SIROCCO ADDAPTS

A Data-Driven Automated Pipeline & Tracking System

User Guide

Version 1.0.2

Rishi Nag
rn202@cam.ac.uk

2/8/2011

Table of Contents

Introduction

This document details the how to process smallRNA samples using the SIROCCO ADDAPTS pipeline.

A separate Administration Guide is provided detailing how to install the system, and configure tools and genomes. In this guide *Arabidopsis thaliana* will be provided as an example.

To view data no restrictions are required.

You will need to ensure before you progress.

- You have a URL to log on to

- A username and password

About SIROCCO

SIROCCO is an EU-funded collaboration over 20 research institutes investigating the role of Silencing RNAs as organisers and coordinators of complexity in eukaryotic organisms.

The goals of the SIROCCO project are

- to exploit RNA silencing mechanisms in order to improve human and plant health

- to reveal how siRNAs play a role in developmental biology and disease defense

- to provide fundamental insights into the genetic networks that underlie normal and diseased embryonic development and adult growth

For further details on the SIROCCO project and more bioinformatics tools, check the website at http://www.sirocco-project.eu

Licence

This code is released under GPL v3. See http://www.gnu.org/copyleft/gpl.html for more details.

Overview

SIROCCO ADDAPTS is essentially two components, connected together by a common database.

The first is the TRACKING system – a LIMS where the users can enter sample information and view results once the processing has completed. This is a web front end that can be viewed in a standard web browser. Your SIROCCO ADDAPTS administrator should provide the URL you need to access this.

The other component is the PIPELINE – this processes the samples as they arrive, running pre-alignment processes such as adaptor removal, genome alignment and then conversion into various file formats suitable for use as other items.

Some configuration occurs within the database and some occurs within files.The Administration Guide documents how to configure the relevant files and databases.

No user account is required to browse data. If you wish to be set up with an account to create and edit data please contact your SIROCCO ADDAPTS administrator.

Concepts

Project Administration

To facilitate tracking of funding, institutions and personnel there are

- Organisations – meant to represent the institutions. At the very least there will need to be an organisation representing the sequencing centre. This is detailed within the SIROCCO ADDAPTS Administration Guide. You can also add further institutions to cover your lab, funding bodies etc.

- People – individuals running projects, or controlling the ADDAPTS installation. There are three levels of user

  - administrators – can add/edit/delete configuration data

  - local - standard users – can add/edit sample related data

  - external - anonymous – can browse data

- Projects – multiple experiments often form part of a project.

Samples

There are four items to represent data sent for sequencing.

- Biosample – represents the biological sample.

- Library – represents the Biosample prepared for sequencing.

- Sequencing Sample – represents what gets sent to the sequencing centre – it may have one or many libraries prepared for sequencing.

- Sequencing Run – represents an individual sequencing run on the sequencing machine.

When data has returned from the sequencing centre a program needs to be run to let the system know that the data is now available. The pipeline will then start processing the file and the results can be viewed.

