

# DEploid Documentation

*Release DEploid-py*

Sha (Joe) Zhu

Nov 15, 2018

---

## Contents

---

<b>1</b>	<b>Description</b>	<b>2</b>
<b>2</b>	<b>Installation</b>	<b>4</b>
<b>3</b>	<b>Reporting Bugs</b>	<b>5</b>
<b>4</b>	<b>Citing DEploid</b>	<b>6</b>
<b>5</b>	<b>dEploid</b>	<b>8</b>
5.1	mcmc module . . . . .	8
5.2	vcf module . . . . .	8
	<b>Bibliography</b>	<b>9</b>

A software that deconvolutes mixed genomes with unknown proportions.

# CHAPTER 1

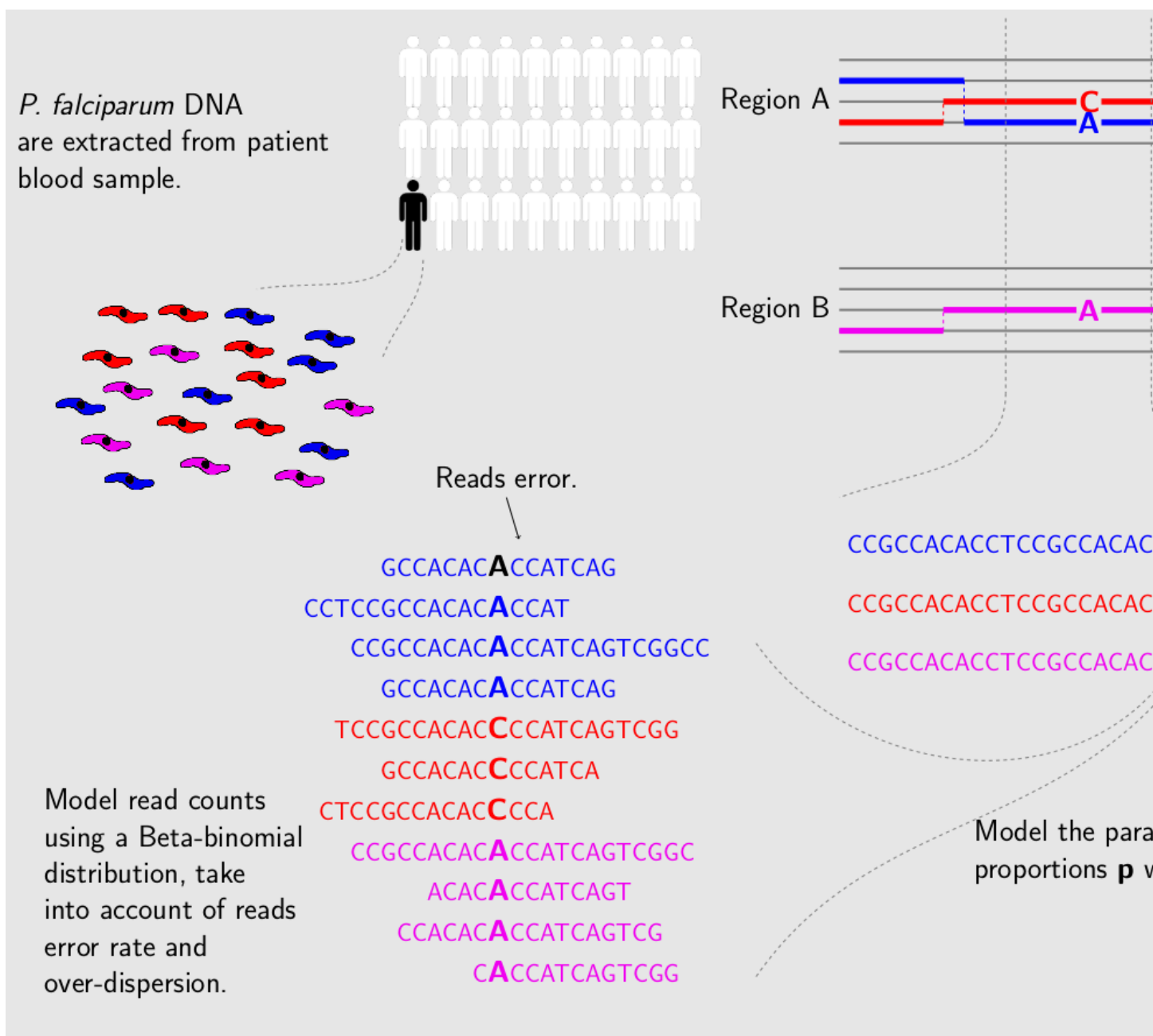
---

## Description

---

dEpl<sub>oid</sub> is designed for deconvoluting mixed genomes with unknown proportions. Traditional ‘phasing’ programs are limited to diploid organisms. Our method modifies Li and Stephen’s [\[Li2003\]](#) algorithm with Markov chain Monte Carlo (MCMC) approaches, and builds a generic framework that allows haplotype searches in a multiple infection setting.

dEpl<sub>oid</sub> is primarily developed as part of the Pf3k project, from which this documentation will take examples from for demonstration. The Pf3k project is a global collaboration using the latest sequencing technologies to provide a high-resolution view of natural variation in the malaria parasite *Plasmodium falciparum*. Parasite DNA are extracted from patient blood sample, which often contains more than one parasite strain, with unknown proportions. dEpl<sub>oid</sub> is used for deconvoluting mixed haplotypes, and reporting the mixture proportions from each sample.



## CHAPTER 2

---

### Installation

---

```
$ pip install dEploid
```

## CHAPTER 3

---

### Reporting Bugs

---

If you encounter any problem when using `dEpl0id`, please file a short bug report by using the [issue tracker](#) on GitHub or email `joe.zhu (at) well.ox.ac.uk`.

Please include the output of `dEpl0id -v` and the platform you are using `dEpl0id` on in the report. If the problem occurs while executing `dEpl0id`, please also include the command you are using and the random seed.

Thank you!

## CHAPTER 4

---

### Citing DEploid

---

