Notes on the fly project

J. Ignacio Lucas Lledó

December 14, 2016

1 Delivery of oligonucleotides

The oligonucleotides designed on 2016-06-09 were ordered to biomers.net, and delivered. They are dry. Table 1 specifies the yield and the corresponding amount of buffer required to dilute them to 100 pmol/μl.

2 DNA extraction test

On September 13th 2009, Dr. Carazo, Mrs. Sultanova, Mr. Andiç and I started the DNA extraction from 8 flies, following the GBS protocol. Table 2 summarizes the procedure. Not that we needed 4 people to run this, but they all were interested in going through the protocol, which is great. The protocol is applied to each fly individually, with the aim of measuring how much DNA we can get from one fly. Four of the flies had their oviduct previously removed by Mrs. Sultanova. We want to make sure that the ovariectomy does not significantly reduce DNA yield.

We ended up measuring DNA concentration 3 times, due to problems with the Qubit standards. The error 'Incorrect standards' was apparently caused by the fact that the reagent had been kept in the fridge. I thought that it would be enough to wait for it to thaw, to bring it to room temperature. Plus, it is a very small amount of it that gets dissolved in the buffer, at room temperature. For whatever reason, the kit worked once we kept the reagent at room temperature. I thank Rebeca Dominguez, from the Evolutionary Genetics lab, her help to solve the problem.

The slight difference in average DNA concentration between control and ovariectomy samples is not significant, which is good news. I expect three samples, namely 1, 3, and 6, to have enough DNA left to run the rest of the protocol with them. I could add sample 5 to see what we get from it.

Table 1: Name and available amount (yield) of oligonucleotides received from biomers.net. 'Volume' makes reference to the volume required to dissolve the corresponding oligonucleotides at 100 pmol/μl (μM).

Name	Yield (nmol)	Volume (μl)	Name	Yield (nmol)	Volume (μl)
Top1.1	82.6	826	Bot1.1	64.8	648
Top1.2	106.8	1068	Bot1.2	110.4	1104
Top1.3	94.9	949	Bot1.3	92.6	926
Top1.4	85.6	856	Bot1.4	75.8	758
Top2.1	78.9	789	Bot2.1	90.7	907
Top2.2	81.9	819	Bot2.2	97.4	974
Top2.3	100.5	1005	Bot2.3	84.2	842
Top2.4	88.4	884	Bot2.4	82.3	823
Top3.1	98.6	986	Bot3.1	87.1	871
Top3.2	95.2	952	Bot3.2	96.6	966
Top3.3	100.2	1002	Bot3.3	87.9	879
Top3.4	95.6	956	Bot3.4	78.7	787
A1.1	60.2	602	A2	65.9	659
A1.2	52.6	526	R1	78.7	787
A1.3	56.7	566	R2	60.0	600
A1.4	51.0	510			

Table 2: Summary of the DNA extraction test.

Sample	Treatment	Elution (μl)	Conc.(ng/μl)	Yield (ng)
1	–	100	4.240	424
2	–	100	0.978	98
3	–	100	3.600	360
4	–	100	1.790	179
5	ovariectomy	100	1.990	199
6	ovariectomy	100	3.980	398
7	ovariectomy	100	1.140	114
8	ovariectomy	100	1.530	153

3 Literature review

Original motivation of classic life span QTL mapping studies was the elucidation of metabolic pathways involved in aging [Nuzhdin et al., 2005]. I suspect that the biomedical interest in the mechanisms of aging preceded the interest in naturally occurring variation, as well as the real effect of the QTL identified in natural conditions. Several sex-specific life span QTL exist in autosomes of *D. melanogaster* [Nuzhdin et al., 1997, Pasyukova et al., 2000]. These QTL were originally identified using recombinant inbred lines (RIL) descended from two strains [Nuzhdin et al., 1997]. In this design, the dominance of the QTL is not known, since only homozygous genotypes are produced [Lynch and Walsh, 1998, page 432]. However, the quantitative complementation test with deficencies used to refine the QTL [Pasyukova et al., 2000] seems to assume that one of the alleles is at least partially recessive.

Crow [2008] mentions the genetic basis of heterosis as one of the main mid-century controversies in population genetics. He is very clear that the debate between dominance and overdominance was resolved in favor of the dominance hypothesis. That is, both the inbreeding depression and the hybrid vigor (heterosis) are mostly due to recessive deleterious alleles that become either expressed upon inbreeding or concealed upon hybridization. A review [Charlesworth and Willis, 2009] cites another work by James F. Crow that I cannot access (volume 9 of the Oxford Series in Evolutionary Biology), in which the author shows evidence of 30% of autosomes isolated from natural populations of *D. melanogaster* being lethal in homozygosis. The same cannot be true for sex chromosomes, of course. Among the 70% of natural chromosomes that are not recessive lethals, homozygosity reduces survival to adulthood in 84%.

Full-sib mating during 10 generations is expected to produce an inbreeding coefficient of 0.886. This is the probability at any locus that the two copies inherited from the parents are identical by descent; that is, they are exact copies of the same chromosome in one of the recent ancestors. But how is this identity-by-descent distributed along a chromosome? It is not random, but clustered, because of linkage. We expect an alternative sequence of fragments that are either identical by descent or not. That is, the inbreeding itself already produces a mosaic of heterozygosity and homozygosity. The question is how long the pieces are. Too short IBD tracts would require very dense markers. According to Franklin [1977], previous authors (Bennet and Fisher) had supposed that 'after a long number of generations of inbreeding the number of heterogenic segments will be approximately distributed as a Poisson variable, and that the distribution of the length of each segment (x) will tend to a negative exponential $(1/a)e^{-x/a}$.' However, this assumptions

Table 3: Digestion reactions.

Sample	DNA (μl)	H_2O (μl)	10× buffer (μl)	NspI (μl)	Total (μl)
1	80	207	32	1	320
3	80	207	32	1	320
5	80	207	32	1	320
6	80	207	32	1	320

are not satisfactory, especially for the first few generations of inbreeding. Taking into account that there is no crossing-over in *Drosophila males*, the probability that two loci in the same chromosome are homozygous by descent after one generation of sib mating is $(1/32)(3+3\lambda_f+\lambda_f^2+\lambda_f^3)$ [Franklin, 1977], where λ_f is the linkage parameter $1-2y_f$, representing the probability of recombination in the female.

