



HAL
open science

CCQTL: facilitating QTL mapping in the Collaborative Cross

Rémi Planel, Victoire Baillet, Jean Jaubert, Christian A. J. Vossheirich, Rachel Torchet, Marion Rincel, Pascal Campagne, Xavier Montagutelli

► **To cite this version:**

Rémi Planel, Victoire Baillet, Jean Jaubert, Christian A. J. Vossheirich, Rachel Torchet, et al.. CCQTL: facilitating QTL mapping in the Collaborative Cross. 20th Annual Meeting of the Complex Trait Community | Rat Genome & Models, University of Tennessee Health Science Center [Memphis, Tennessee], Oct 2023, Memphis, TN, United States. pasteur-04588091

HAL Id: pasteur-04588091

<https://pasteur.hal.science/pasteur-04588091v1>

Submitted on 25 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CCQTL: facilitating QTL mapping in the Collaborative Cross

Rémi Planel*, **Victoire Baillet***, Vincent Guillemot, Jean Jaubert, Christian Vosshenrich,
Rachel Torchet, Marion Rincel, Pascal Campagne, Xavier Montagutelli

INSTITUT
PASTEUR

Collaborative Cross initiative at Institut Pasteur

- The Collaborative Cross: a precious resource for dissecting the genetic basis of complex traits
 - Recombinant inbred strains: repeatability
 - High genetic diversity: phenotypic variation for large # of traits
 - Well-characterized (haplotypes, genomes, ...)
- Institut Pasteur CC Initiative
 - Coordination: Mouse Genetics Laboratory (X. Montagutelli)
 - ~35 CC strains imported and bred at IP, available to all IP scientists
 - 21 "CC-Club" members, belonging to 5 Departments: geneticists and beyond: immunologists, neuroscientists, ...



CCQTL: rationale

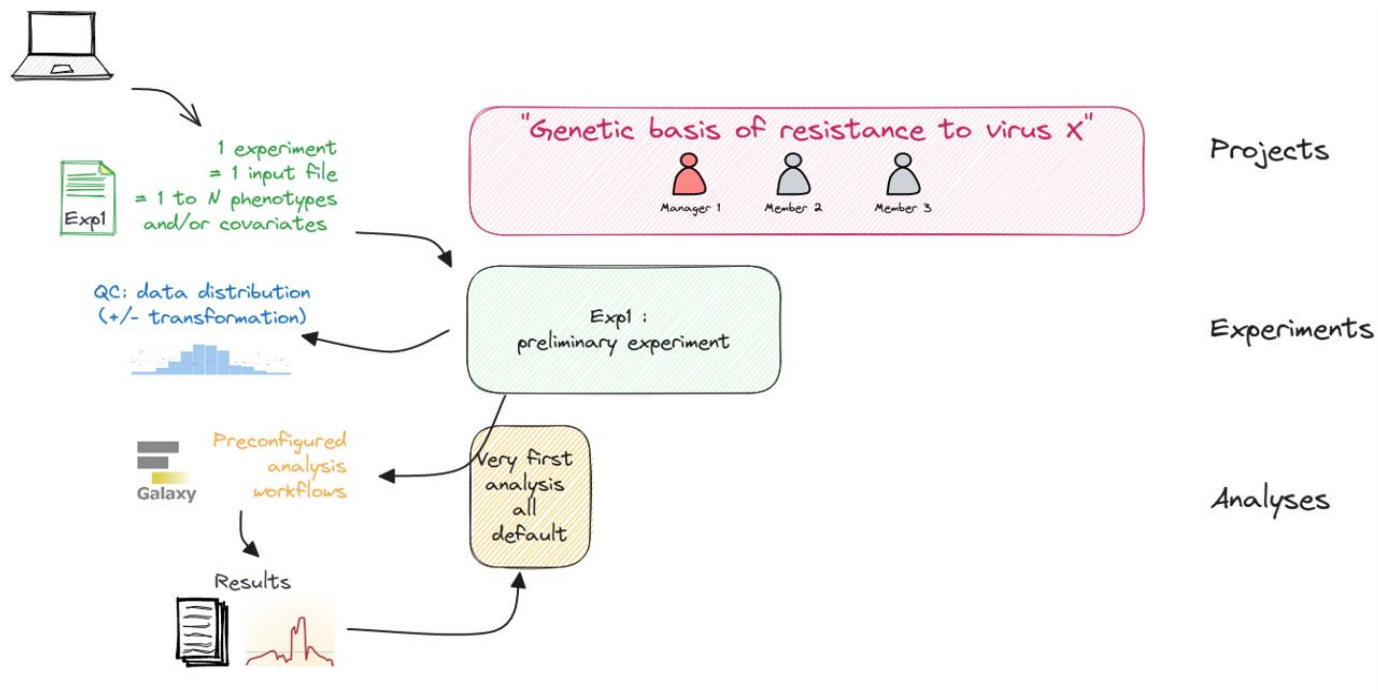
- QTL mapping combines (heavy) phenotyping experiments with analyses calling on quantitative skills.
- Experimentalists often need support for both code and statistics to benefit from best practices in the field.
- Analysing QTL mapping experiments involves going back and forth on the data and often require informed decisions that can only be made by the experimentalist, making it more complex to "outsource".



