

# Paper\_Review

Nathan

July 29, 2016

## Contents

<b>1</b>	<b>I recognize but not 100% understand (AKA squishy bio-stuff)</b>	<b>2</b>
<b>2</b>	<b>What is this? (AKA stuff I haven't heard off)</b>	<b>2</b>
<b>3</b>	<b>Grammar?</b>	<b>2</b>
3.1	Page 5 . . . . .	2
3.2	Page 6 . . . . .	2
3.3	Page 8 . . . . .	2
3.4	Page 10 . . . . .	3
3.5	Page 11 . . . . .	3
<b>4</b>	<b>General thoughts</b>	<b>3</b>
4.1	Page 5 . . . . .	3
4.2	Page 6 . . . . .	3
4.3	Page 8 . . . . .	4
4.4	Page 9 . . . . .	4
4.4.1	More recent methods... . . . .	4
4.5	Pages 17-20 . . . . .	4
<b>5</b>	<b>Other</b>	<b>4</b>
5.1	General question on bacteria . . . . .	4
5.2	Page 6 . . . . .	5

## 1 I recognize but not 100% understand (AKA squishy bio-stuff)

- Haplotypes (I assume this is a pretty straightforward thing for most biologists though right?)

## 2 What is this? (AKA stuff I haven't heard off)

- Nucleotide
- DHFR & AIMP1 I've seen these mentioned but might just be going over my head
- Same as above but Bowtie2?
- Haplotype (I put this in twice because it's mentioned so much and changed what I thought I knew and is mentioned in more detail after a while)

## 3 Grammar?

### 3.1 Page 5

1. Diversity present in the enzymes and small peptides in microbial communities . . . This entire sentence is a little awkward to read, something more like: "Diversity in the enzymes and small peptides found in microbial communities are a valuable. . . "

### 3.2 Page 6

1. The problem of haplotype discover . . . is a post-assembly problem  
Could be worded better, feels odd to read
2. There's a [?] that has no references yet to software that can perform these functions

### 3.3 Page 8

There's a few areas on this page that have ". . . " sections that I don't really understand these areas :s But I assume this is because they're unfinished so no need to worry

### 3.4 Page 10

- "Markov chain models .. have been proposed in 2006" Is the wording alright for that, wouldn't "were proposed by Wang in 2006..."

### 3.5 Page 11

- "The evidence on those reads" could make better sense as "The evidence on these reads"?
- Silly little things in the bullet points like not having caps or periods on the sentences.
- When talking about N haplotypes, possibly change it to the N symbol to stand out a bit more?
- What is the defining feature that Hansel and Gretel use to determine "too small to be useful" is it a hard % of data. If so

how do you know it's safe to ignore / even consider in the first place

## 4 General thoughts

I might be wrong, but doesn't anything between two ',' need to not be a requirement for the before and after parts of the sentence?

### 4.1 Page 5

- About midway through the page there is a bit about assembly of reads from metagenomes, another thing I'm most likely wrong about /

misunderstanding but it isn't clear if creating a reference from a metagenome refers to making one metagenomic sample a reference for other unique samples or does the sample be used for itself?

### 4.2 Page 6

- Gene-level variation in genomes... the margin of error given for contigs, is it worth giving a % example here of the kind of accuracy you aim for regularly?

Especially if it is a low % it would be good to have more figures!

- Last sentence "Most existing haplotyping... for single-species" isn't this a little bit repeated with what's written a few paragraphs ago

"Most assemblers are designed for single species genomes ... and aim to produce a single sequence". Might be worth while rewording slightly just to avoid sounding like repeating. I do realise it's slightly different but sounds very similar.

### 4.3 Page 8

- Is it worth while expanding upon / reworking the  $P(h_1, h_2 \mid M)$  so that it's easier to read?

All the data is there for the function but has explanation before and after?  
I could just be nitt picky on this

- Maybe I missed something but last paragraph talks about "Wang" but there doesn't seem to be a mention of them before or decent explanation of their work.

Good chance that I'm wrong about this so perfectly happy for you to tell me I'm wrong on this

### 4.4 Page 9

#### 4.4.1 More recent methods...

This section is really nice to read and all of the stuff makes sense (to me, which is saying something)

### 4.5 Pages 17-20

Explanations here are really good, the introduction of the diagrams really clear up. Picture == 1000 words and all that

## 5 Other

### 5.1 General question on bacteria

Just a thought. Rumen has a load of stuff floating around and I was under the impression that these bacteria have a cool feature in that they can absorb/copy beneficial blocks of DNA from others. So is this something that affects these regions of interest?

## 5.2 Page 6

1. Gene level variation in metagenomes This section is really nice as it uses brackets to refer to stuff. It's just a bit of contrast to some other sections of the paper, where there is some lack of clarification. It goes from talking about the metahaplome with no explanation to saying that DNA sequences are reads?