

From: Lau, Matthew K. matthewklau@fas.harvard.edu
Subject: Re: ant genomes
Date: February 14, 2017 at 5:15 PM
To: Caroline Cusick ccusick@broadinstitute.org
Cc: James Bochicchio jboch@broadinstitute.org



Great, I also just saw that Sarah emailed me with some suggestions as well.

Thanks!

Matt

On Feb 14, 2017, at 10:27 AM, Caroline Cusick
<ccusick@broadinstitute.org> wrote:

Hi Matt,

I heard back from our analysts and they said the following,

"You're on the right track with the GATK website. Also try the forums (<http://gatkforums.broadinstitute.org/gatk/>), there are people that maintain them and they can be very helpful. It will be a process to figure out how to run GATK for your particular organism. I would suggest that you call SNPs genome-wide first, and then look to see where they fall to dive into individual genes. Also, it's important that you pick your best reference genome, and align data to that for analysis purposes. By best, I typically recommend the genome that looks most complete (maybe the largest?). Also, I should point out, GATK is mostly supported for human, so whenever one wants to utilize it for an alternative genome it's always a process to figure out the exact optimal parameters. The GATK people should be able to help out. Alternatively, I know you mentioned this but I forget the answer, if your data is haploid, then you could try using Pilon to call SNPs: <https://github.com/broadinstitute/pilon/wiki>."

let me know if you have any other questions!

thanks,
Caroline

On Tue, Feb 14, 2017 at 9:10 AM Caroline Cusick

<ccusick@broadinstitute.org> wrote:

Hi Matt,

Happy belated new year to you too!

I'm going to pass your question on to our analyst (Terry) and will let you know his response as soon as I get it.

thanks!

Caroline

On Wed, Feb 8, 2017 at 10:23 AM Lau, Matthew K.

<matthewklau@fas.harvard.edu> wrote:

Hi Caroline, happy much belated New Year. A couple of hopefully quick things:

1) I've finished moving and backing up the raw files for the genomes. So, if you need the space on the server, feel free to free it up.

2) I've been going over the sequence quality stats, and I'm starting to get into doing some comparative analysis of the sequences. I started looking at the GATK pages and the Genome analysis sections were listed as under-development. Can you point me in the right direction of info for using GATK to search for target genes and for comparing my genomes (e.g. genetic distances)?

Thanks,

Matt