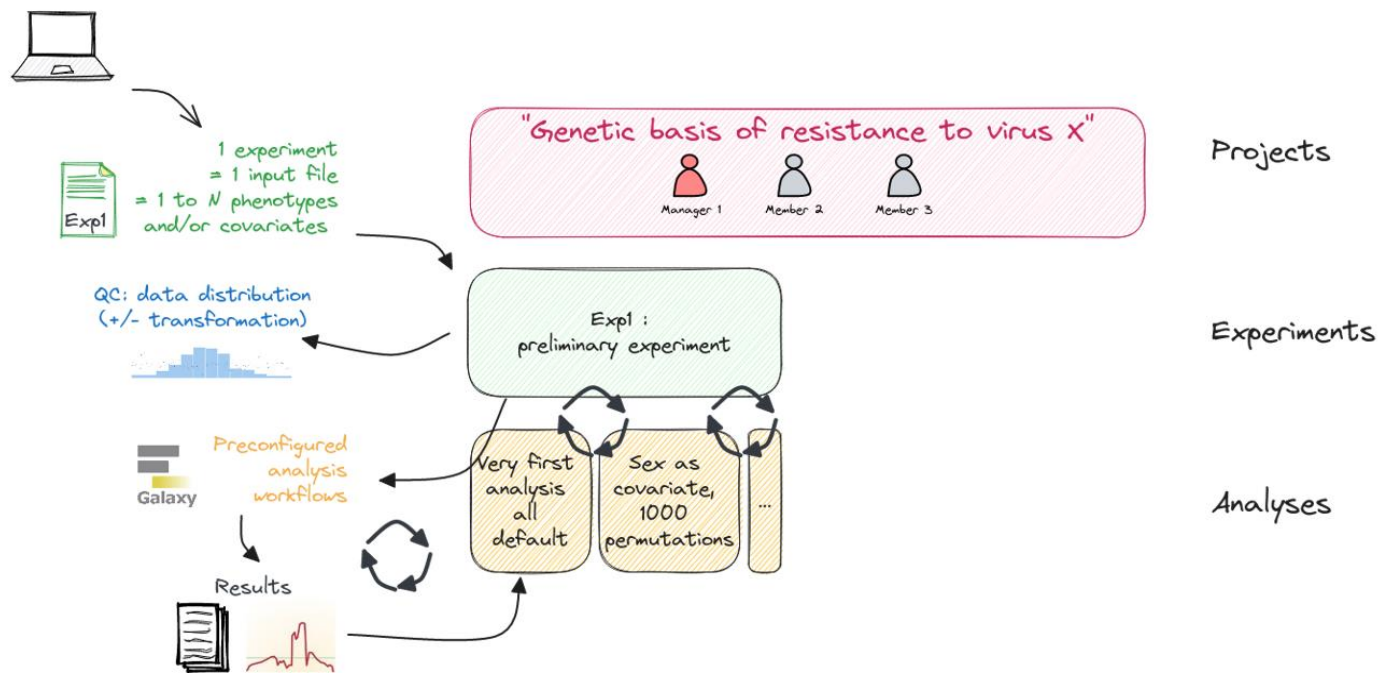
Make the biologists autonomous in analysing their QTL mapping experiments in the CC

