ToolFinder

Discovery of bioinformatics software in an Australian infrastructure context
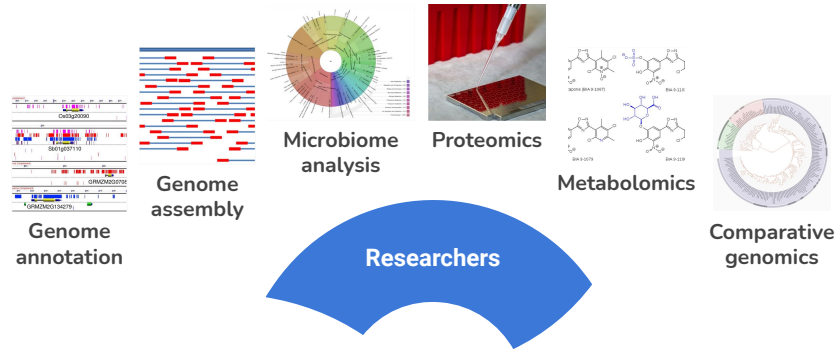
Ove J. R. Gustafsson, Jeffrey Christiansen, Tiffanie Nelson, Nigel Ward, Dale Roberts, Brian Davis, Marco De La Pierre, Tracy Chew, Dominique Gorse, Gareth Price, Andrew Lonie

Australian BioCommons

Community

democratised access to **community-supported** tools and workflows available on fit-for-purpose infrastructure
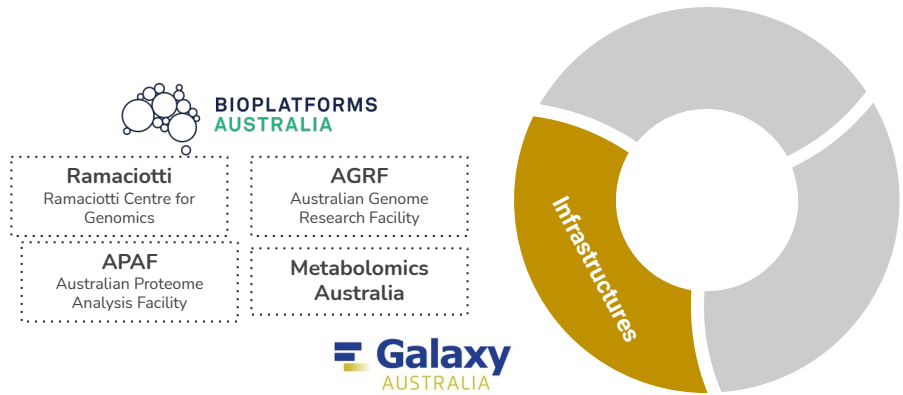
Australian
BioCommons

Community
Science drivers

Community
Infrastructures

Community

Community
Integration via BioCommons activities

Community

Increasing transparency - what's available and where?

Community
Increasing transparency - what's available and where?

*To share **learnings from development of tools** for cloud computing*

*an **index of available tools** in one place*

Researchers

Infrastructures

Science drivers

*adoption of **registry platform** to address FAIR requirements for publication*

*access to **publicly available pipelines or workflows** available on the BioCommons GitHub*

*a **shared wiki** of pipelines, workflows, sequencing platforms and other relevant information around compute.*

Australian
BioCommons

ToolFinder origin

Community engagement to ID tools used by Australian communities

→

Tools identified by communities

Australian **BioCommons**

ToolFinder origin

Tools identified by communities

Tools installed on Australian infrastructures

Engagement with infrastructure providers to ID tools that are available locally

Australian BioCommons

ToolFinder origin

Tools identified by communities

?