The suggestion given by Dr. Carazo of using simulations is actually good. I started learning to use simuPOP.

4 Enzymatic digestion of DNA samples 1, 3, 5, and 6

I run the digestions as described on table 3, for above 4 hours, at 37°C. Together with Muhammed Andiç, we also cleaned up the reaction with magnetic beads. We used a 1:1 ratio of beads to DNA, and eluted with water (40 μl).

5 Literature review

From page 489 of White [1973]: "The adaptive significance of achiasmatic meiosis is not entirely clear. Obviously it has been evolved repeatedly in at least ten major groups of animals. In those cases where achiasmatic meiosis is confined to one sex, we have far too little information as to the chiasma frequency in the other sex. If it were generally true that in the 'chiasmatic sex' there is only one chiasma per bivalent, we might assume that selection for a low level of genetic recombination *per se* had led to total loss of chiasmata in one sex. If, on the other hand, absence of chiasmata in one sex is generally accompanied by a fairly high chiasma frequency in the other, this

explanation obviously would not hold. Only evidence from species that had acquired achiasmatic mechanisms fairly recently in their phylogeny would be significcant from this point of view (*Tipula caesia* would be relevant – *Drosophila* spp. would not). The only evidence on this point seems to be the grasshopper *Thericles whitei*, with no chiasmata in the male and a very low chiasma frequency in the female; it obviously supports the hypothesis of selection for low levels of recombination."

However, selection for low levels of recombination would not explain why the achiasmatic sex is nearly always the heterogametic sex [Burt et al., 1991]. White [1973] also mentions the hypothesis of achiamatic spermatogenesis being an adaptation to tolerate inversion-heterozygosity. However, no evidence supports this hypothesis.

More recently, Lenormand et al. [2016] mention two more hypotheses: first, that achiasmy may have evolved as a side effect of selection to suppress recombination between the sex chromosomes Haldane [1922], Huxley [1928]. And second, that it evolved as a way to promote tight linkage without suppressing recombination on the X or Z chromosome [Lenormand and Dutheil, 2005].

Well, I cannot review now all the hypotheses now. If anything, I should relate the results of the simulations to the theoretical expectations of the variance in inbreeding coefficient [Cockerham and Weir, 1968, Franklin, 1977]. It seems that multiple linked loci in sex chromosomes have received relatively little attention in the literature [Bennett, 1963, Owen, 1988].

6 Annealing of adapters

On September 28th, Muhammed and I prepared 100 ml of 1× TE buffer, following a recipe online: 1.0 ml 1M Tris pH 8.0, 0.2 ml EDTA 0.5M, and 98.8 ml H_2O. We also prepared annealing buffer, according to the protocol, and suspended all the oligonucleotides in TE buffer to 100 μM, according to the volume specified on table 1. Finally, we prepared 10 μl of annealed adapters of each type.

7 Quantification with Agilent Bioanalyser

Amparo Martínez, from the Sequencing Service, sent the results from the Bioanalyser run. The four samples have a very similar profile (Figure 1), which looks like the result of a successful digestion followed by a more aggressive selection of long fragments that I expected from a 1:1 ratio of magnetic beads to

Table 4: Distribution of DNA fragments between 40 and 9000 bp. 'Size' refers to the average fragment size. For total mass of DNA left over, a volume of 35 μl is assumed to be left. The average fragment mass (Frag.Mass) is calculated in two ways. First, by dividing the overall concentration by the overall molarity (Frag.Mass.1); and then, by applying the formula for the molecular mass of double stranded DNA to the average length of the fragments (i.e., $607.4x + 157.9$ g/mol; Frag.Mass.2).

Sample	Size (bp)	Conc. (pg/μl)	Molarity (pmol/l)	Left over (ng)	Frag.Mass.1 (ng/fmol)	Frag.Mass.2 (ng/fmol)
1	2206	1425.92	3042.0	49.91	0.468744	1.34008
3	2239	1439.86	2484.9	50.40	0.579444	1.36013
5	2022	1113.97	3509.8	38.99	0.317388	1.22832
6	2376	1523.30	2066.7	53.32	0.737069	1.44334

DNA. Next time, I should use a more generous proportion of beads. Table 4 shows some additional details. I note that the amount of DNA available is now an order of magnitude lower than what I had for the *Coregonus* samples.

8 Ligation of adapters

In folder 2016-10-04, I calculate the composition of the ligation reactions to have a 10-fold excess of adapters to DNA fragment ends. On table 5 I show the composition and final concentrations of the working solutions of the four adapters, required to prepare the ligation reactions specified on table 6. Importantly, I decide to assign adapters 1.1, 1.2, 1.3, and 1.4 to samples 1, 3, 5, and 6, respectively. See table 7.

Just to have a graphic representation of the adapters and the expected DNA fragments, this is how they should look like after ligation (adapter 1.1 represented):

```
ACACTCTTTCCCTACACGACGCTCTTCCGATTGATCCAGTCATGYNNNNNNNNNNRCATGACTGGATCAATCGGAAGAGCACACGTCTGAACTCCAGTCAC
   |||      |       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||       |      |||
CACTCTTTCCCTACACGACGCTCTTCCGATTGATCCAGTCATGYNNNNNNNNNNRCATGACTGGATCAATCGGAAGAGCACACGTCTGAACTCCAGTCAC
```

The codeword is in red, and the genomic fragment in blue. Note that the restriction site, RCATGY, is not present in the ligation product (R is G or