- Web interface to easily perform QTL mapping and visualize the results
- Database system to keep the data organised
- Reproducibility through Galaxy-powered, pre-configured analysis workflows

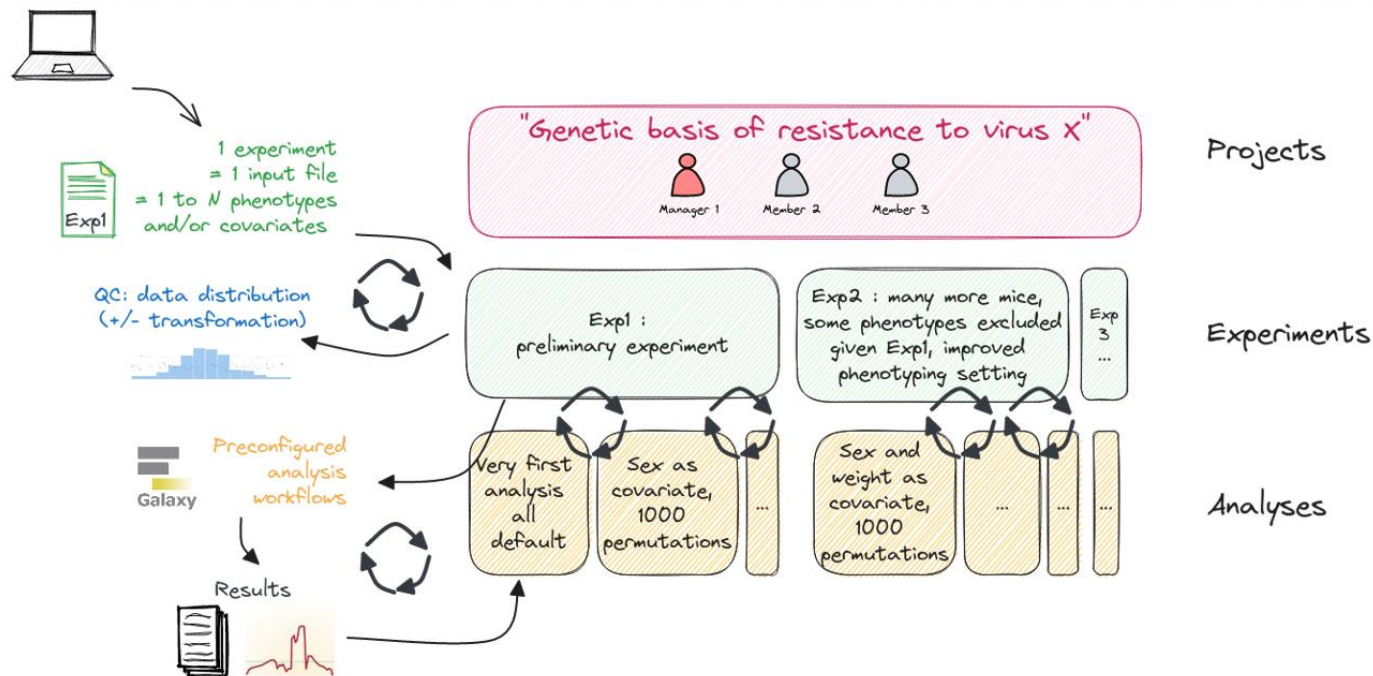
CCQTL in practice



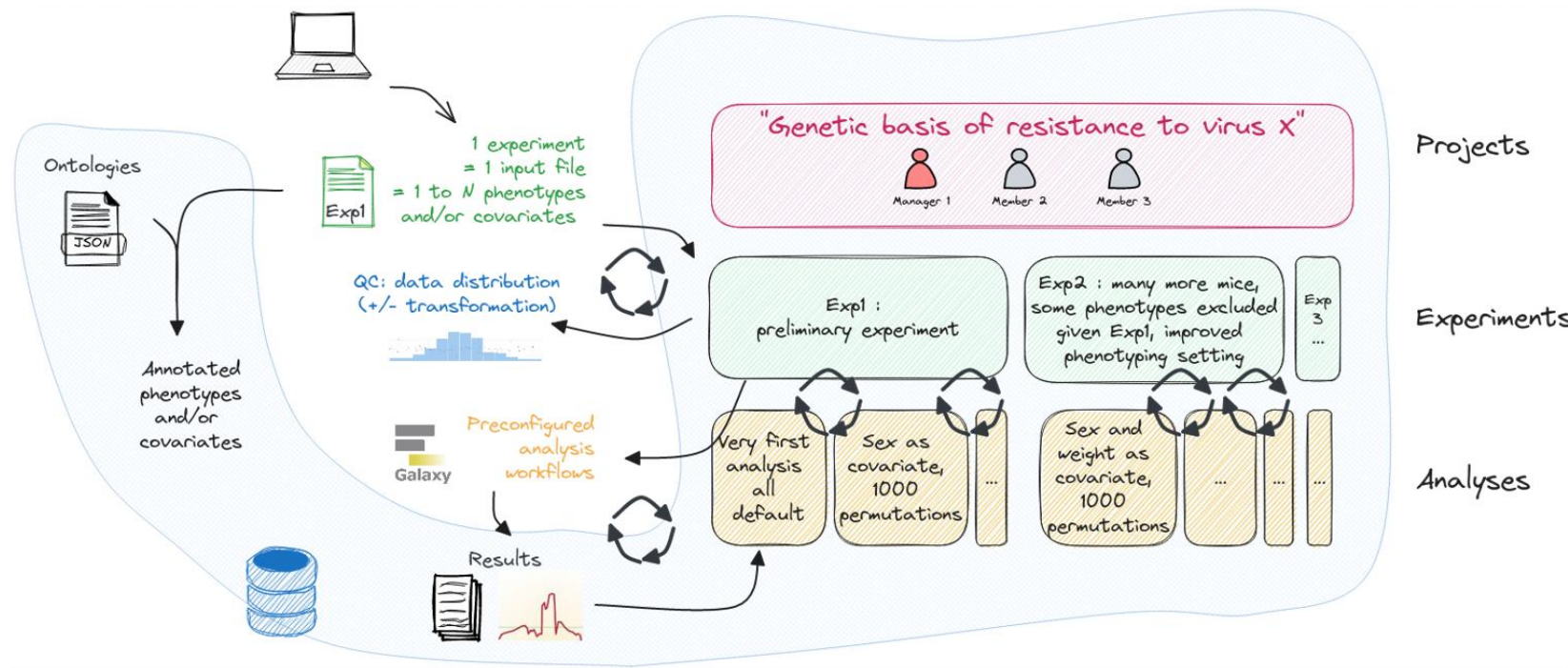
CCQTL in practice



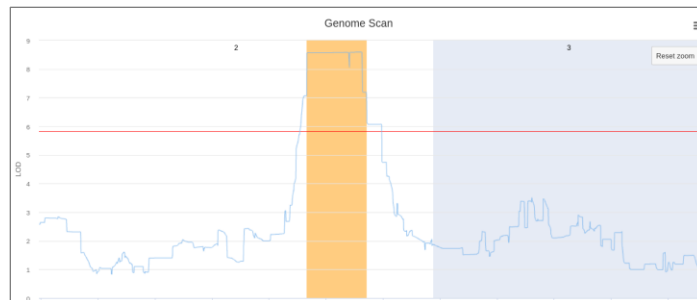
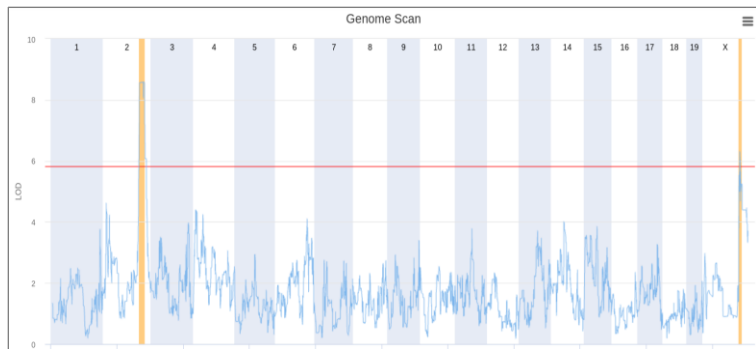
CCQTL in practice