If you use dEploid with the flag `-ibd`, please cite the following paper:

Zhu, J. S., J. A. Hendry, J. Almagro-Garcia, R. D. Pearson, R. Amato, A. Miles, D. J. Weiss, T. C. D. Lucas, M. Nguyen, P. W. Gething, D. Kwiatkowski, G. McVean, and for the Pf3k Project. (2018) The origins and relatedness structure of mixed infections vary with local prevalence of *P. falciparum* malaria. *bioRxiv*, doi: <https://doi.org/10.1101/387266>.

Bibtex record::

```
@article {Zhu387266,
author = {Zhu, Sha Joe and Hendry, Jason A. and Almagro-Garcia, Jacob and Pearson,
↪Richard D. and Amato, Roberto and Miles, Alistair and Weiss, Daniel J. and Lucas,
↪Tim C.D. and Nguyen, Michele and Gething, Peter W. and Kwiatkowski, Dominic and
↪McVean, Gil and },
title = {The origins and relatedness structure of mixed infections vary with local
↪prevalence of P. falciparum malaria},
year = {2018},
doi = {10.1101/387266},
publisher = {Cold Spring Harbor Laboratory},
URL = {https://www.biorxiv.org/content/early/2018/08/09/387266},
eprint = {https://www.biorxiv.org/content/early/2018/08/09/387266.full.pdf},
journal = {bioRxiv}
}
```

If you use dEploid in your work, please cite the program:

Zhu, J. S. J. A. Garcia G. McVean. (2017) Deconvolution of multiple infections in *Plasmodium falciparum* from high throughput sequencing data. *Bioinformatics* btx530. doi: <https://doi.org/10.1093/bioinformatics/btx530>.

Bibtex record::

```
@article {Zhubtx530,
author = {Zhu, Sha Joe and Almagro-Garcia, Jacob and McVean, Gil},
title = {Deconvolution of multiple infections in Plasmodium falciparum from
↪high throughput sequencing data},
```

(continues on next page)



(continued from previous page)

```
year = {2017},  
doi = {10.1093/bioinformatics/btx530},  
URL = {https://doi.org/10.1093/bioinformatics/btx530},  
journal = {Bioinformatics}  
}
```

## CHAPTER 5

---

dEpl0id

---

### 5.1 mcmc module

### 5.2 vcf module

---

## Bibliography

---

- [Li2003] Li, N. and M. Stephens (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165(4), 2213–2233.