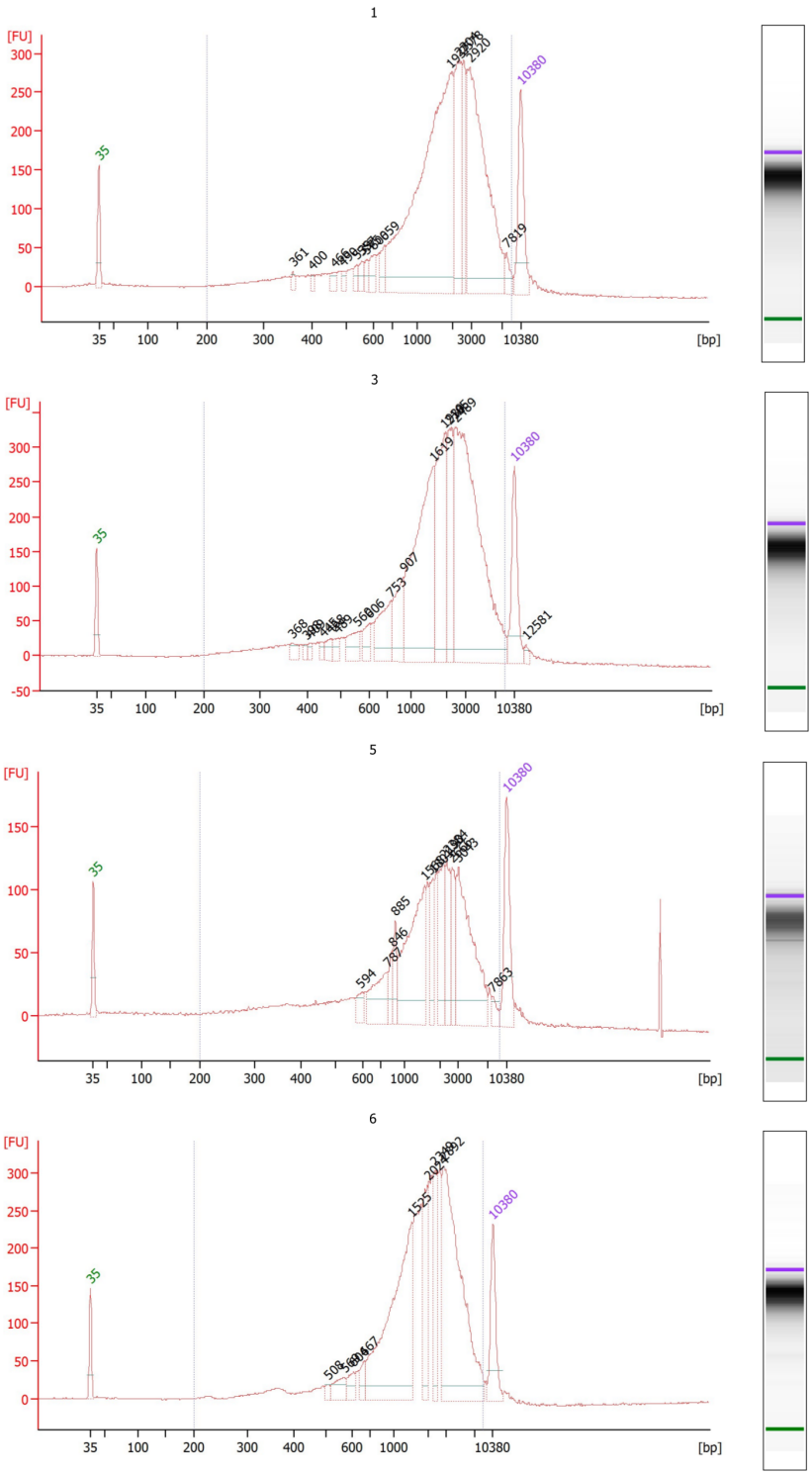


Figure 1: Electropherograms from the high-sensitivity DNA assay of the four samples.

Table 5: Composition and final concentration of the working solutions of adapters for the ligation reactionson table 6. The buffer must be 1× annealing buffer. Amounts are μl, except the final molarity, in the last column, in pmol/μl.

Sample	Adapter	1× buffer	15 μM Stock	Total	pmol/μl
1	1.1	8.580	1.420	10.00	2.13
3	1.2	8.840	1.160	10.00	1.74
5	1.3	8.362	1.638	10.00	2.46
6	1.4	9.036	0.964	10.00	1.45

Table 6: Volumes of reactants in the ligation reactions. All in μl. T4 ligase is at 2000000 U/ml.

Sample	DNA	10× Buffer	Adapter	1.5 M NaCl	T4	H_2O	Total
1	35	4.40	1.00	1.10	2.00	0.50	44.00
3	35	4.40	1.00	1.10	2.00	0.50	44.00
5	35	4.40	1.00	1.10	2.00	0.50	44.00
6	35	4.40	1.00	1.10	2.00	0.50	44.00

Table 7: Adapters and codewords assigned to the DNA samples.

Sample	Adapter	Codeword
1	1.1	TTGATCCAGT
3	1.2	GATCAGGCAGT
5	1.3	CCAGCTTGT
6	1.4	AGCTGAAT

Table 8: Current composition of the ligation reaction, after necessary adjustments. All amounts are in μl. T4 ligase was actually at 400000 U/ml.

Sample	DNA	10× Buffer	Adapter	1.5 M NaCl	T4	H_2O	Total
1	35	4.62	1.00	1.10	4.00	0.50	46.22
3	35	4.62	1.00	1.10	4.00	0.50	46.22
5	35	4.62	1.00	1.10	4.00	0.50	46.22
6	30	4.02	1.00	1.10	4.00	0.25	40.37

A, and Y is C or T). This allows me to digest a second time, and get rid of chimeric fragments.

On October 5th, I prepared the working solution of the annealed adapters, and the following day I set up the ligation reaction. I then realized the available T4 DNA ligase was at 400000 U/ml, instead of 2000000 U/ml. I decided to double the amount of T4 ligase, and adjust the amount of buffer. I also noticed that sample 6 did not have more than 30 μl of DNA. Thus, I changed the composition of the reaction accordingly. However, I note that I used 1 μl of adapters at a slightly higher concentration than necessary to achive the 10-fold excess, because the working solution of adapters was already prepared, and I didn't want to re-calculate the whole thing. See table 8 for the current composition of the ligation reaction.

9 Second digestion

The second digestion is meant to break chimeric fragments. The adapters will not be cleaved because they were designed not to reproduce the cut site. I do not want to clean up the ligation reaction. In the worse case, I will not get rid of the chimeras. But I don't want to loose DNA. So, I count the total ligation volumes (table 8) as DNA volumes in the second digestion reactions specified on table 9.