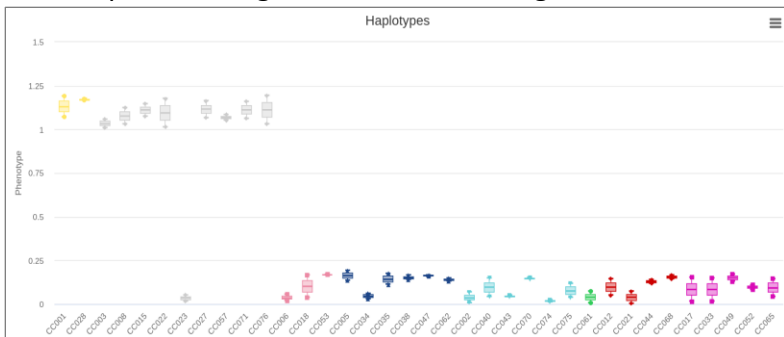
CCQTL in practice



CCQTL offers multiple plotting features for QTL exploration



Which parental origin associates with high/low trait values ?



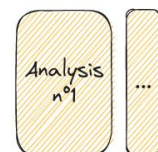
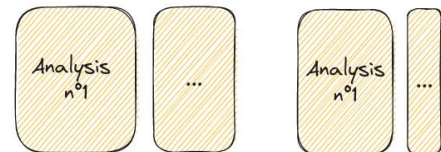
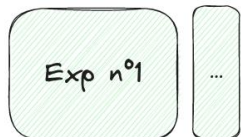
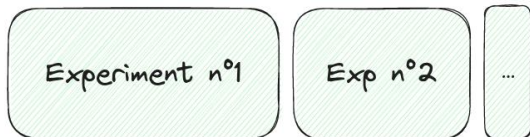
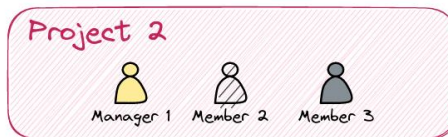
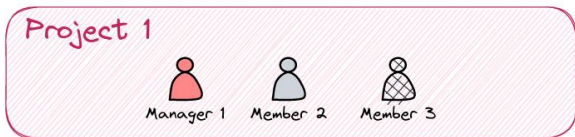
EFFECTS HAPLOTYPES
 SNPS TOP SNPS GENES

Do we have obvious candidate genes in the QTL interval ?

Genes

id	Chromosome	Start	Stop	Strand	Type	Biotype	Name	MGI Name	Alias	Database cross-reference
MGI:C57BL6J_1914221	2	154222011	154373038	-	gene	protein coding gene	Cdk5rap1	MGI:1914221	CDK5 regulatory subunit associated protein 1	NCBI:46971 ENSEMBL:ENSMUSG0000027468
MGI:C57BL6J_1918675	2	154307227	154320646	+	gene	protein coding gene	Epf1rb9b	MGI:1918675	BPI fold containing family (B2c) member 9B	NCBI:43245 ENSEMBL:ENSMUSG0000027796
MGI:C57BL6J_101772	2	154376313	154408107	-	gene	protein coding gene	Smta1	MGI:101772	synthrophin-like acidic 1	NCBI:30468 ENSEMBL:ENSMUSG0000027488
MGI:C57BL6J_562847	2	154416499	154421329	-	gene	lincRNA gene	Gm39962	MGI:562847	predicted gene-like 39962	NCBI:10254431
MGI:C57BL6J_6366846	2	154426601	154430434	+	gene	lincRNA gene	Gm52564	MGI:6366846	predicted gene-like 52564	NCBI:115449332
MGI:C57BL6J_1333833	2	154436098	154539356	+	gene	protein coding gene	Ctfa2t2	MGI:1333833	CBFA2/IRANK1 translocation partner 2	NCBI:12706 ENSEMBL:ENSMUSG0000027626
MGI:C57BL6J_5010952	2	154507683	154508290	-	pseudogene	pseudogene	Gm18767	MGI:5010952	predicted gene-like 18767	NCBI:100475897