Tools installed on Australian infrastructures

Australian BioCommons
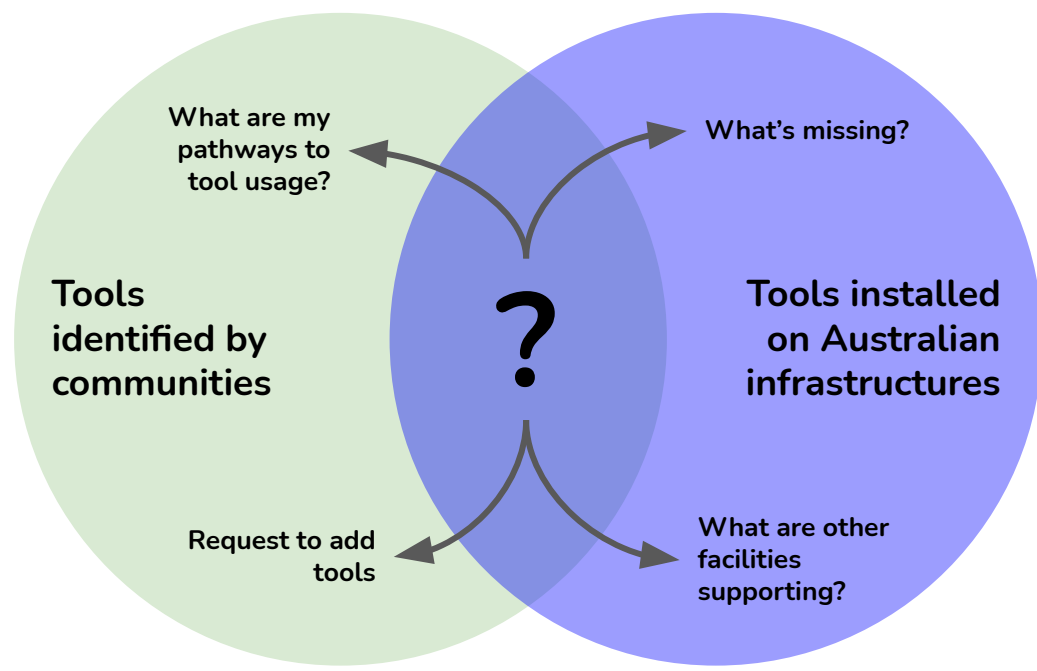
ToolFinder origin

ToolFinder is...

A **interactive, searchable & sortable list** of bioinformatics tools of interest to the Australian bioinformatics community, that

- provides information about **installed tool versions** across Australian computational infrastructures

- **displays standardised metadata** sourced from international registries

- **provides pathways** to these registries through direct links

Australian
BioCommons

ToolFinder

ToolFinder

ToolFinder

| Copy | CSV | Column visibility | | | | | | | | | | Search: | bcftools |

Accessible?

Tool metadata						Availability on Australian compute infrastructures					
Tool / workflow name	bio.tools documentation	Description	EDAM topics	Containers available? (BioContainers)	Available in Galaxy toolshed	Galaxy Australia	NCI (Gadi)	Pawsey (Zeus)	Pawsey (Magnus)	QRIScloud / UQ-RCC (Flashlite, Awoonga, Tinaroo)
BCFtools	bcftools	BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF. All commands work transparently with both VCFs and BCFs, both uncompressed and BGZF-compressed.	Genetic variation, DNA polymorphism, GWAS study, Genotyping experiment	bcftools	bcftools	Yes	1.9 1.12	1.6		1.12 1.3
HTSlib	htslib	The main purpose of HTSlib is to provide access to genomic information files, both alignment data (SAM, BAM, and CRAM formats) and variant data (VCF and BCF formats). The library also provides interfaces to access and index genome reference data in FASTA format and tab-delimited files with genomic coordinates. It is utilized and incorporated into both SAMtools and BCFtools.	Data management	htslib			1.9 1.12	1.2.1 1.6		
SAMTools	samtools	SAMtools and BCFtools are widely used programs for processing and analysing high-throughput sequencing data. They include tools for file format conversion and manipulation, sorting, querying, statistics, variant calling, and effect analysis amongst other methods.	Mapping, Sequence analysis, Sequencing, Rare diseases	samtools	samtools	Yes	1.9 1.10 1.12	0.1.19 1.6	0.1.19 1.6	0.1.17 1.10 1.2 1.3 1.6 1.9

Showing 1 to 3 of 3 entries (filtered from 466 total entries)

Previous | 1 | Next

Building ToolFinder
Primary process

Starting from scratch

Clone repository to desktop

australianbiocommons.github.io

Website for the Australian BioCommons GitHub: contains links to partner compute infrastructures, tool and workflow install status, and a help section.

TeX ☆ 2

ToolFinder generation

Australian **BioCommons**

Building ToolFinder
Primary process

ToolFinder generation

Clone repository to desktop

Download locally curated list of bioinformatics tools

Building ToolFinder
Primary process

Clone repository to desktop

ToolFinder generation

Download locally curated list of bioinformatics tools

Develop web page appearance & backend in Rstudio

Building ToolFinder
Primary process

ToolFinder generation

Clone repository to desktop

Download locally curated list of bioinformatics tools

Develop web page appearance & backend in Rstudio

Push commits to repo

Australian BioCommons

Building ToolFinder
Primary process

ToolFinder generation

- Clone repository to desktop
- Download locally curated list of bioinformatics tools
- Develop web page appearance & backend in Rstudio
- Push commits to repo
- Review content + further development