10 Size selection

I plotted the expected shape of the electropherogram, using data from the in silico digestion with the pattern RCATGY (2016-04-12). The result, shown in figure 2, is remarkably similar to the observed electropherograms after digestion (figure 1).

In any case, from those figures it is not immediate to know the number

Table 9: Second digestion reactions.

Sample	DNA (μl)	H_2O (μl)	10$\times$ buffer (μl)	NspI (μl)	Total (μl)
1	46.22	119.17	18.49	1	184.88
3	46.22	119.17	18.49	1	184.88
5	46.22	119.17	18.49	1	184.88
6	40.37	103.96	16.15	1	161.48

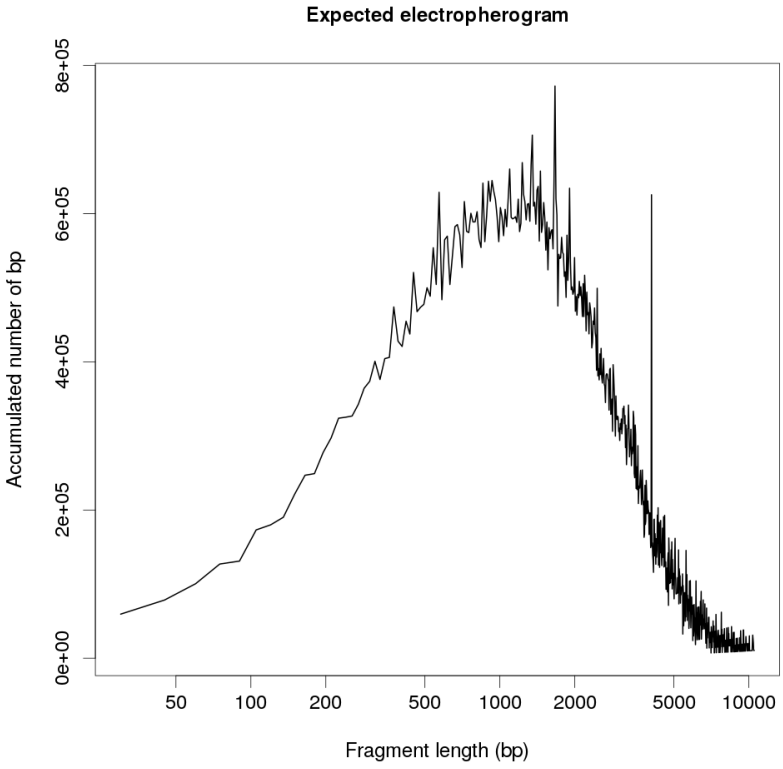


Figure 2: Expected shape of the electropherogram of digested DNA.

Table 10: Volumes of DNA sample, bead suspension and ethanol to be added during right-side selection (Beads 1, 0.600:1 ratio), left-side selection (Beads 2; 0.645:1 ratio), and washing steps (ethanol). All in μl.

Samples	DNA	Beads 1	Beads 2	Ethanol
1, 3, 5	184.88	110.93	8.32	2×305.00
6	161.48	96.89	7.27	2×305.00

of loci in a size range. The use of magnetic beads for size selection is tricky, because very precise volumes need to be measured, some DNA is lost, and the range is wider and less reproducible than if selected with a Blue Pippin. So, wide size ranges are convenient in this case. Therefore, I should move the range in the direction of lower number of fragments, so that a tighter control in the number of fragments is possible. After looking at the ratios of beads to DNA required for each range, I realize I cannot aim at fragments just between 500 and 800 bp, but rather between 450 and 1000 bp. Because I will use 300 bp, paired end reads, I stick to this range. Finally, I use beads to sample ratios of 0.600 (right-side selection at $\sim$ 900 bp) and 0.645 (left-side selection at $\sim$ 350 bp. I expect less than 33000 loci, according to results from the 2016-04-12/RCATGY folder. Given that the total volumes are either 184.88 or 161.48 μl, this translates into the volumes specified on table 10. I elute with 10 μl.

I have two different batches of magnetic beads. One is a free, 5-ml sample of NucleoMag® NGS Clean-up and Size Select (Machery-Nagel), sent by Purificación Carrasco (Cultek S.L.U.) on July 1st 2016. A 50 ml bottle was offered then for €689.11 (VAT included). The other is a smaller sample of the CleanPCR product by CleanNA, sent on September 21st 2016 by Carmen Llorca (Labclinics S.A.). The 50 ml bottle of CleanPCR is offered for €475.20 + 21.0% VAT = €574.99.

I used CleanPCR with sample 1, and NucleoMag with samples 3, 5, and 6, for comparison. It is possible that the two batches produce different fragment size ranges [Meyer and Kircher, 2010].

11 Amplification PCR. October 25, 2016

After size selection I do not expect to have more than 2 ng of DNA per sample. I will not quantify before amplification, in order to save DNA and because I would not trust so small measures.

I realized I did not have the Phusion® High-Fidelity PCR Master Mix with HF Buffer, but simply the Phusion High-Fidelity PCR Kit. Thus, the PCR reaction was like in table 11. I regreted the mistake, because I suspect pipetting accuracy was far from perfect with new tips that I don't trust. In addition, I note that by mistake I used a primer concentration of 0.2 μM, instead of the recommended 0.5 μM.

When reading the manual of Phusion High-Fidelity PCR Kit, I noticed the recommendation of running a 2-steps PCR with primers > 20 nucleotides and with $T_m \geq 69$°C. While the melting temperatures of the portions of the primers that anneal during the first and second cycle don't seem to be larger than 69°C (table 12), actually when using the ThermoFisher T_m calculator, they are. Thus, a 2-step PCR programme is recommended, with a unique temperature of 72°C for both annealing and extension. I finally gave only 20 s for these steps, which in retrospect seem too short (table 13).

