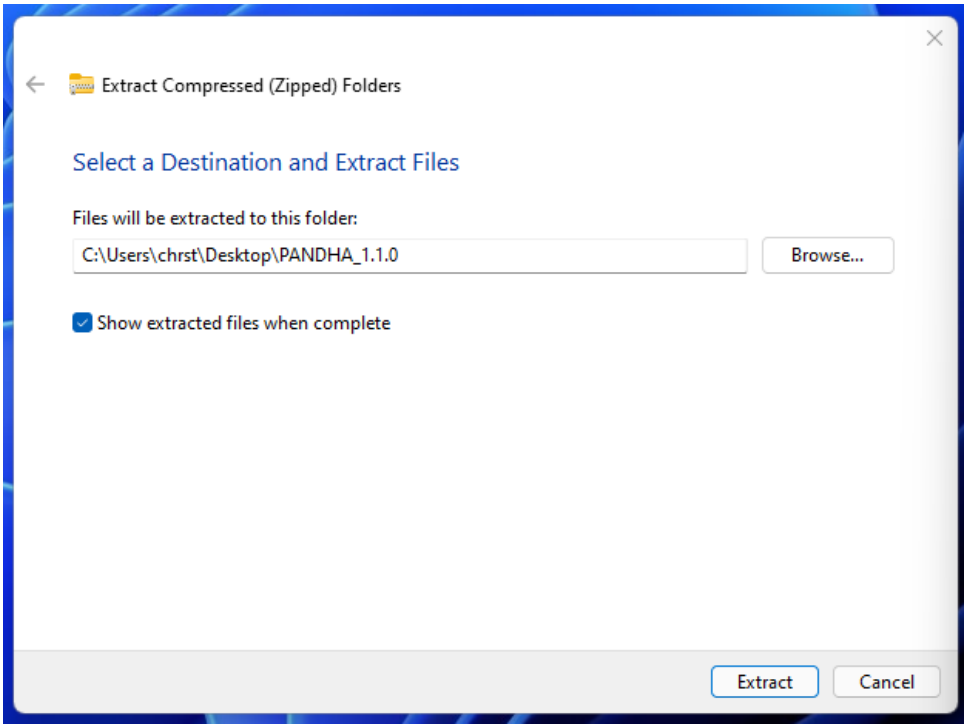
1. Download latest version from (PANDHA_1.1.0)

https://usepa.sharepoint.com/sites/DOEORNLCPAD/Shared%20Documents/General/Release%20Candidate/PANDHA_1.1.0.zip

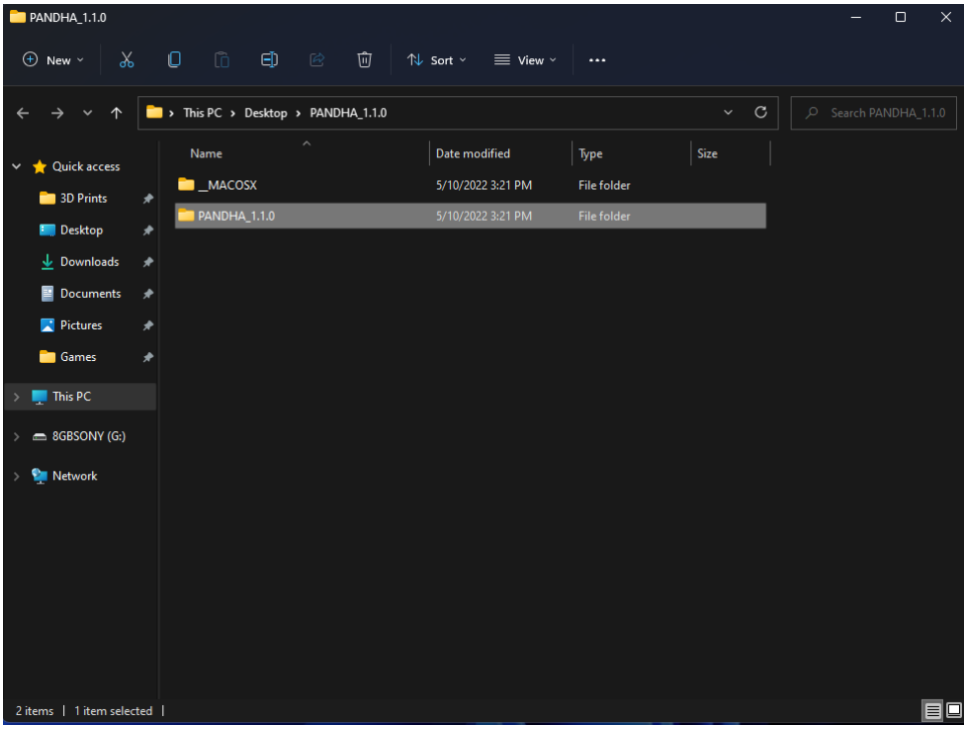
2. Right click ‘Windows.zip’ and select ‘Extract All’