Australian BioCommons

Building ToolFinder

Community information

ToolFinder generation

Review content + further development

Clone repository to desktop

Push commits to repo

Download locally curated list of bioinformatics tools

Develop web page appearance & backend in Rstudio

Genome assembly

Proteomics

Community roadmaps

Direct engagements

BioCommons projects

Australian BioCommons

Building ToolFinder

Community information

ToolFinder generation

Review content + further development

Clone repository to desktop

Push commits to repo

Download locally curated list of bioinformatics tools

Develop web page appearance & backend in Rstudio

Community roadmaps

Direct engagements

BioCommons projects

Genome assembly

Proteomics

ToolFinder > 460 tools
bio.tools > 22000 tools

Australian **BioCommons**

Building ToolFinder
Infrastructure integration

Review content + further developments

Clone repository to desktop

ToolFinder generation

Download locally curated list of bioinformatics tools

Push commits to repo

Develop web page appearance & backend in Rstudio

EDAM ontology
EDAM is an ontology of bioscientific data analysis and data management

elixir bio.tools

NCI AUSTRALIA

pawsey

NCI apps-service

Infrastructure tool lists

Galaxy AUSTRALIA

QCIF

Australian BioCommons

ToolFinder

Copy | CSV | Column visibility

Search: bcftools

Tool metadata						Availability on Australian compute infrastructures					
Tool / workflow name	bio.tools documentation	Description	EDAM topics	Containers available? (BioContainers)	Available in Galaxy toolshed	Galaxy Australia	NCI (Gadi)	Pawsey (Zeus)	Pawsey (Magnus)	QRIScloud / UQ-RCC (Flashlite, Awoonga, Tinaroo)
BCFtools	bcftools	BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF. All commands work transparently with both VCFs and BCFs, both uncompressed and BGZF-compressed.	Genetic variation, DNA polymorphism, GWAS study, Genotyping experiment	bcftools	bcftools	Yes	1.9 1.12	1.6		1.12 1.3
HTSlib	htslib	The main purpose of HTSlib is to provide access to genomic information files, both alignment data (SAM, BAM, and CRAM formats) and variant data (VCF and BCF formats). The library also provides interfaces to access and index genome reference data in FASTA format and tab-delimited files with genomic coordinates. It is utilized and incorporated into both SAMtools and BCFtools.	Data management	htslib			1.9 1.12	1.2.1 1.6		
SAMTools	samtools	SAMtools and BCFtools are widely used programs for processing and analysing high-throughput sequencing data. They include tools for file format conversion and manipulation, sorting, querying, statistics, variant calling, and effect analysis amongst other methods.	Mapping, Sequence analysis, Sequencing, Rare diseases	samtools	samtools	Yes	1.9 1.10 1.12	0.1.19 1.6	0.1.19 1.6	0.1.17 1.10 1.2 1.3 1.6 1.9

Showing 1 to 3 of 3 entries (filtered from 466 total entries)

Previous | 1 | Next

Australian BioCommons

ToolFinder

Copy	CSV	Column visibility							Search: bcftools

						Availability on Australian compute infrastructures				
Tool / workflow name	bio.tools documentation	Description	EDAM topics	Containers available? (BioContainers)	Available in Galaxy toolshed	Galaxy Australia	NCI (Gadi)	Pawsey (Zeus)	Pawsey (Magnus)	QRIScloud / UQ-RCC (Flashlite, Awoonga, Tinaroo)
BCFtools	bcftools	BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF. All commands work transparently with both VCFs and BCFs, both uncompressed and BGZF-compressed.	Genetic variation, DNA polymorphism, GWAS study, Genotyping experiment	bcftools	bcftools	Yes	1.9 1.12	1.6		1.12 1.3
HTSlib	htslib	The main purpose of HTSlib is to provide access to genomic information files, both alignment data (SAM, BAM, and CRAM formats) and variant data (VCF and BCF formats). The library also provides interfaces to access and index genome reference data in FASTA format and tab-delimited files with genomic coordinates. It is utilized and incorporated into both SAMtools and BCFtools.	Data management	htslib			1.9 1.12	1.2.1 1.6		
SAMTools	samtools	SAMtools and BCFtools are widely used programs for processing and analysing high-throughput sequencing data. They include tools for file format conversion and manipulation, sorting, querying, statistics, variant calling, and effect analysis amongst other methods.	Mapping, Sequence analysis, Sequencing, Rare diseases	samtools	samtools	Yes	1.9 1.10 1.12	0.1.19 1.6	0.1.19 1.6	0.1.17 1.10 1.2 1.3 1.6 1.9

Showing 1 to 3 of 3 entries (filtered from 466 total entries)