Focus on: database and permissions



Project:

research project, several members with distinct roles

Experiment:

data, organized by rounds of acquisition

Analysis:

QTL mapping analysis (with all parameters being recorded)

- Permissions are managed at the Project level (compartmentalisation)
- Within each project: manager vs. maintainer roles

Focus on: preconfigured analysis workflows

← ANALYSIS | RUN ANALYSIS

1 Select an experiment

2 Select a workflow

ccqtl-wf v0.0.1_beta.2

Steps

- 2 ccqtl_make_inputs 0.0.1_beta.2+galaxy0
- 3 ccqtl_genome_scan 0.0.1_beta.2+galaxy0
- 4 ccqtl_permutations 0.0.1_beta.2+galaxy1
- 5 ccqtl_find_peaks 0.0.1_beta.2+galaxy0
- 6 ccqtl_refine_region 0.0.1_beta.2+galaxy0

CONTINUE

3 Set workflow parameters
ccqtl-wf v0.0.1_beta.2

Exp n°1

Workflow: QTL mapping for continuous traits in CC mice

Focus on: preconfigured analysis workflows

The screenshot shows a web interface for running an analysis. At the top, there is a navigation bar with a back arrow and the text 'ANALYSIS' and 'RUN ANALYSIS'. Below this, there are three numbered steps: 1. 'Select an experiment' (with a pencil icon), 2. 'Select a workflow' (with a checkmark icon), and 3. 'Set workflow parameters' (with a gear icon). Under step 2, a workflow named 'ccqtl-wf' is selected, with a version indicator 'v0.0.1_beta.2'. Below the workflow name, a list of steps is shown: 2. 'ccqtl_make_inputs' (v0.0.1_beta.2+galaxy0), 3. 'ccqtl_genome_scan' (v0.0.1_beta.2+galaxy0), 4. 'ccqtl_permutations' (v0.0.1_beta.2+galaxy1), 5. 'ccqtl_find_peaks' (v0.0.1_beta.2+galaxy0), and 6. 'ccqtl_refine_region' (v0.0.1_beta.2+galaxy0). A blue 'CONTINUE' button is located at the bottom left of the workflow selection area.

Exp n°1

Workflow: QTL mapping for continuous traits in CC mice

Steps:

- r/qt2 inputs creation
- Haplotype reconstruction
- Genome scan (LMM, LOCO kinship)
- Permutations-derived significance
- LOD peaks identification
- Founder allele effects estimation
- Local SNP association mapping
- List of genes in support interval

Analysis n°1



(+ RDS files)

Focus on: preconfigured analysis workflows

← ANALYSIS | RUN ANALYSIS

1 Select an experiment

2 Select a workflow

ccqtl-wf v0.0.1_beta.2

Steps

- 2 ccqtl_make_inputs 0.0.1_beta.2+galaxy0
- 3 ccqtl_genome_scan 0.0.1_beta.2+galaxy0
- 4 ccqtl_permutations 0.0.1_beta.2+galaxy1
- 5 ccqtl_find_peaks 0.0.1_beta.2+galaxy0
- 6 ccqtl_refine_region 0.0.1_beta.2+galaxy0

CONTINUE

3 Set workflow parameters
ccqtl-wf v0.0.1_beta.2

Exp n°1

Workflow: QTL mapping for continuous traits in CC mice

Steps:

- r/qtI2 inputs creation
- Haplotype reconstruction
- Genome scan (LMM, LOCO kinship)
- Permutations-derived significance
- LOD peaks identification
- Founder allele effects estimation
- Local SNP association mapping
- List of genes in support interval

Tunable parameters (if not default):

- Marker set (Mega/GigaMUGA)
- Genome version/annotation (m38)
- CC genotype data
- Additive covariate for genome scan
- Permutations # / stratification
- Significance level / Bonferroni correction for multiple phenotypes
- LOD drop-off value

Analysis n°1



(+ RDS files)

Focus on: preconfigured analysis workflows

← ANALYSIS | RUN ANALYSIS

1 Select an experiment

2 Select a workflow

ccqtl-wf v0.0.1_beta.2

Steps

- 2 ccqtl_make_inputs v0.0.1_beta.2+galaxy0
- 3 ccqtl_genome_scan v0.0.1_beta.2+galaxy0
- 4 ccqtl_permutations v0.0.1_beta.2+galaxy1
- 5 ccqtl_find_peaks v0.0.1_beta.2+galaxy0
- 6 ccqtl_refine_region v0.0.1_beta.2+galaxy0

CONTINUE

3 Set workflow parameters
ccqtl-wf v0.0.1_beta.2

Exp n°1

Workflow: QTL mapping for continuous traits in CC mice

Steps:

- r/qtI2 inputs creation
- Haplotype reconstruction
- Genome scan (LMM, LOCO kinship)
- Permutations-derived significance
- LOD peaks identification
- Founder allele effects estimation
- Local SNP association mapping
- List of genes in support interval

Analysis n°1



(+ RDS files)

1 workflow = 1 type of analysis allowing easy extensions (binary traits, CC-RIX, ..) & versioning

Tunable parameters (if not default):

- Marker set (Mega/GigaMUGA)
- Genome version/annotation (m38)
- CC genotype data
- Additive covariate for genome scan
- Permutations # / stratification
- Significance level / Bonferroni correction for multiple phenotypes
- LOD drop-off value

Focus on: preconfigured analysis workflows

← ANALYSIS | RUN ANALYSIS

1 Select an experiment

2 Select a workflow

ccqtl-wf v0.0.1_beta.2

Steps

- 2 ccqtl_make_inputs 0.0.1_beta.2+galaxy0
- 3 ccqtl_genome_scan 0.0.1_beta.2+galaxy0
- 4 ccqtl_permutations 0.0.1_beta.2+galaxy1
- 5 ccqtl_find_peaks 0.0.1_beta.2+galaxy0
- 6 ccqtl_refine_region 0.0.1_beta.2+galaxy0

CONTINUE

3 Set workflow parameters
ccqtl-wf v0.0.1_beta.2

Exp n°1

Workflow: QTL mapping for continuous traits in CC mice

Steps:

- r/qtI2 inputs creation
- Haplotype reconstruction
- Genome scan (LMM, LOCO kinship)
- Permutations-derived significance
- LOD peaks identification
- Founder allele effects estimation
- Local SNP association mapping
- List of genes in support interval

Analysis n°1



(+ RDS files)

1 workflow = 1 type of analysis allowing easy extensions (binary traits, CC-RIX, ..) & versioning

Tunable parameters (if not default):

- Marker set (Mega/GigaMUGA)
- Genome version/annotation (m38)
- CC genotype data
- Additive covariate for genome scan
- Permutations # / stratification
- Significance level / Bonferroni correction for multiple phenotypes
- LOD drop-off value

R package: ccqtl
(+/- wrapper around R/qtI2)



Take-home message

CCQTL, a comprehensive web interface for QTL mapping in the Collaborative Cross:

- Seamless procedure from data exploration to candidate genes: current practices in the field accessible to non-coders
- Absolute requirement for record tracking: Galaxy's built-in reproducibility
- Database structure: safe and organized storage of both raw phenotypic data and analysis results

Deployable anywhere:



License GPLv3

Acknowledgments

Bioinformatics and Biostatistics Hub

Victoire Baillet

Pascal Campagne

Vincent Guillemot

Rémi Planel

Rachel Torchet



Collaborative Cross initiative @ IP

Jean Jaubert

Caroline Manet

Xavier Montagutelli

Marion Rincel

Christian Voshenrich

CC-club members



Miscellaneous

Collaborative Cross & R/qrtl2 communities

CTC-RG 2023 organizing committee

More on CCQTL ?