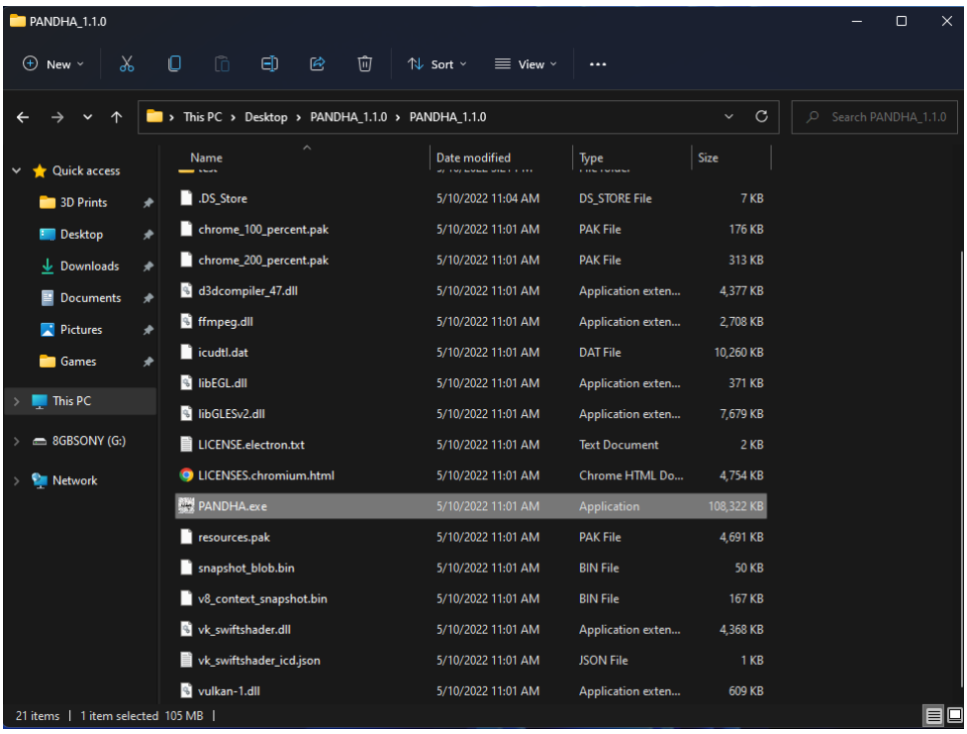
3. Select location and hit *Extract*



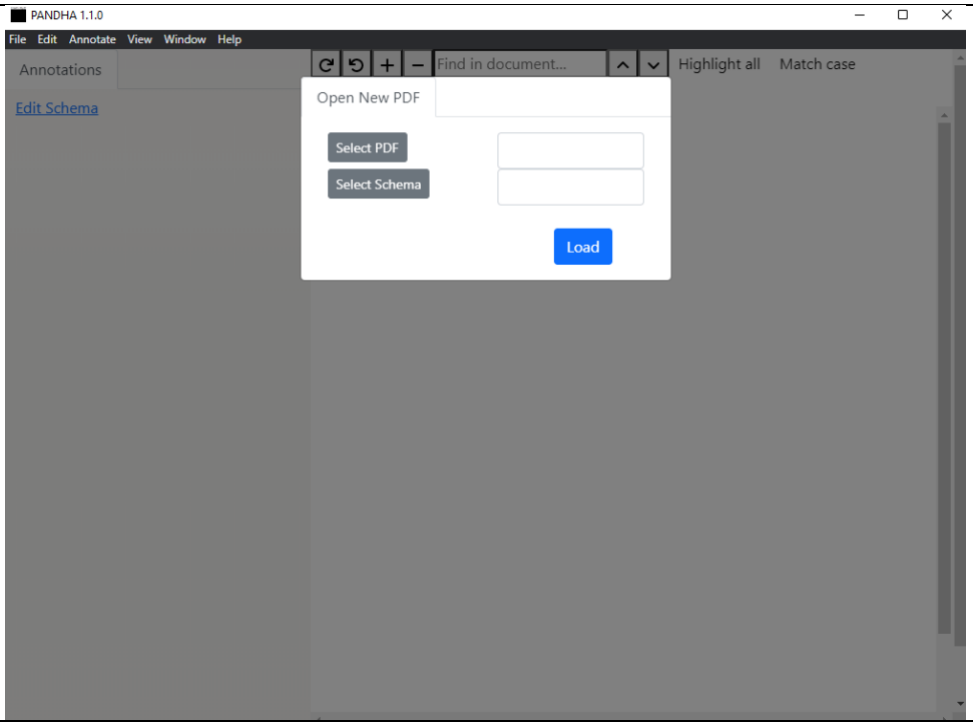
4. From the extracted location double click the *PANDHA_1.1.0*Folder



5. Double click *PANDHA.exe* to start the application



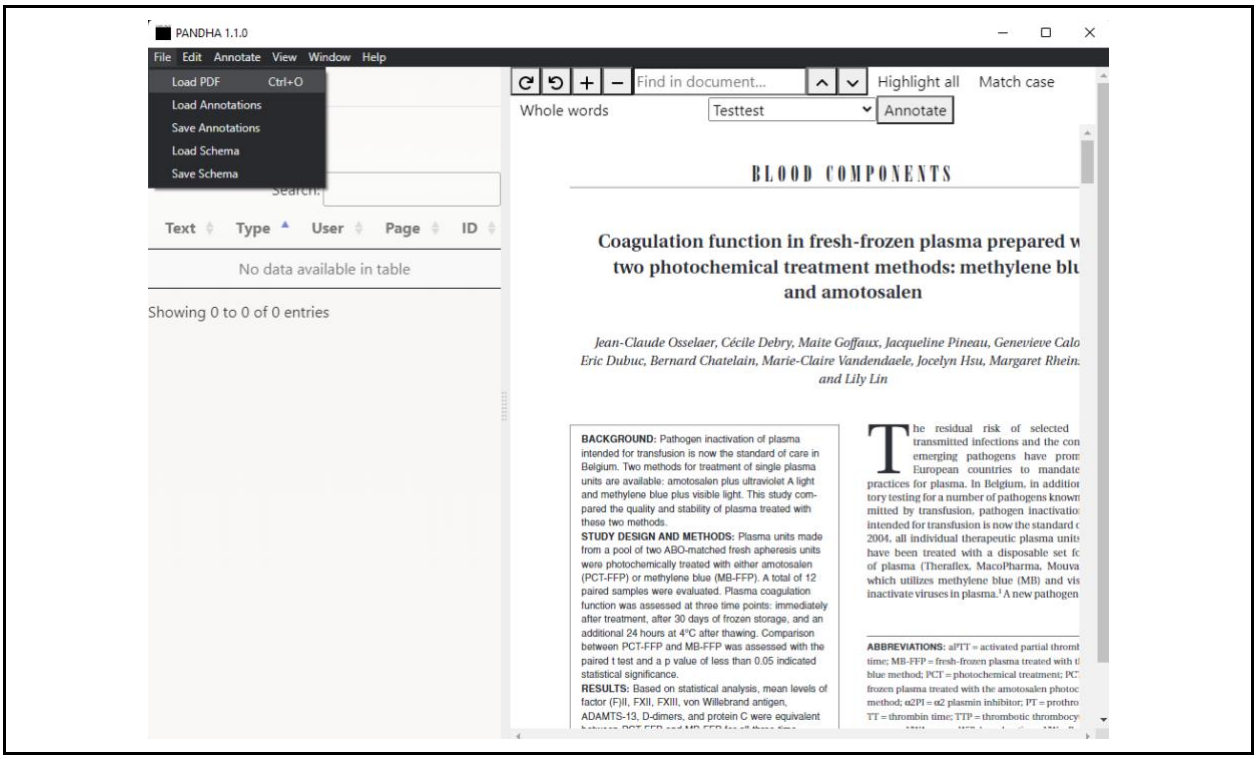
7. Select the PDF and Schema (tags.json is including in the pandha/test folder) and click *Load*



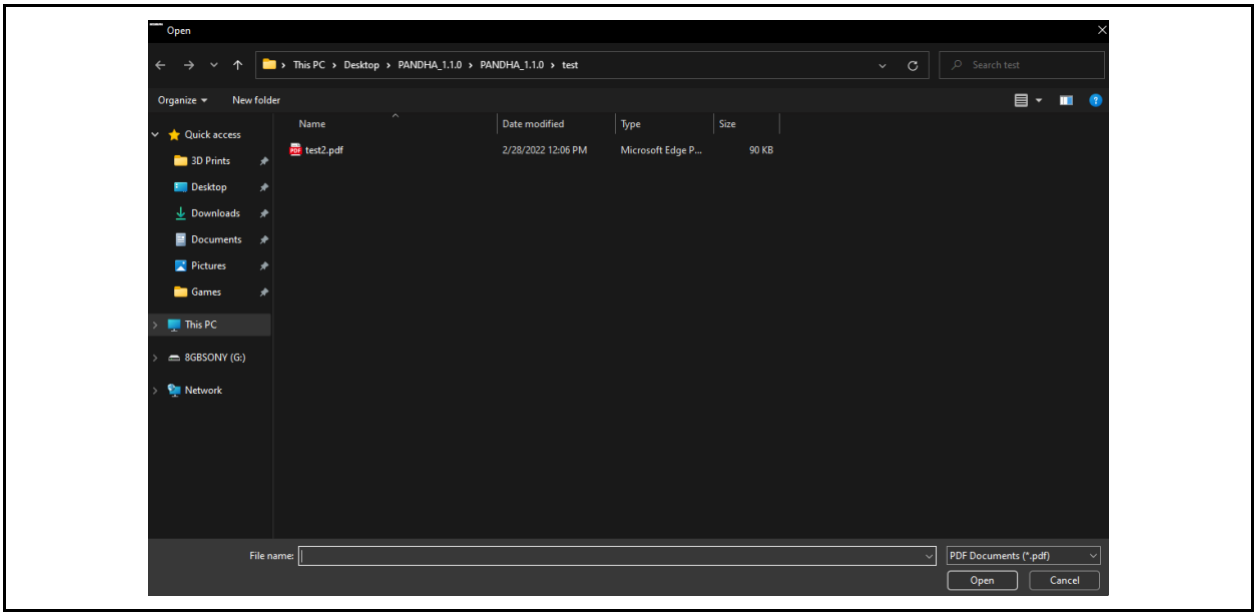
2 Application

2.0 Load PDF

1. Click *File* in the menu bar and select *Load PDF*.

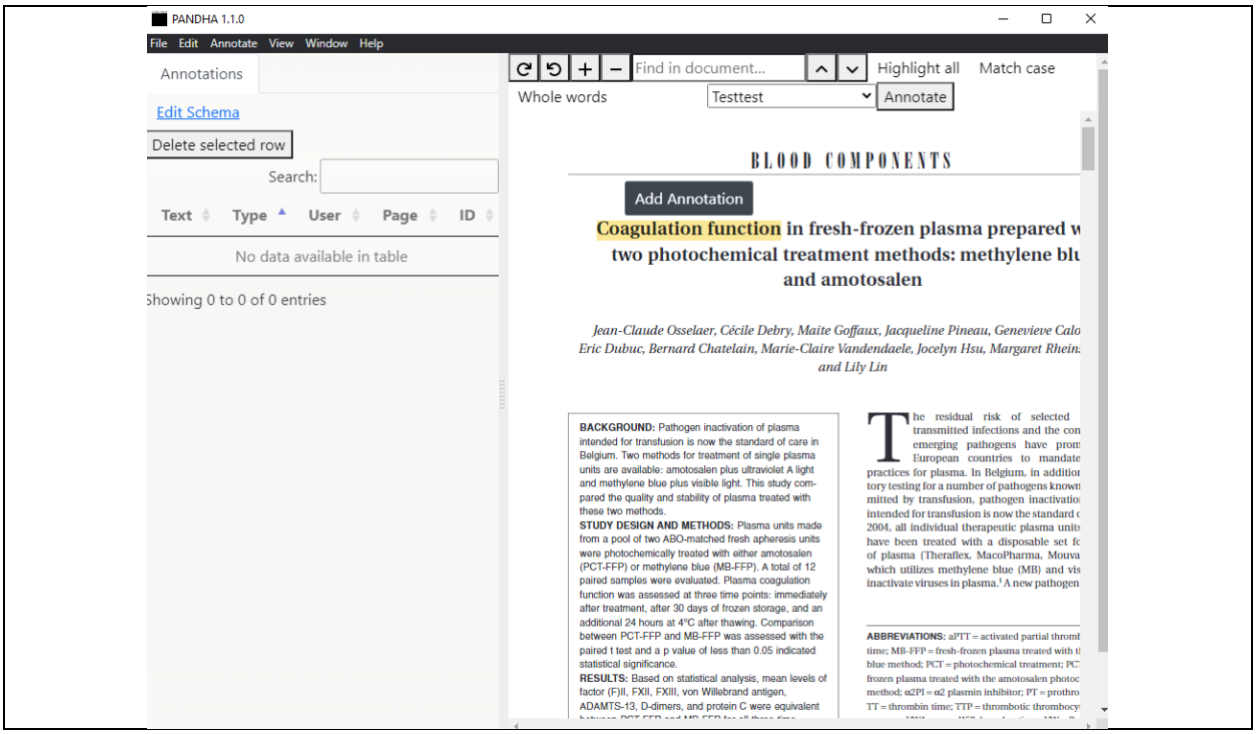


2. Select the PDF file from your computer.