Previous | 1 | Next

ToolFinder

Copy	CSV	Column visibility

						Galaxy AUSTRALIA	**NCI AUSTRALIA**	**PAWSEY**	**PAWSEY**	**QCIF**
Tool metadata										
Tool / workflow name	bio.tools documentation	Description	EDAM topics	Containers available? (BioContainers)	Available in Galaxy toolshed					
BCFtools	bcftools	BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF. All commands work transparently with both VCFs and BCFs, both uncompressed and BGZF-compressed.	Genetic variation, DNA polymorphism, GWAS study, Genotyping experiment	bcftools	bcftools	Yes	1.9 1.12	1.6		1.12 1.3
HTSlib	htslib	The main purpose of HTSlib is to provide access to genomic information files, both alignment data (SAM, BAM, and CRAM formats) and variant data (VCF and BCF formats). The library also provides interfaces to access and index genome reference data in FASTA format and tab-delimited files with genomic coordinates. It is utilized and incorporated into both SAMtools and BCFtools.	Data management	htslib			1.9 1.12	1.2.1 1.6		
SAMTools	samtools	SAMtools and BCFtools are widely used programs for processing and analysing high-throughput sequencing data. They include tools for file format conversion and manipulation, sorting, querying, statistics, variant calling, and effect analysis amongst other methods.	Mapping, Sequence analysis, Sequencing, Rare diseases	samtools	samtools	Yes	1.9 1.10 1.12	0.1.19 1.6	0.1.19 1.6	0.1.17 1.10 1.2 1.3 1.6 1.9

Showing 1 to 3 of 3 entries (filtered from 466 total entries)

Previous | 1 | Next

ToolFinder

Copy	CSV	Column visibility

Tool metadata

Tool / workflow name	bio.tools documentation	Description	EDAM topics	Containers available? (BioContainers)	Available in Galaxy toolshed	Galaxy AUSTRALIA	NCI AUSTRALIA	PAWSEY	PAWSEY	QCIF
BCFtools	bcftools	BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF. All commands work transparently with both VCFs and BCFs, both uncompressed and BGZF-compressed.	Genetic variation, DNA polymorphism, GWAS study, Genotyping experiment	bcftools	bcftools	Yes	1.9 1.12	1.6		1.12 1.3
HTSlib	htslib	The main purpose of HTSlib is to provide access to genomic information files, both alignment data (SAM, BAM, and CRAM formats) and variant data (VCF and BCF formats). The library also provides interfaces to access and index genome reference data in FASTA format and tab-delimited files with genomic coordinates. It is utilized and incorporated into both SAMtools and BCFtools.	Data management	htslib			1.9 1.12	1.2.1 1.6		
SAMTools	samtools	SAMtools and BCFtools are widely used programs for processing and analysing high-throughput sequencing data. They include tools for file format conversion and manipulation, sorting, querying, statistics, variant calling, and effect analysis amongst other methods.	Mapping, Sequence analysis, Sequencing, Rare diseases	samtools	samtools	Yes	1.9 1.10 1.12	0.1.19 1.6	0.1.19 1.6	0.1.17 1.10 1.2 1.3 1.6 1.9

Information about tools you can **use _today_**

Showing 1 to 3 of 3 entries (filtered from 466 total entries)

Previous | 1 | Next

Australian **BioCommons**

ToolFinder

The future...

UX   Integrations   Reusability   WorkflowFinder

Tool metadata						Availability on Australian compute infrastructures				
Tool / workflow name	bio.tools documentation	Description	EDAM topics	Containers available? (BioContainers)	Available in Galaxy toolshed	Galaxy Australia	NCI (Gadi)	Pawsey (Zeus)	Pawsey (Magnus)	QRIScloud / UQ-RCC (Flashlite, Awoonga, Tinaroo)
BCFtools	bcftools	BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF. All commands work transparently with both VCFs and BCFs, both uncompressed and BGZF-compressed.	Genetic variation, DNA polymorphism, GWAS study, Genotyping experiment	bcftools	bcftools	Yes	1.9 1.12	1.6		1.12 1.3
HTSlib	htslib	The main purpose of HTSlib is to provide access to genomic information files, both alignment data (SAM, BAM, and CRAM formats) and variant data (VCF and BCF formats). The library also provides interfaces to access and index genome reference data in FASTA format and tab-delimited files with genomic coordinates. It is utilized and incorporated into both SAMtools and BCFtools.	Data management	htslib			1.9 1.12	1.2.1 1.6		
SAMTools	samtools	SAMtools and BCFtools are widely used programs for processing and analysing high-throughput sequencing data. They include tools for file format conversion and manipulation, sorting, querying, statistics, variant calling, and effect analysis amongst other methods.	Mapping, Sequence analysis, Sequencing, Rare diseases	samtools	samtools	Yes	1.9 1.10 1.12	0.1.19 1.6	0.1.19 1.6	0.1.17 1.10 1.2 1.3 1.6 1.9

Showing 1 to 3 of 3 entries (filtered from 466 total entries)   Previous [1] Next

https://australianbiocommons.github.io/2_1_workflows.html

Australian
BioCommons

Thanks!

Any questions?

You can email me at: johan@biocommons.org.au