Using ADDAPTS Tracking Component

To access the homepage just type the URL provided by your SIROCCO ADDAPTS administrator into your browser.
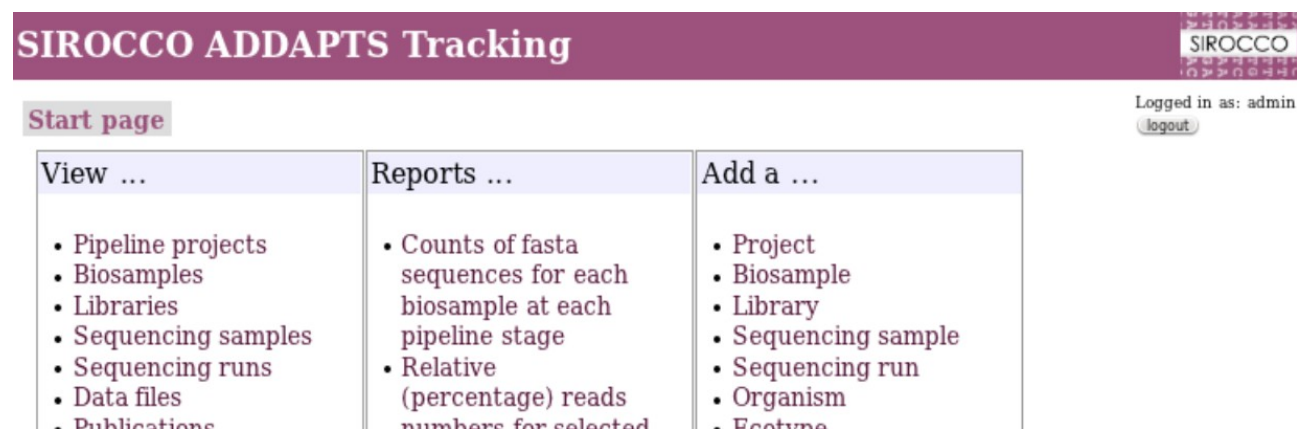
A page like the one below should appear.
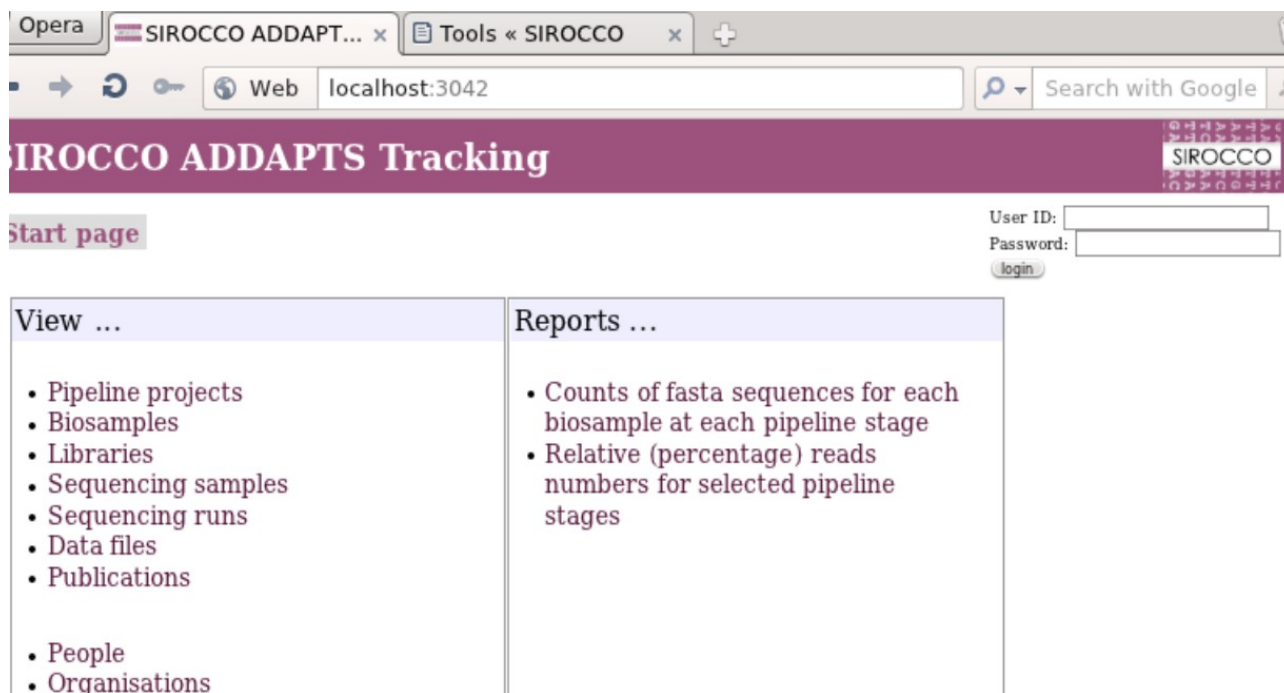
You can return to the home page at any time by clicking on the banner at the top of the page.