After setting up the PCR reaction, I received Sibelle's Vilaça suggestion. She cites Meyer and Kircher [2010], a protocol using the same PCR kit that I use, and in a similar application. They recommend the exact same reaction that I set up (table 11), including the 0.2 μM primers. However, the programme they suggest is different from mine (see table 13). It is unclear if the difference is justified by the use of slightly shorter primers, or if they just overrode Meyer and Kircher [2010] recommendation for some reason.

Because we have 4 differently indexed first primers, I prepared four primer mixtures (table 14). During preparation, I commited the mistake of adding 1 μl of 100 μM primer A1.1 directly to the template DNA of sample 1, instead of adding it to the corresponding primer mix. Because that would be a too high primer concentration, I decided to clean up sample 1's DNA with magnetic beads again, using a 0.7 beads to DNA ratio. Again, I used the CleanPCR brand, although the comparison with the other brand may be affected by this additional step.

12 Clean-up and Bioanalyser quantification. October 26, 2016.

I cleaned up the PCR reactions using a 0.8:1 beads to DNA ratio. Again, sample 1 was treated with the CleanPCR suspension, and the others with Machery Nagel's NucleoMag. I eluted with 10 μl of water. Then, I took the samples to the Sequencing Services to be quantified with the High Sensitivity Bioanalyser chip.

On October 27$^{\text{th}}$, A. Martínez sent me the electropherograms, shown on

Table 11: PCR reactions in 50 μl. 10 μl of 10 μM Primer Mix may be prepared with 1 μl 100 μM Primer 1 + 1 μl 100 μM Primer 2 + 8 μl nuclease-free water.

Component	Volume (μl)	Final conc.
H_2O	27.5	
5× HF buffer	10.0	1×
10 mM dNTPs	1.0	200 μM each
10 μM Primer Mix	1.0	0.2 μM
Template DNA	10.0	
Phusion DNA Pol.	0.5	0.02 U/μl
Total	50.0	

Table 12: Amplification primers. These properties were calculated with an online oligonucleotide properties calculator Kibbe [2007]. Different properties are obtained with the ThermoFisher Tm calculator.

	Primer 1a	Primer 1b	Primer 1c	Primer 1d	Primer 2
Length (bp)	63	63	63	63	55
Molecular weight (g/mol)	19511.7	19545.7	19536.7	19536.7	16719.9
Basic T_M First cycle (°C)	65.7	65.7	65.7	65.7	64.4
52 mM Na^+ T_M 1^{st} cycle (°C)	75.0	75.0	75.0	75.0	73.3
Basic T_M Later cycles (°C)	75.7	76.4	76.4	76.4	74.3
52 mM Na^+ T_M Later (°C)	87.2	88.0	88.0	88.0	85.7

Table 13: PCR programme. In parenthesis, values recommended by Meyer and Kircher [2010], if different from the ones used.

Step	Temp.(°C)	Time (s)	
Initial denaturation	98	30	
Denaturation	98	10	12×
Annealing	72 (60)	10 (20)	
Extension	72	10 (20)	
Final extension	72	600	
Hold	4	∞	

Table 14: Indexing of samples.

Sample	Primer 1	Primer 2	Index
1	A1.1	A2	CTTGAGTC
3	A1.2	A2	GAACGCTG
5	A1.3	A2	GCCAGGTT
6	A1.4	A2	GCGTTAGC

figure 3. The figure confirms that the ligation of adapters, the size selection, and the PCR worked more or less as expected. The only disappointment is that the Machery Nagel magnetic beads have quite different affinity for DNA than expected, which resulted in a shift of the selected size range towards larger fragments in samples 3, 5, and 6. I therefore, need to repeat the size selection on those samples, using the CleanPCR product.

13 Second size selection of samples 3, 5, and 6. November 3, 2016

I need to remove the large DNA fragments (right-side size selection). I will use the same suspesion to DNA ratio used before for right-side selection, namely 0.600:1. I have 9 μl left of each sample, which means I should use 5.4 μl of CleanPCR suspension to trap the large fragments and leave the small ones in the supernatant. Unfortunately, the supernatant will contain polyethylene glycol (PEG) and salt. Thus, I need to also apply a second load of suspension be able to elute the DNA in water. The second ratio does not need to be very tight, just to remove the very small primers that may be left over. Thus, in order to optimize the yield of small fragments, I will set it to 0.9:1. This translates in 2.7 additional μl of the CleanPCR suspension.

I performed the second size selection on samples 3, 5, and 6 without problems. I also run the amplification PCR just as the first one (table 11), but in 25.0 instead of 50.0 μl. I did change pipette tips. The FisherBrand tips that I was testing did not fit well the pipette, and were not accurate. The VWR tips work better with my pipettes (Nichipet EX II).

On November 8$^{\text{th}}$, A.M. sent the results of the Bioanalyzer posterior to the size selection and amplification. It immediately became clear that I had forgotten to clean up the PCR reactions, since the primers were still there. I show the results on figure 4. In addition, the bimodal nature of the size distribution suggests that the beads became saturated during right-side

