

# BayPass 2.3: tutoriel adapté au séquençage en pool

Jérôme OLIVARES<sup>1</sup>

2023-06-14

<sup>1</sup>INRAE, UR-1115 PSH, 228 route de l'aérodrome, 84914 Avignon, France

# Contents

<b>Résumé</b>	<b>3</b>
Abstract . . . . .	3
<b>introduction</b>	<b>4</b>
<b>1 Cross-references</b>	<b>5</b>
1.1 Chapters and sub-chapters . . . . .	5
1.2 Captioned figures and tables . . . . .	5
<b>2 Parts</b>	<b>7</b>
<b>3 Footnotes and citations</b>	<b>8</b>
3.1 Footnotes . . . . .	8
3.2 Citations . . . . .	8
<b>4 Blocks</b>	<b>9</b>
4.1 Equations . . . . .	9
4.2 Theorems and proofs . . . . .	9
4.3 Callout blocks . . . . .	9
<b>5 Sharing your book</b>	<b>10</b>
5.1 Publishing . . . . .	10
5.2 404 pages . . . . .	10
5.3 Metadata for sharing . . . . .	10

# Résumé

Ce tutoriel détaille d'une part la manière de générer les fichiers d'entrées du logiciel BayPass à partir de données de séquençage en pool, d'autre part le paramétrage optimal du logiciel BayPass et en fin propose une méthode d'exploration des résultats d'analyses produits. Un pipeline d'analyse mixant des packages sous Rstudio et des lignes de commandes Linux, est décrit afin de guider pas à pas l'utilisateur tout au long du processus depuis les données brutes jusqu'à la liste finale de loci/variants candidats. Ce tutoriel est à destination des étudiants et des bioinformaticiens débutants.

## mots clefs

Logiciel BayPass, séquençage en pool, GWAS, études d'associations pangénomiques, Rstudio.

## Abstract

This tutorial describes how to generate BayPass software input files from pooled sequencing data, how to optimise BayPass software settings and how to explore the analysis results produced. An analysis pipeline combining Rstudio packages and Linux command lines is described in order to guide the user step by step through the process from the raw data to the final list of candidate loci/variants. This tutorial is aimed at students and novice bioinformaticians.

Translated with [www.DeepL.com/Translator](http://www.DeepL.com/Translator) (free version)

## Keywords

BayPass software, pooled sequencing, GWAS, genome-wide association study, Rstudio

# introduction

Dans un contexte agronomique actuel de réduction de l'utilisation des pesticides ou de réchauffement climatique, analyser et comprendre les bases génétiques de l'adaptation des organismes aux méthodes de lutttes qui leurs sont opposées ou à l'évolution de leur environnement est un enjeu majeur des recherches de ces dernières années.

Les études d'associations pangénomiques (GWAS en anglais pour genome-wide association study) adossées aux techniques de séquençage haut débit (NGS) permettant le séquençage de génome complet, sont un outil de choix pour ce type d'analyse. L'étude au niveau populationnel demandant le séquençage d'un grand nombre d'individus, les couts d'analyses étaient initialement très élevés et ces études étaient souvent réservées aux organismes dit "modèles", humain en tête. Néanmoins il a été démontré depuis, que le séquençage en pool d'individus c'est-à-dire en mélangeant de manière équimolaire l'ADN d'un grand nombre d'individus (50 à 100) issus d'une même population permettait non seulement de réduire drastiquement les couts puisqu'on ne réalise qu'un seul séquençage mais aussi que la découverte des points de mutations (SNP) et l'estimation de leur fréquence allélique étaient souvent plus efficaces et précises [Futschik and Schlötterer, 2010]. Parmi les logiciels à même d'analyser ces fréquences alléliques on compte Baypass [Gautier, 2015] qui évalue par une approche bayésienne, la différenciation des SNP en liaison avec une covariable environnementale en tenant compte de la structure et de la parenté entre les populations en estimant la covariance ( $\Omega$ ) des fréquences alléliques. Baypass 2.3 a la particularité supplémentaire de pouvoir calculer un contraste des fréquences d'allèles SNP entre deux groupes de populations spécifiés par un caractère binaire ; sensible ou résistant par exemple.

La documentation disponible de BayPass 2.3 décrit par le menu les algorithmes de fonctionnement et les différents paramètres du logiciel mais reste néanmoins, et assez logiquement, succincte sur les étapes en amont et en aval. A notre connaissance un seul tutoriel est disponible sur le web [Nielsen, 2020] et décrit de manière plus détaillée la préparation des données brute et l'analyse en mode « poolseq » de Baypass 2.3., mais il nécessite des bases quelque peu avancées de codage sous R et en environnement bash. Si les bio-informaticiens chevronnés ne rencontreront pas de difficultés particulières, il n'en va pas forcément de même pour bon nombre d'agents que les orientations des recherches associées à la baisse des coûts de séquençage ont aiguillé vers les voies de la génomique. Cet article a pour but de détailler les différentes étapes d'une analyse de type GWAS/poolseq avec le logiciel BayPass 2.3 et d'éclairer les points qui sont habituellement peu explicités car considérés comme évident et se veut, au final, le plus proche possible de la citation de Talleyrand : « Si cela va sans le dire, cela ira encore mieux en le disant ».

# Chapter 1

## Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

### 1.1 Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
  - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world = # Hello world {#hello-world}`.
  - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.
2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter 1.
  - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

### 1.2 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 1.1.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Don't miss Table 1.1.

```
knitr::kable(
  head(pressure, 10), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

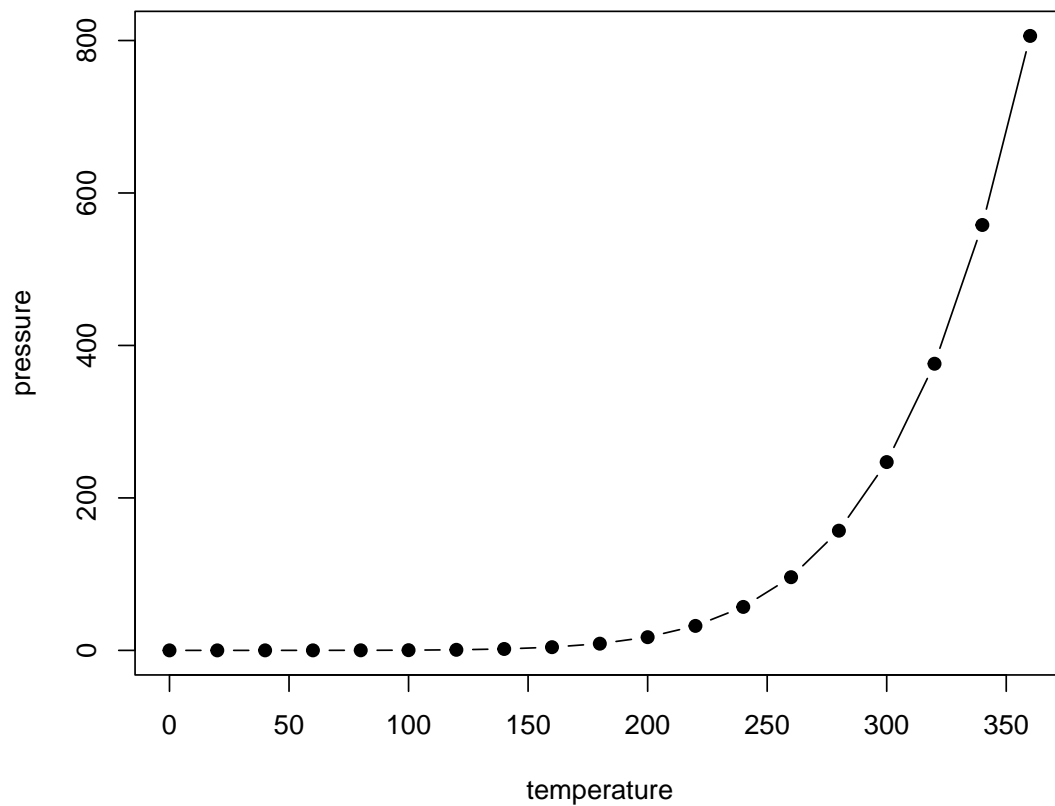


Figure 1.1: Here is a nice figure!

Table 1.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

## Chapter 2

# Parts

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART\*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

## Chapter 3

# Footnotes and citations

### 3.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one <sup>1</sup>.

### 3.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [?] (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [?] (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

---

<sup>1</sup>This is a footnote.



# Chapter 4

## Blocks

### 4.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (4.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (4.1).

### 4.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 4.1.

**Theorem 4.1.** *For a right triangle, if  $c$  denotes the length of the hypotenuse and  $a$  and  $b$  denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

### 4.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

## Chapter 5

# Sharing your book

### 5.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

### 5.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

### 5.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book- all links shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

<https://pkgs.rstudio.com/bookdown/reference/gitbook.html>

Or use:

```
?bookdown::gitbook
```

# Bibliography

- Andreas Futschik and Christian Schlötterer. The Next Generation of Molecular Markers From Massively Parallel Sequencing of Pooled DNA Samples. *Genetics*, 186(1):207–218, September 2010. ISSN 1943-2631. doi: 10.1534/genetics.110.114397. URL <https://academic.oup.com/genetics/article/186/1/207/6063740>.
- Mathieu Gautier. Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, 201(4):1555–1579, December 2015. ISSN 1943-2631. doi: 10.1534/genetics.115.181453.
- Erica S. Nielsen. Pool-Seq Analyses: PoolFstat & BayPass, 2020. URL <https://esnielsen.github.io/post/pool-seq-analyses-poolfstat-baypass/>.