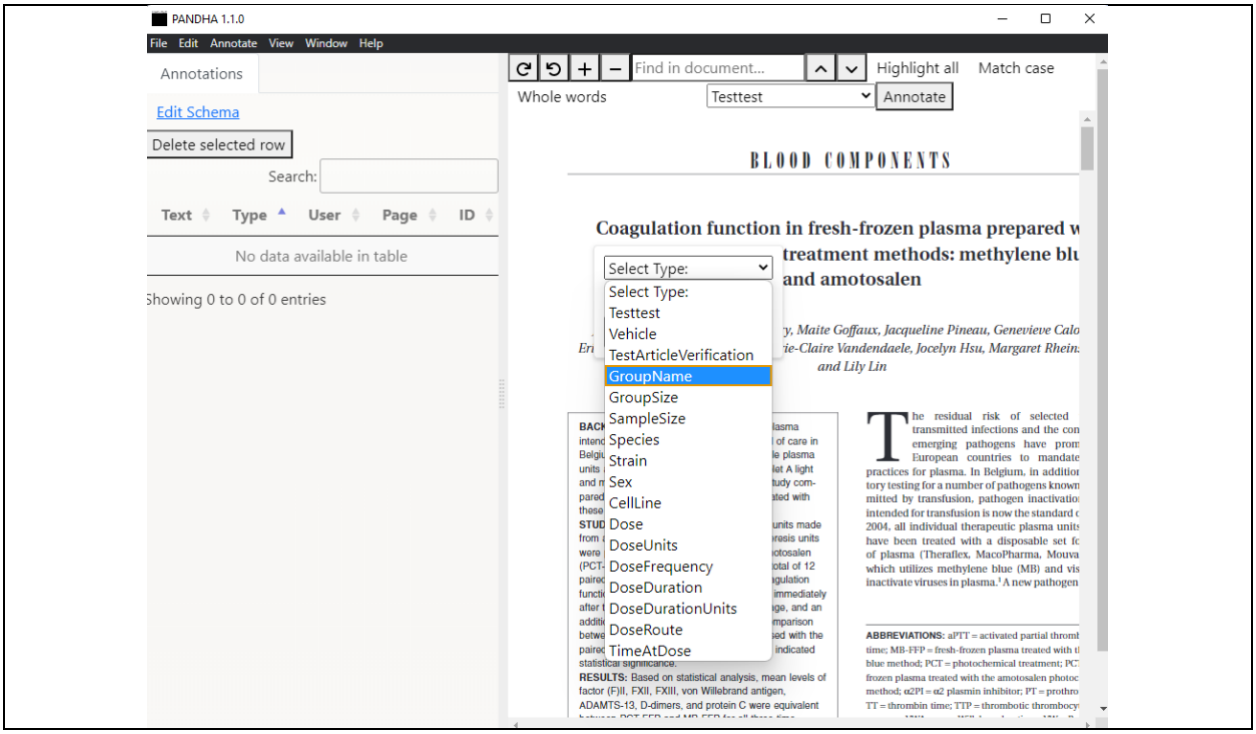


2.1 Annotate PDF

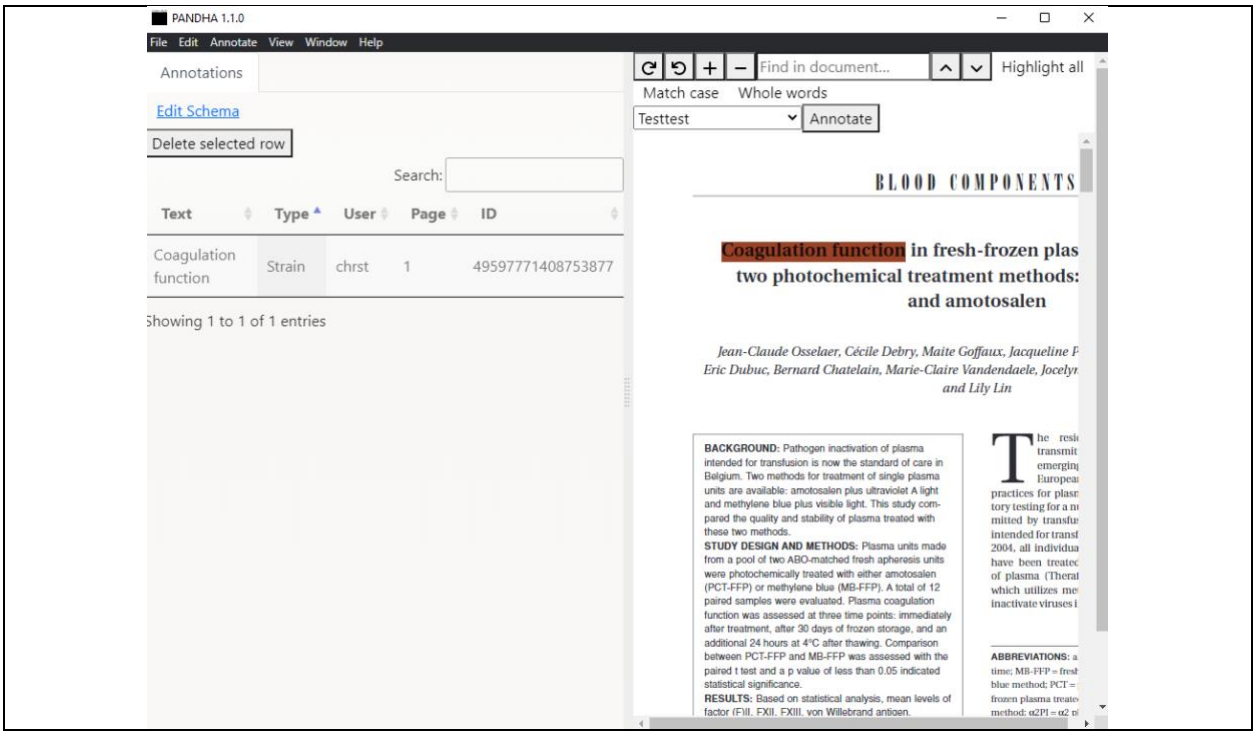
1. Highlight text you wish to annotate and select *Add Annotation*.



2. Select the annotation type.

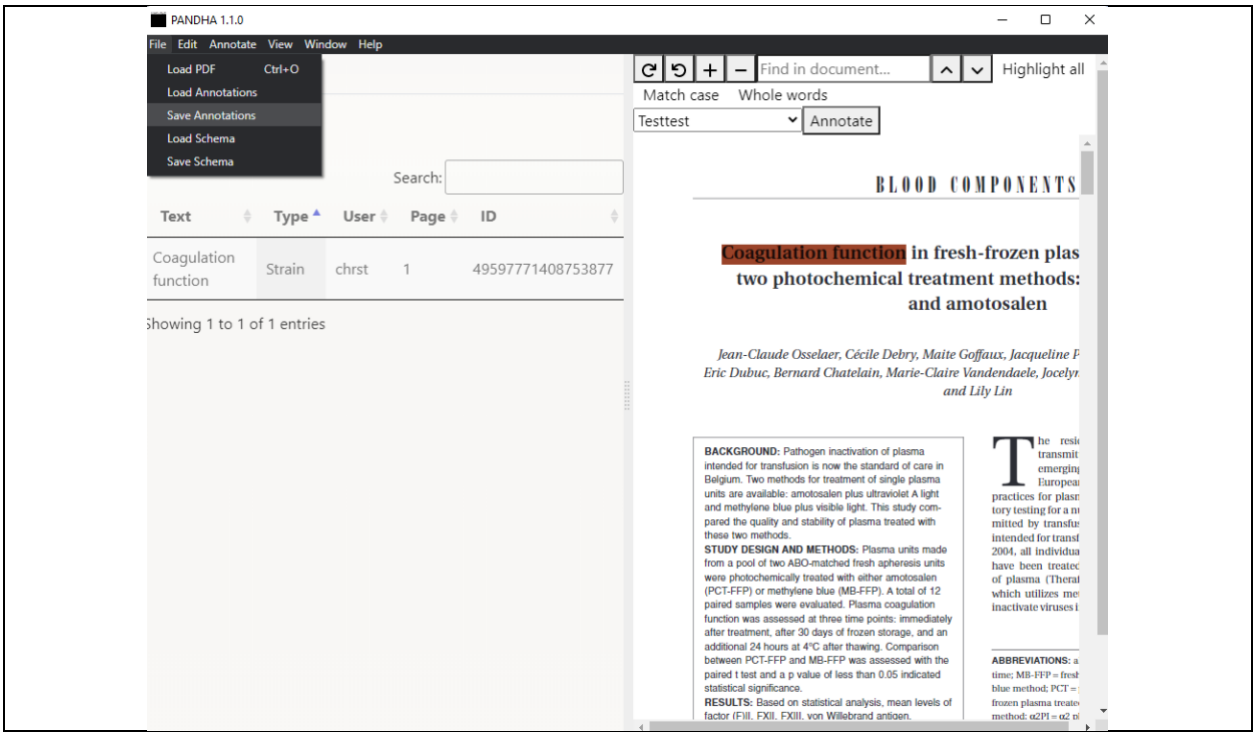


3. Hit save

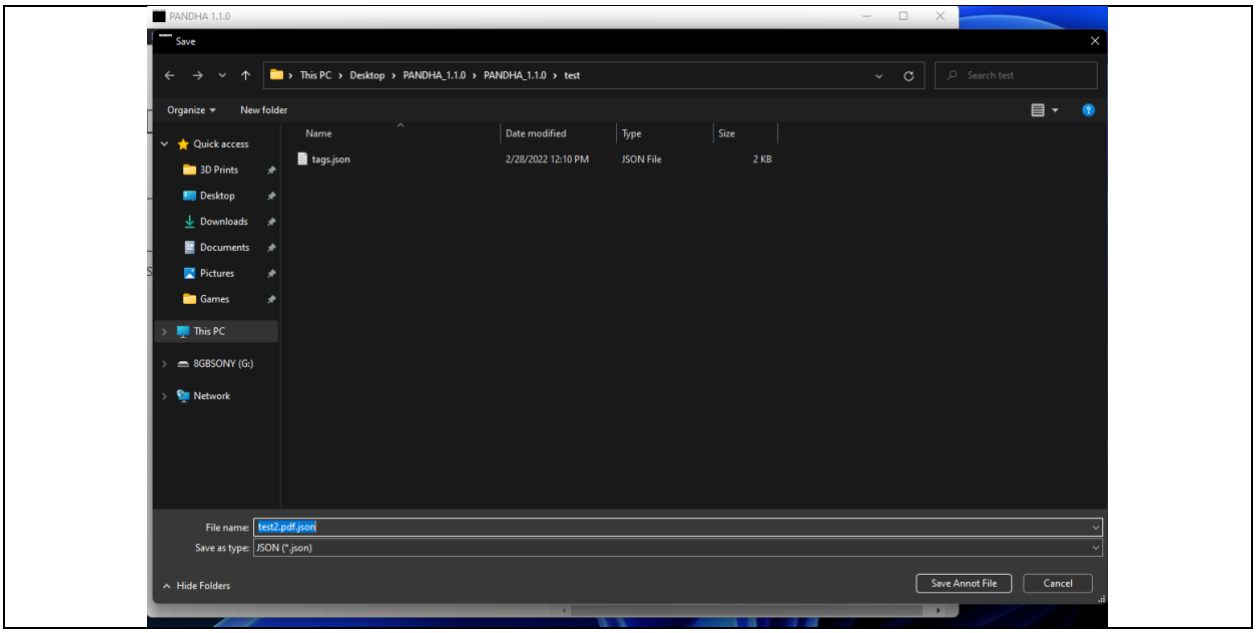


2.4 Save Annotations

1. Click *File* in the menu bar and select *Save Annotations*.

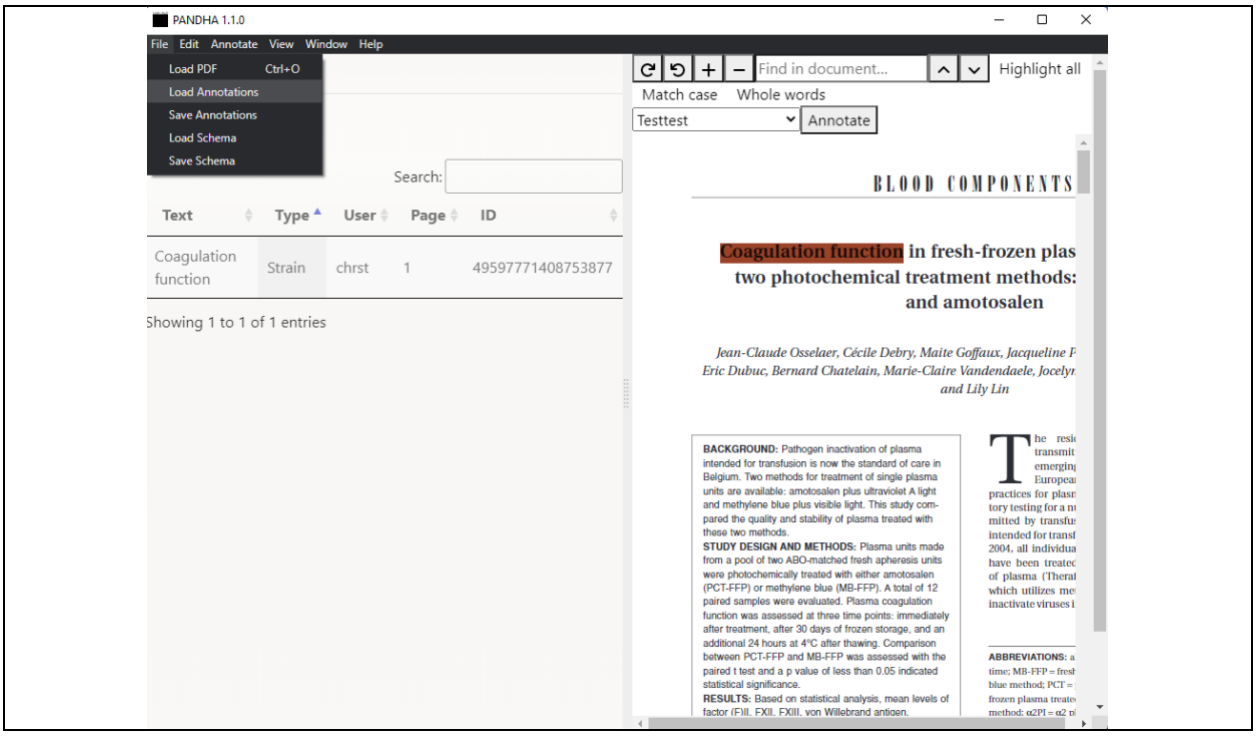


2. Select a save location on your computer and click *Save Annot File*.

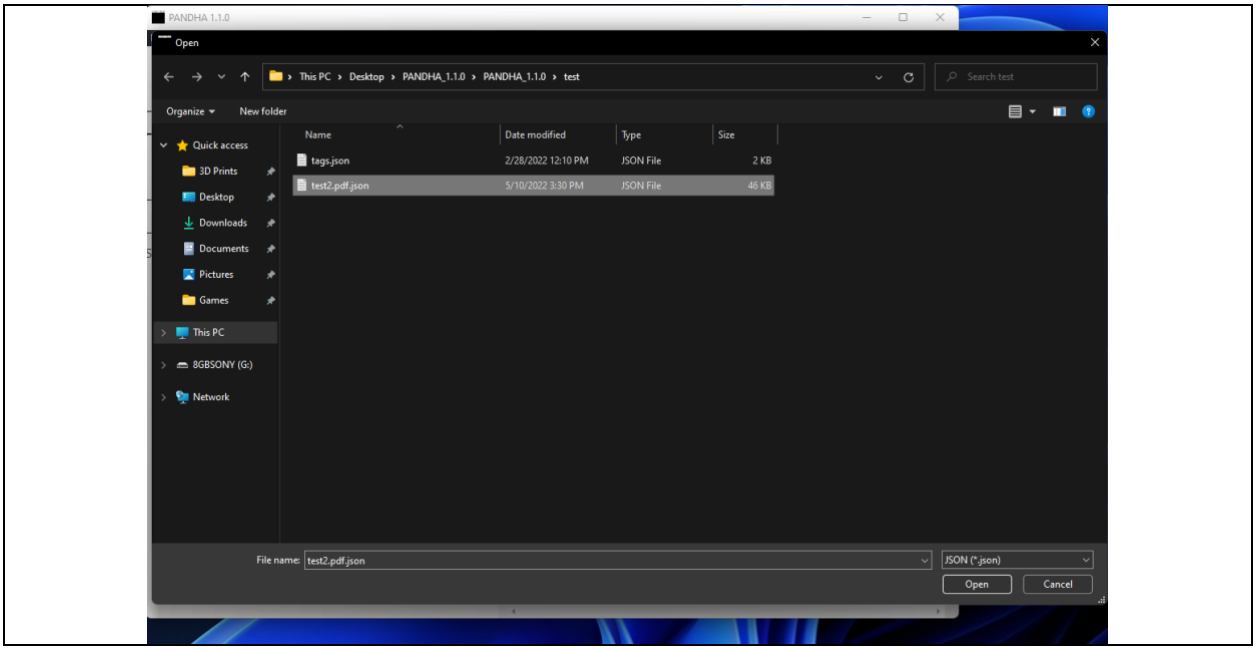


