

ddRAD Phase 2

Gus Dunn

2015-04-30

1 Main costs and limiting factors

After talking with AndreaG and JoshuaR, I believe that the main cost determinate will simply be the sequencing. There is plenty of the special primer material still available which would be the other main cost. The reagents, as the PLOS paper points out are very cheap: NEB enzymes/buffers plus high-fi Taq reagents.

The current cost to do a similar type of sequencing strategy according to the [YCGA website](#) is \$1887 for HiSeq 2000 or 2500: 75 bp, paired-end sequencing (1 lane, 2 x 75).

2 How many individuals to do

AndreaG did 24 per lane, and according to the results in our current paper:

Downstream analysis of the ddRAD reads indicates that we achieved an average of 32X coverage for each individual fly.

I am going to refer to the depth-coverage per individual as DCI from now on.

Our paper doesn't yet mention how many RAD-regions (genomic chunks) we got. It probably should as this is talked about in the PLOS ddRAD paper and is related to the proportion of the genome we can *see* in our data. I asked AndreaG, and she didn't remember of the top of her head but was going to look it up. However, since we do not intend to deviate from her protocol, I only bring it up in the interest of completeness.

At current prices and using AndreaG's 24 individuals per plate, 5 lanes would cost \$9,435 and yield 120 individuals. If we could tolerate lower than 32X DCI, we can expand the number of individuals. Assuming a linear relationship between individuals per lane and DCI: allowing our DCI to drop to 19.2X (40 flies per lane) would increase our yield to 200 individuals: a DCI of 15.36X (50 flies per lane) yields 250 individuals. And of course adding another lane would bump that to 300 and change the cost to \$11,322.

Since we want to shoot for at least ~100 infected individuals, we should plan to sequence *at least* ~200 flies. We can do that easily as long as we can tolerate DCI ~15X and/or add an additional lane(s).

3 Minimum depth-coverage per individual

I am still looking for a theoretical minimal DCI threshold needed in order to draw solid conclusions from our data. I am opening up correspondence with Deren Eaton (pyrad – upstairs neighbor) and probably Brant K. Peterson (corresponding author on the PLOS paper) regarding this. However, I do not think that 15X is very close to this limit. Especially since we will have $\sim 15X * 200$ or more individuals when it is all said and done.

4 Planing collecting efforts

I am worried that our existing specimens stores from the south may not represent comparable breadth or depth as compared to our existing specimens from the northern and to some degree western stores, particularly when considering infections. I think that as the project continues we need to insist on some level of reconnaissance protocols (regularly executed low-volume surveying expeditions that cover a wide swath of our population landscapes) that will allow us to *swiftly (in the same season)* deploy high-volume collection expeditions to regions where sufficient infections are detected.

I am double checking our numbers, but I would currently focus that type of reconnaissance in the north-west and south and send high-volume expeditions to the one that looks most promising for this season. Then repeat or adjust recon focus-regions for Fall according to the results of the Summer.

5 ddRAD phase 2 experimental goals

Once we decide how many flies we plan to sequence, I will propose to you the populations and fly numbers to represent each to make sure that the project's goals are covered. From there, I will give you a set of milestones and completion goals. After I run through the process in my own hands and get the preliminary quality control results from the single and double digestion trails, I will adjust the completion goals based on real-life.