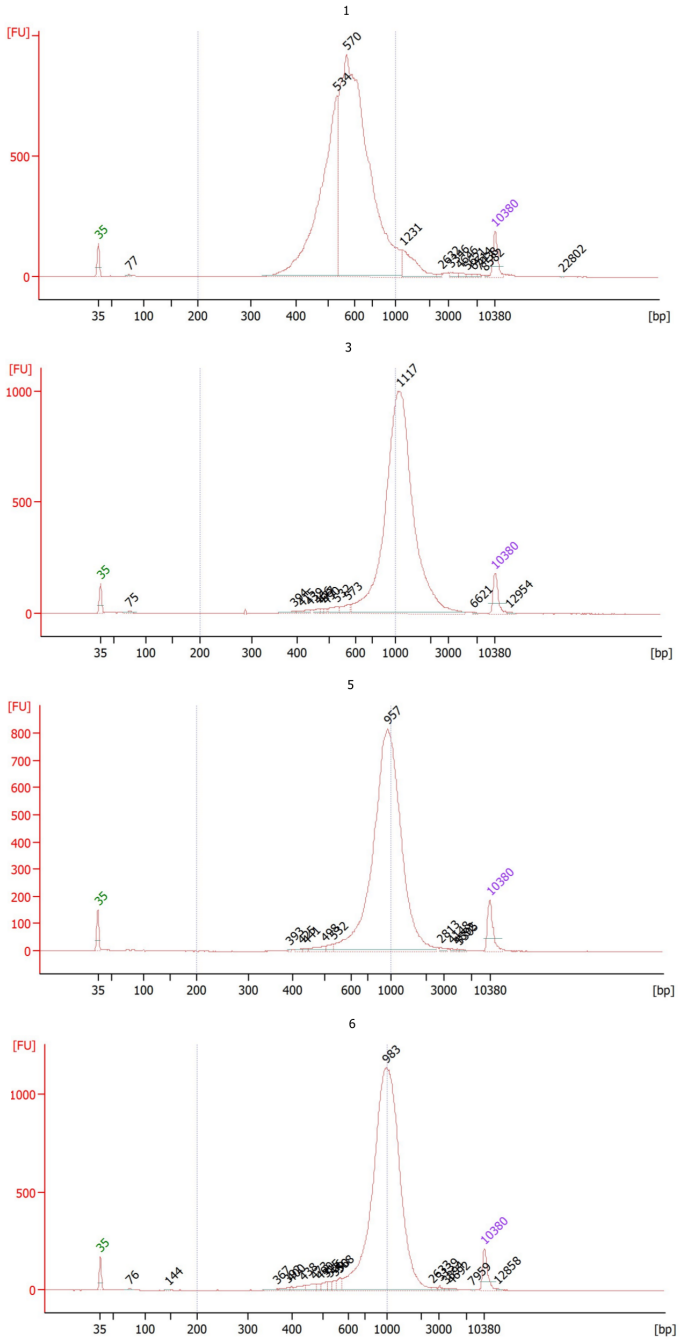


Figure 3: Electropherograms from the high-sensitivity DNA assay of the four samples after size selection and PCR amplification.

selection. Because the DNA volume was so low (9 μl), the right proportion of beads did not contain enough bead surface to adsorbe all the large DNA fragments. Also, because the DNA concentration of large fragments was quite high after the PCR.

14 Third attempt of size selection with samples 3, 5, and 6. November 9$^{\text{th}}$

I had original DNA volume of 24.0 μl: the PCR product minus 1 μl used in the HS DNA chip. I added 16.0 μl of H_2O (total DNA volume of 40.0 μl), in order to make sure that there would be enough magnetic beads in suspension to adsorbe all the large DNA fragments. As before, I used a 0.6:1 bead suspension to DNA ratio for right-side selection, and a 0.645:1 ratio for left-side selection. That amounts to 24.0 μl of beads first, plus 1.8 μl afterwards. I eluted with 17.75 μl of water, instead of 10.0, in order to save the pipetting of 7.75 μl of water for the PCR.

On November 14, A.M. sent the results of the Bioanalyser (figure 5). This time, the libraries have enough DNA and in an acceptable size range. Obviously, after so many rounds of PCR, I expect samples 3, 5, and 6 to have a terrible bias. Still, I will go ahead with the sequencing, in order to estimate the number of fragments sequenced at least from sample 1, and to compare them to the expectation.

15 Preparation of sequencing primers. November 15

A.M. requests 600 μl of 0.5 μM sequencing primers (300 pmol). In addition, she asked about the indexing primer. I had assumed it would be just the standard one. In fact, the TruSeq 'Multiplexing Index Read Sequencin Primer' is two nucleotides longer than what I need. Fortunately, the two extra nucleotides are in 5', and will not prevent correct priming (see representation below).

I prepare 1 ml of each sequencing primer at 5 μM, which is equivalent to 5000 pmol. The primers are at 100 μM. Thus, the dilution factor is 20, which means that I need $1000/20 = 50$ μl of primer and 950 μl of water.

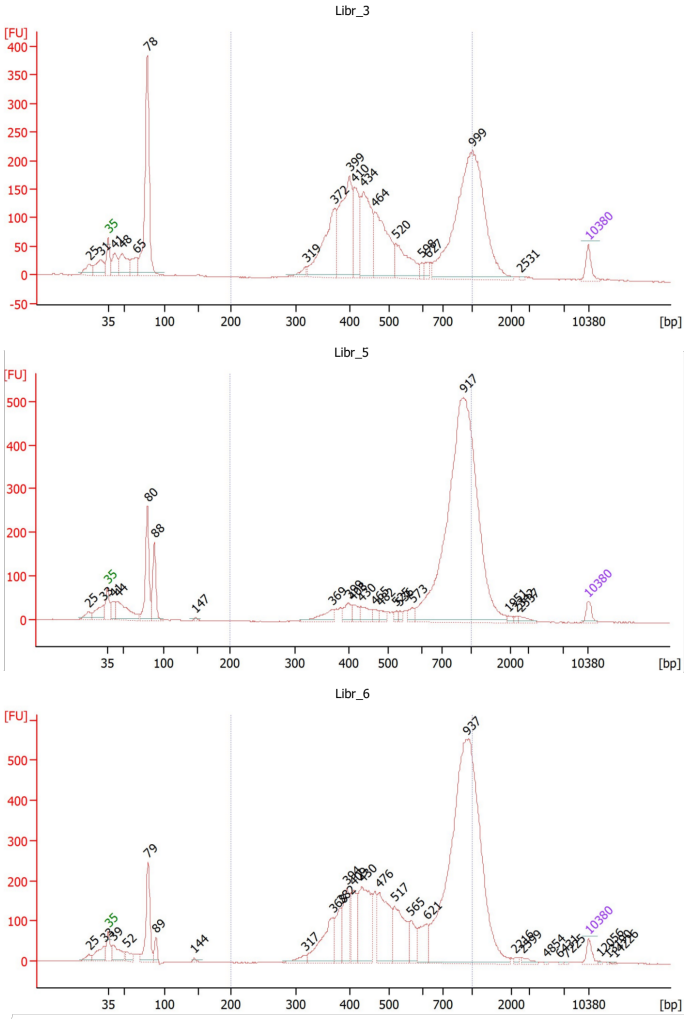


Figure 4: Electropherograms from the HS DNA assays after the second size selection performed on samples 3, 5, and 6. We did not get rid of the the large fragments, and the primers also show up, because I failed to clean up the PCR reaction.