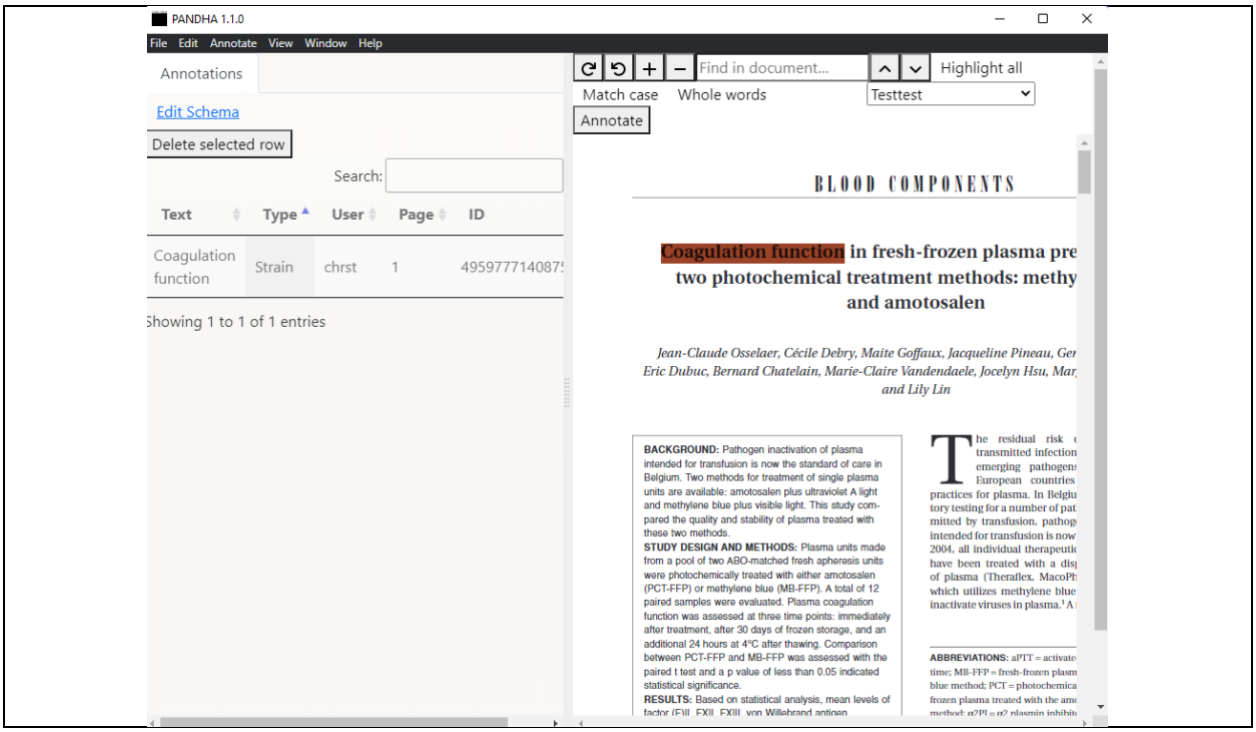
2.5 Load Annotations

1. Click *File* in the menu bar and select *Load Annotations*.



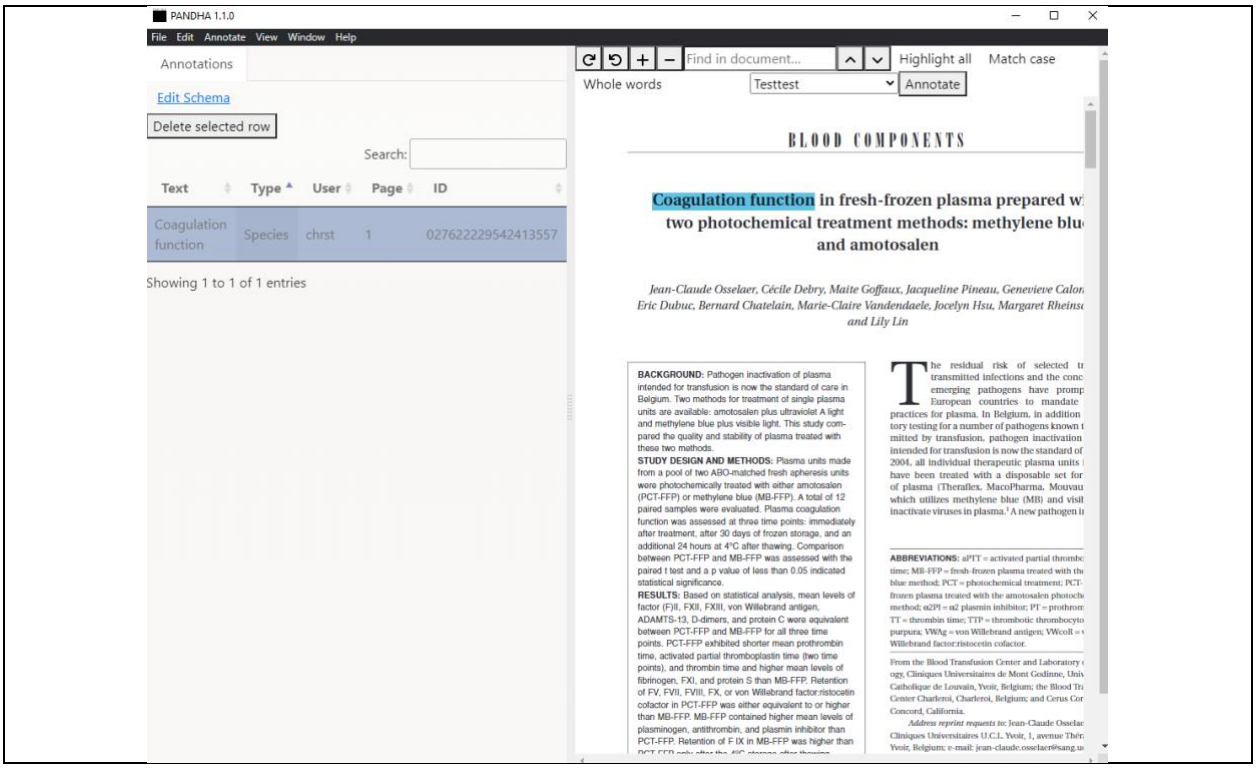
2. Select an annotation file



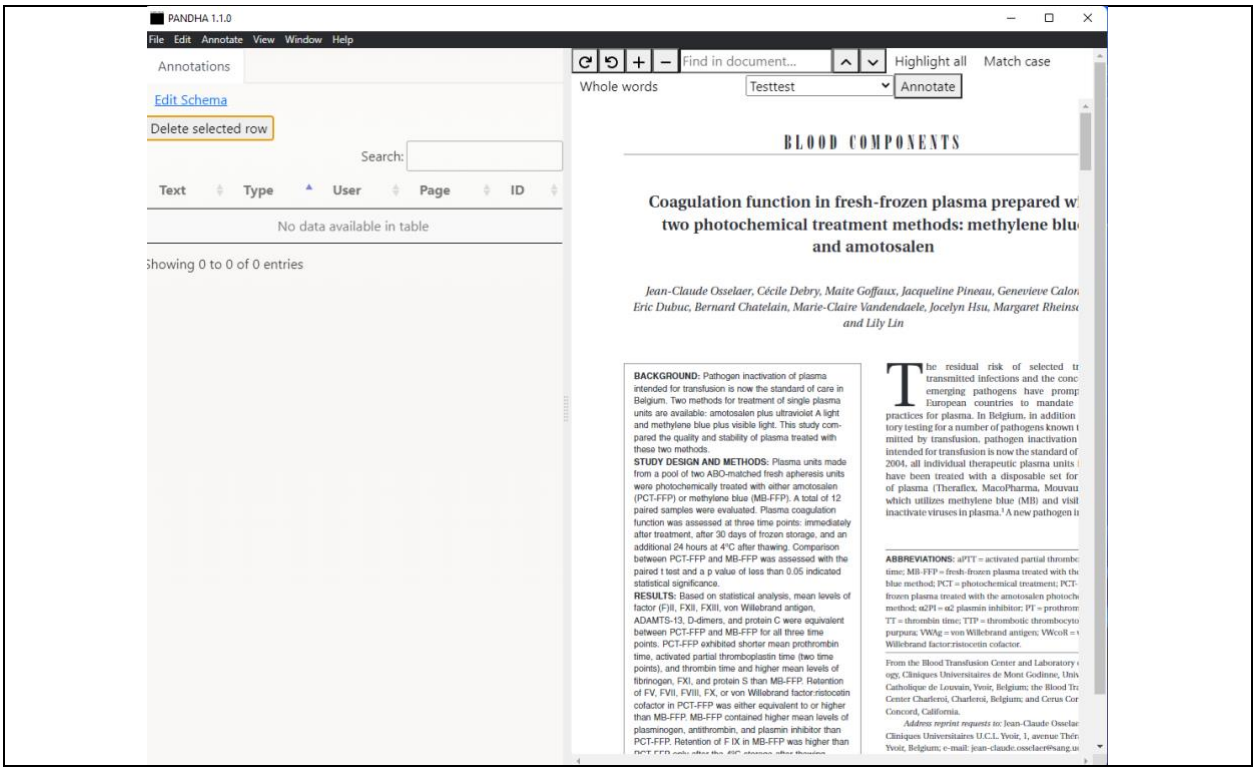


2.6 Delete Annotations

1. Select annotation you wish to delete from the table in the side bar.



2. Click *Delete selected row* button.



2.6 Edit Schema

1. Click *Edit Schema* hyper-link

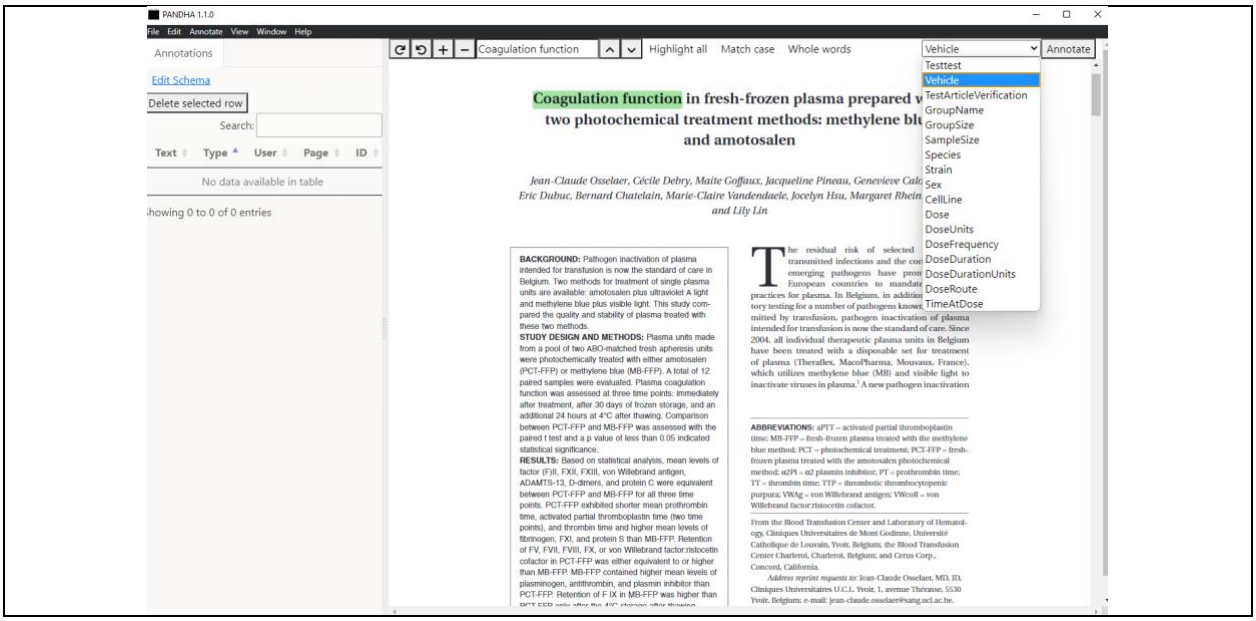
- Change existing entity
    - Click the text of any entity to edit that entities type.
    - Click the color selector to change the annotation color.
    - Click the trash can icon to delete that entity.
- Add new entity type
    - Click Add Entity Type to add a new entity.

2.7 Auto Annotation

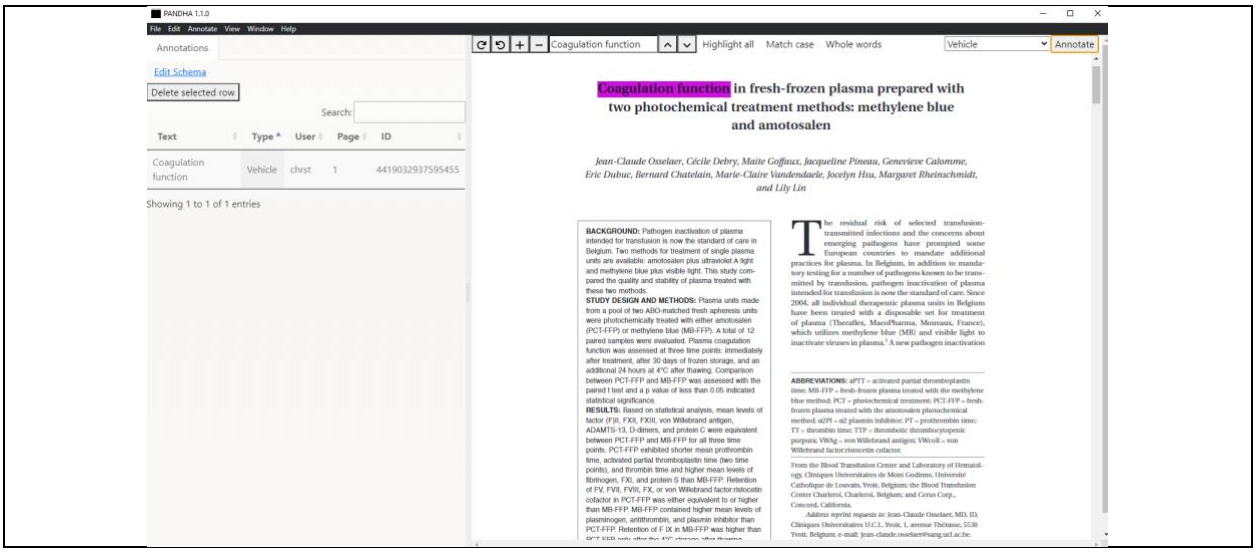
1. Type word or phrase to be searched for in *Find in document* search bar

2. Using the arrows (Up or Down) a green highlight will cycle through matches found in the document.

3. Select entity type from the dropdown box.



4. Click Annotate to add an annotation for the current selection.


PANDHA 1.1.0
File Edit Annotate View Window Help
Annotations
Edit Schema
Delete selected row
Search:
Text
Type
User
Page
ID
Coagulation function
Vehicle
chrst
1
4419032937595455
Showing 1 to 1 of 1 entries
Coagulation function
Highlight all
Match case
Whole words
Vehicle
Annotate
Coagulation function in fresh-frozen plasma prepared with two photochemical treatment methods: methylene blue and amotosalen
Jean-Claude Osselaer, Cécile Debry, Maite Goffaux, Jacqueline Pineau, Geneviève Calomme, Eric Dubuc, Bernard Chatelain, Marie-Claire Vandendaele, Jocelyn Hsu, Margaret Rheinschmidt, and Lily Lin
BACKGROUND: Pathogen inactivation of plasma intended for transfusion is now the standard of care in Belgium. Two methods for treatment of single plasma units are available: amotosalen plus ultraviolet A light and methylene blue plus visible light. This study compared the quality and stability of plasma treated with these two methods.
STUDY DESIGN AND METHODS: Plasma units made from a pool of two ABO-matched fresh apheresis units were photochemicaly treated with either amotosalen (PCT-FFP) or methylene blue (MB-FFP). A total of 12 paired samples were evaluated. Plasma coagulation function was assessed at three time points: immediately after treatment, after 30 days of frozen storage, and an additional 24 hours at 4°C after thawing. Comparison between PCT-FFP and MB-FFP was assessed with the paired t test and a p value of less than 0.05 indicated statistical significance.
RESULTS: Based on statistical analysis, mean levels of factor (F)II, FXII, FXIII, von Willebrand antigen, ADAMTS-13, D-dimers, and protein C were equivalent between PCT-FFP and MB-FFP for all three time points. PCT-FFP exhibited shorter mean prothrombin time, activated partial thromboplastin time (two time points), and thrombin time and higher mean levels of fibrinogen, FXI, and protein S than MB-FFP. Retention of FV, FVII, FVIII, FX, or von Willebrand factor:ristocetin cofactor in PCT-FFP was either equivalent to or higher than MB-FFP. MB-FFP contained higher mean levels of plasminogen, antithrombin, and plasmin inhibitor than PCT-FFP. Retention of F IX in MB-FFP was higher than
The residual risk of selected transfusion-transmitted infections and the concerns about emerging pathogens have prompted some European countries to mandate additional practices for plasma. In Belgium, in addition to mandatory testing for a number of pathogens known to be transmitted by transfusion, pathogen inactivation of plasma intended for transfusion is now the standard of care. Since 2004, all individual therapeutic plasma units in Belgium have been treated with a disposable set for treatment of plasma (Theraflex, MacoPharma, Mouvaux, France), which utilizes methylene blue (MB) and visible light to inactivate viruses in plasma.[1] A new pathogen inactivation
ABBREVIATIONS: aPTT = activated partial thromboplastin time; MB-FFP = fresh-frozen plasma treated with the methylene blue method; PCT = photochemical treatment; PCT-FFP = fresh-frozen plasma treated with the amotosalen photochemical method; α2PI = α2 plasmin inhibitor; PT = prothrombin time; TT = thrombin time; TTP = thrombotic thrombocytopenic purpura; VWAg = von Willebrand antigen; VWcoR = von Willebrand factor:ristocetin cofactor.
From the Blood Transfusion Center and Laboratory of Hematology, Cliniques Universitaires de Mont Godinne, Université Catholique de Louvain, Yvoir, Belgium; the Blood Transfusion Center Charleroi, Charleroi, Belgium; and Cerus Corp., Concord, California.
Address reprint requests to: Jean-Claude Osselaer, MD, ID, Cliniques Universitaires U.C.L. Yvoir, 1, avenue Thérasse, 5530 Yvoir, Belgium; e-mail: jean-claude.osselaer@uang.ucl.ac.be.