To browse samples use the links on the left.

Logging In

Log in to the system using details provided by your ADDAPTS administrator using the login fields on the top right corner of the home page.

Setting Up A Sample In The Database

This section details the data entry to register an sample having been prepared, and being ready or already having been sent off, to your sequencing centre. It is good practise to do it early when samples have been just prepared or just sent to the facility, not when it has been returned.

Prerequisites

1. Before starting this process you should ensure the organism you require to process has been set up. This will need to be discussed with your ADDAPTS administrator

2. Ensure you are logged in - once you are logged in a third 'Add a ...' column will appear on the home page.

3. Ensure you have read the Concepts – Sample section of this document to understand the steps you are about to undergo.

4. Ensure the ADDAPTS system has been setup to cope with the Ecotype and Tissue types you require for your organism. It is a good idea to clarify this across the lab before configuring the system. The neccesary values can be added using the 'Add a Ecotype' and 'Add a Tissue' options from the third column on the home page

5. If the sample is intended to be part of a project, create a project here.

6. Ensure the sequencing centre has been set up as an organisation and as a process_conf. Clarify with the ADDAPTS adminstrator that this is the case.

7. Any adapters or barcodes you are using have been set up in the system.

The following steps as described are all designed to follow on from one another: Biosample -> Library -> Sequencing Sample -> Sequencing Run. However you can resume the process if neccesary by browsing to the last item you stopped at.

Biosample

Note – for smallrna processing you will need to set the biosample and molecule types appropriately.

New biosample

name	SL1
description	Sample 1
biosample creator	Rishi Nag
genotype	
protocol	unknown
biosample type	small_rnas
molecule type	RNA
treatment type	no treatment
fractionation type	no fractionation
tissue	Leaf
processing requirement	pipeline processing

ecotypes
- ☑ C24 Arabidopsis thaliana
- ☐ Wild Type Arabidopsis thaliana

projects
- ☑ Project A

Library

If following on directly from a newly created Biosample click on 'Create a library from this biosample' on the follow on page, or otherwise via the 'Add a Library' option from the homepage. If doing the latter option ensure the biosample value is set correctly.

New library

identifier	SL1_L1
description	A library that I created
run type	technical replicate
biosample	SL1
barcode	
adaptor	illumina v1 smallrna adaptor
sequencing sample	

At this stage you can leave the sequencing sample field blank. If combining multiple libraries into a single sequencing sample you can enter this here.

Sequencing Sample

If following on directly from a newly created library click on 'Create a sequencing sample from this library' on the follow on page, or otherwise via the 'Add a Sequencing Sample' option from the homepage.

New sequencing sample

identifier	SL1_L1_SQ1
sequencing centre identifier	
	sample creator Rishi Nag ⌄
read length	36
end type	⌄
number of lanes	1

libraries
☑ SL1_L1

Ensure the libraries making up this sequencing sample are all selected. The sequencing centre at this stage will probably not have designated an identifier from their end at this point, so it can be left blank and edited later.

Sequencing Run

New sequencing run

identifier	RUN-1
Sequencing centre	SIROCCO ⌄
Sequencing type	Illumina ⌄
Sequencing sample	SL1_L1_SQ1 ⌄
quality	high ⌄

At this level you can enter data for a sequencing run. The quality can be used to indicate result of a biolyzer machine or the suchlike.

Once you have clicked Submit the tracking system has registered the sample as having been sent off for sequencing.

Processing Sequenced Data with the ADDAPTS Pipeline Component

Your sequencing centre will typically generate a FASTQ file.

The SIROCCO ADDAPTS Administration Guide details how to start processing using the pipeline.

Browsing Results

To examine already processed samples use the links in the left hand column. Lists will appear allowing you to browse results. To view the data for the Biosample you just generated use. View->Biosample.