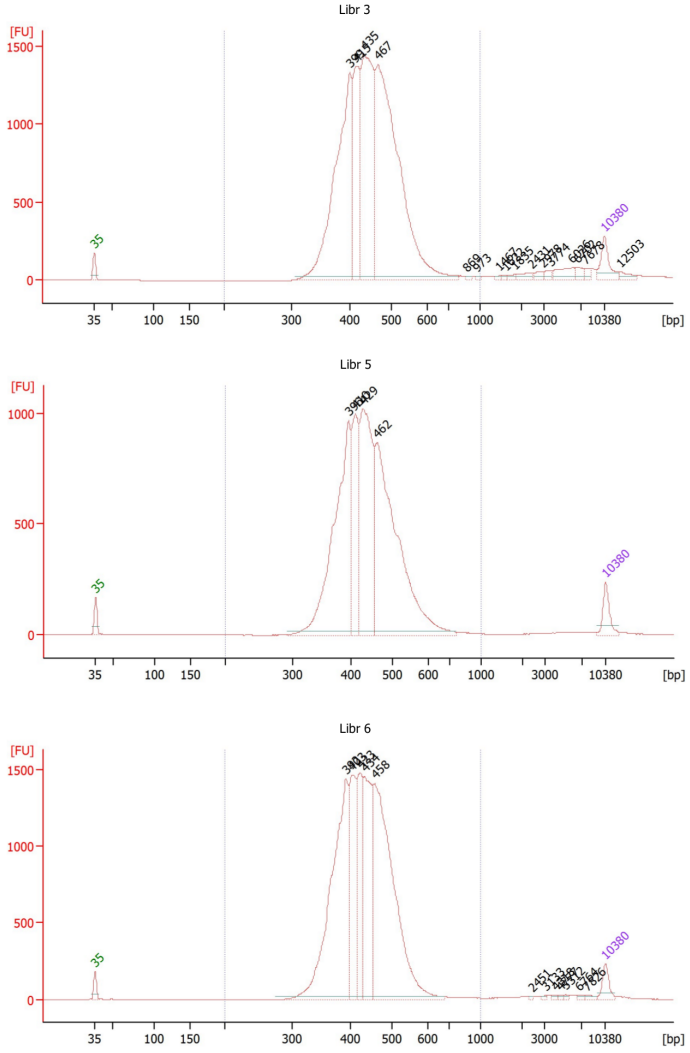
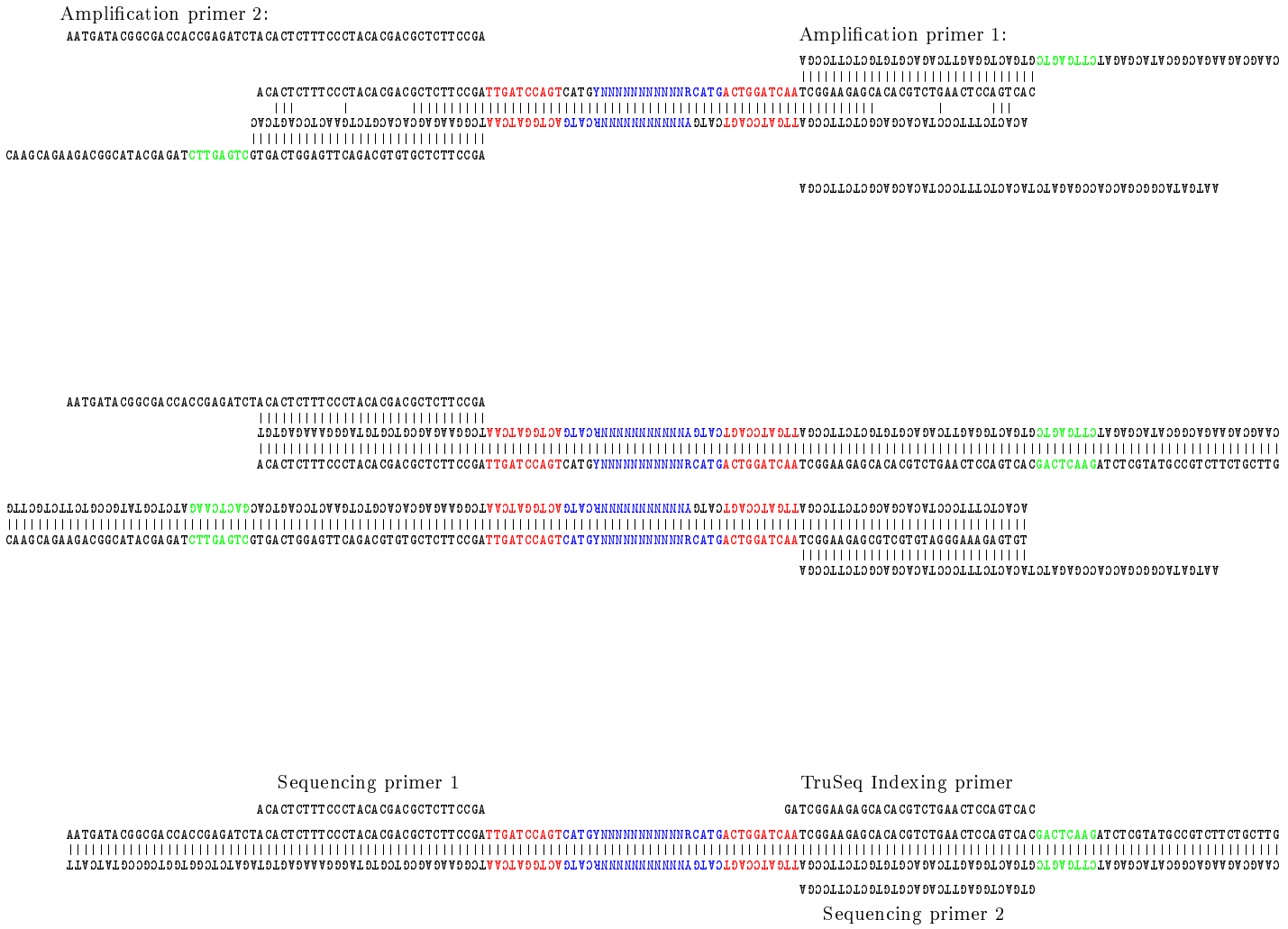


Figure 5: Electropherograms from the HS DNA assay of samples 3, 5, and 6, after the third round of size selection. This time, it seems to have worked, despite of a shift to the left, with respect to sample 1.


Amplification primer 2:
Amplification primer 1:
Sequencing primer 1
TruSeq Indexing primer
Sequencing primer 2

16 Summary of sequencing and pre-processing of reads

Sequenced reads came already identified from the sequencing center, by the indices. In order to check the reliability of indices and barcode, I also demultiplexed them by the barcode, using both Sabre and with Pyrad Eaton [2014]. Pyrad (but not ipyrad) reports what words have been found as barcodes, and in what frequency. That helps identify common versions of the true barcode that the algorithms fail to recognize, e.g. because they involve the deletion of the first base. I keep in separate fastq files reads with a different combination of index and codeword. In most reads (88.13 %), index and codeword coincide. The contradictions between index and codeword need be analysed until their true identity can be decided by the genetic variants.

The merging of forward and reverse reads was very successful, resulting in no unmerged pairs other than those discarded by quality filters. Merged reads were checked for the presence of adapter sequences, and less than 0.1 % had any trace of adapter sequence. The average lengths of merged reads were 394 bp for sample 1, 289 bp for sample 3, 283 bp for sample 5, and 280 bp for sample 6. This roughly corresponds to what was expected from the electropherograms of the libraries (see figure 3 for sample 1, and figure 4 for samples 3, 5, and 6; take into account that the original library included between 126 and 129 additional base pairs, not present in demultiplexed, merged reads).

Overall mapping rates to the *D. melanogaster* reference genome are above 95 %, while unambiguously mapped reads are not more than 85 %. Curiously, around 0.3 % of samples map to the Y chromosome (all samples were females). The Y chromosome is just 2.9 % of the length of the genome, and because of being in one coppy per cell it should be represented by 1.6 % of the reads. I favor the hypothesis that the reads mapping to the Y chromosome are just missmapped. Only less than 2 % of those Y-mapping reads hava a mapping quality of at least 20.

I used bedtools to compare the loci covered by the samples. There are 14149 sites covered at least 6 times in sample 1, 12157 in sample 3, 14529 in sample 5, and 12496 in sample 6. Overall, it is quite good. 9157 of all those are in common. But, for the purpose of estimating runs of homozygosity, it is not necessary for all samples to have the same markers covered. Note that those counts are loci shorter than 1000 nucleotides. For some reason the mappings include reads stretched over much larger regions, which must be spurious.

The bad news is that overall there are almost 50000 sites apparently

covered by at least one read. That is much higher than I intended, and spreads the coverage too much to sequence 24 samples per run. I expect to improve the coverage by using higher cluster densities in the sequencing run and by using the exact same conditions of size selection in all samples. Also, the fact that samples 3, 5, and 6 underwent 36 PCR cycles probably contributed to a biased distribution of coverage, and will be avoided in future runs.

References

J. Bennett. Random Mating and Sex Linkage. *J. Theor. Biol.*, 4(1):28–&, 1963.

A. Burt, G. Bell, and P. Harvey. Sex-Differences in Recombination. *J. Evol. Biol.*, 4(2):259–277, 1991.

D. Charlesworth and J. H. Willis. FUNDAMENTAL CONCEPTS IN GENETICS The genetics of inbreeding depression. *Nat. Rev. Genet.*, 10(11): 783–796, Nov. 2009.

C. C. Cockerham and B. S. Weir. Sib Mating with Two Linked Loci. *Genetics*, 60(3):629–640, Nov. 1968.

J. F. Crow. Mid-Century Controversies in Population Genetics. *Annu. Rev. Genet.*, 42(1):1–16, Nov. 2008.

D. A. R. Eaton. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, page btu121, Mar. 2014.

I. R. Franklin. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theoretical Population Biology*, 11(1):60–80, Feb. 1977.

J. B. S. Haldane. Sex ratio and unisexual sterility in hybrid animals. *Journ. of Gen.*, 12(2):101–109, 1922.

J. S. Huxley. Sexual difference of linkage inGammarus chevreuxi. *Journ. of Gen.*, 20(2):145–156, 1928.

W. A. Kibbe. OligoCalc: an online oligonucleotide properties calculator. *Nucl. Acids Res.*, 35(suppl 2):W43–W46, July 2007.

T. Lenormand and J. Dutheil. Recombination Difference between Sexes: A Role for Haploid Selection. *PLOS Biol*, 3(3):e63, Feb. 2005.

T. Lenormand, J. Engelstädter, S. E. Johnston, E. Wijnker, and C. R. Haag. Evolutionary mysteries in meiosis. *bioRxiv*, page 050831, Apr. 2016.

M. Lynch and B. Walsh. *Genetics and analysis of quantitative traits*. Sinauer Associates, Inc., Sunderland, MA, 1998. ISBN 0-87893-481-2.

M. Meyer and M. Kircher. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*, 2010 (6):pdb.prot5448, June 2010.

S. V. Nuzhdin, E. G. Pasyukova, C. L. Dilda, Z.-B. Zeng, and T. F. C. Mackay. Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster. *Proc Natl Acad Sci U S A*, 94(18):9734–9739, Sept. 1997.

S. V. Nuzhdin, A. A. Khazaeli, and J. W. Curtsinger. Survival Analysis of Life Span Quantitative Trait Loci in Drosophila melanogaster. *Genetics*, 170(2):719–731, June 2005.

R. Owen. Selection at two Sex-Linked Loci. *Heredity*, 60:415–425, June 1988.

E. G. Pasyukova, C. Vieira, and T. F. C. Mackay. Deficiency Mapping of Quantitative Trait Loci Affecting Longevity in Drosophila melanogaster. *Genetics*, 156(3):1129–1146, Nov. 2000.

M. White. *Animal cytology and evolution*. Cambridge University Press, London, 3rd edition edition, 1973. ISBN 0 521 07071 6.