Software Heritage



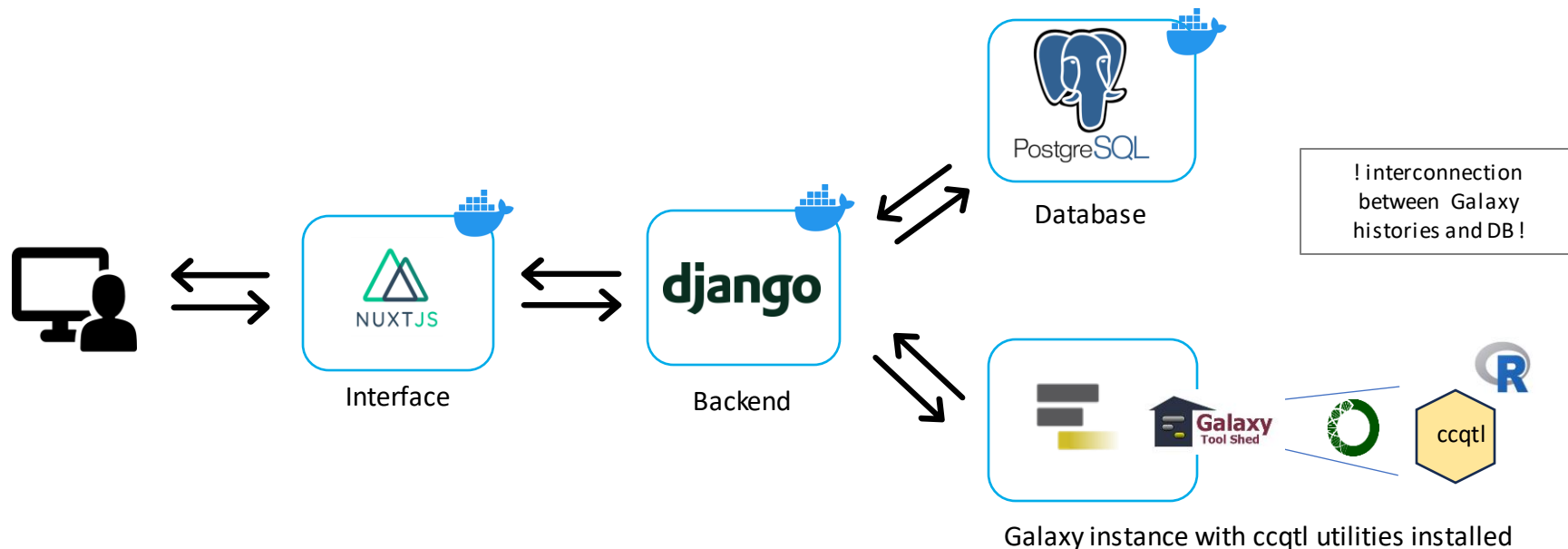
<https://gitlab.pasteur.fr/cc-qt/ccb-qt/db>

Galaxy ?

- Open-source, web-based platform for accessible, reproducible, and transparent data analysis (omics mainly)
 - **Accessible:** Users without programming experience can easily specify parameters and run tools and workflows
 - **Reproducible:** Galaxy captures information so that any user can understand and repeat a complete analysis: built-in
 - **Transparent:** Users can share or publish their analyses (histories, workflows)
- Multiple Galaxy instances (aka Galaxy servers) across the world: main servers (eg usegalaxy.org) along with 100+ additional public servers
- Ease of tools installation from one instance to the next: Galaxy Toolshed
- "Galaxy hacking": making an in-house tool accessible to a broader community by using a Galaxy server for data analysis and workflow execution, hidden behind a web interface (NGPhylogeny, ARIAweb, ...)



CCQTL's architecture guarantees its deployment anywhere



Deployment:

