Exam Genomes and Phenotypes 2020
Evolution of fly glue

This exam was done at home by each student in a limited amount of time due to COVID-19 pandemic.
This exercise is based on unpublished results obtained by my lab, so searches on internet will not be useful here! Please try to be as clear as possible and explain your reasoning in your answers. Answers can be in French or English.

Juste before entering metamorphosis, Drosophila larvae produce a glue to stick themselves on natural substrates (wood, leaves, stones, bear bottles, etc.). Then the animal (the pupa) remains attached to the substrate for several days until the end of metamorphosis where the adult emerges from the pupal case. The glue is made of several proteins named Sgs1 to Sgs8 (for Salivary gland secretory protein). How these proteins confer adhesiveness is unknown.

1) We designed an assay to quantify adhesiveness of single pupae. Fig. 1 shows adhesion strength in various strains of *Drosophila melanogaster* (named A5, B4, A6, etc.) originating from different parts of the world.

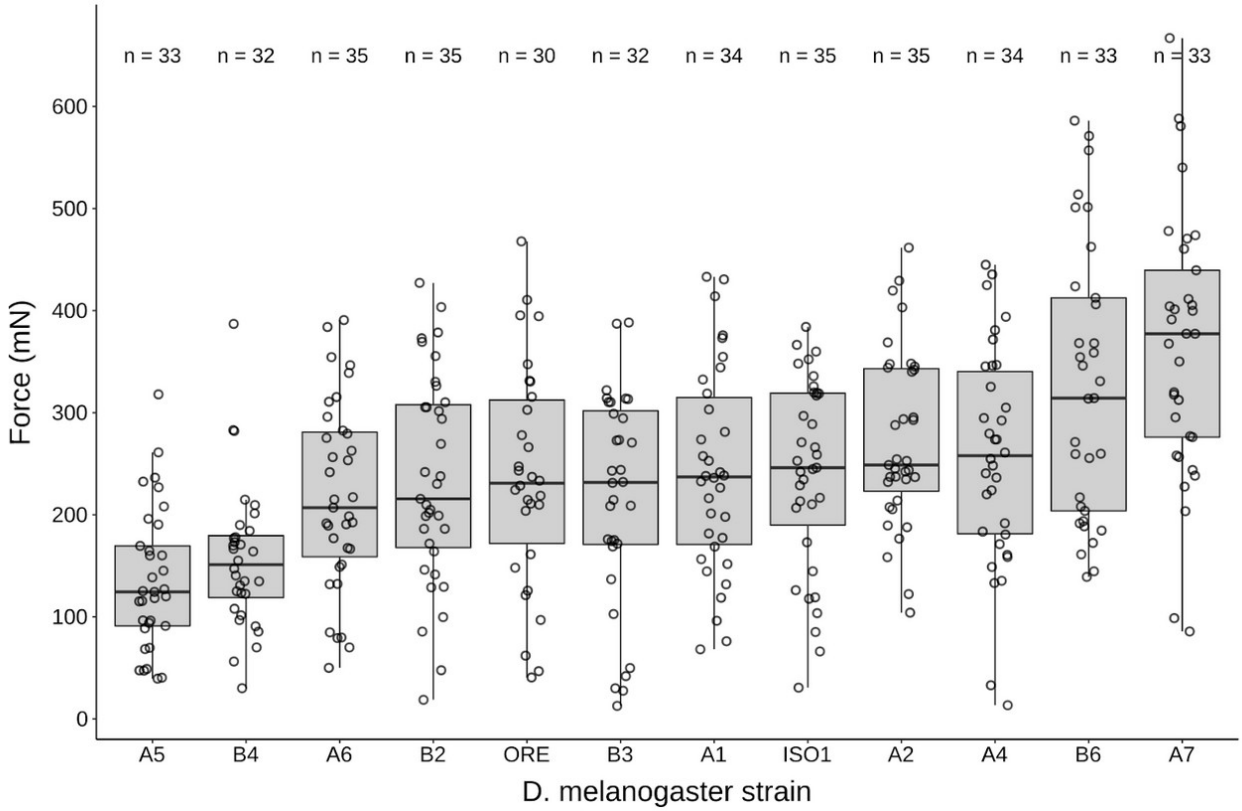

Figure 1. Adhesion strength of DSPR lines. Force indicates the force required to detach a pupa naturally attached to a glass slide. Each dot corresponds to a single pupa and n indicates the total number of pupae tested for each strain. Ends of the boxes define the first and third quartiles. The black horizontal line represents the median. The vertical line on the top of the box extends to the largest value no further than 1.5 * IQR from the upper hinge of the box. The vertical line on the bottom of the box extends to the smallest value at most 1.5 * IQR of the hinge. (IQR: inter-quartile range is the distance between the first and the third quartiles). Data beyond the end of these lines are "outlying" points.

All the experiments were performed at 25°C. What are the mechanisms that can explain the high variability in adhesion values for a given strain?

2) Strain A7 appears to stick more than the other strains, and strain A5 appears to stick less. So we decided to investigate whether mutation(s) in some of the *Sgs* genes could explain their weaker and stronger adhesion, respectively. Complete sequences for all the Sgs genes were found in all strains. Two candidate mutations were found: one in *Sgs5* gene (chromosome 3) for strain A7 and one in *Sgs8* gene (chromosome 3) for strain A5. Below are alignments of parts of the relevant *Sgs* sequences with strain iso-1, which is considered here as the "wild-type" normal strain.

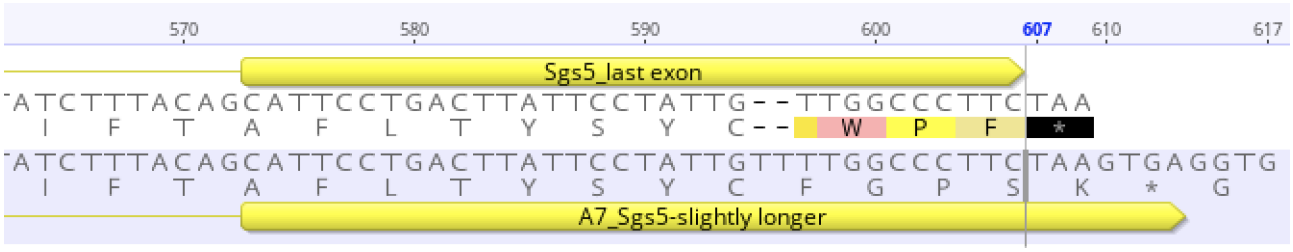

Fig. 2. Alignment of partial sequences of *Sgs5* for strain iso-1 (top) and strain A7 (bottom). The normal amino acid sequence of SGS5 ends with CWPF.

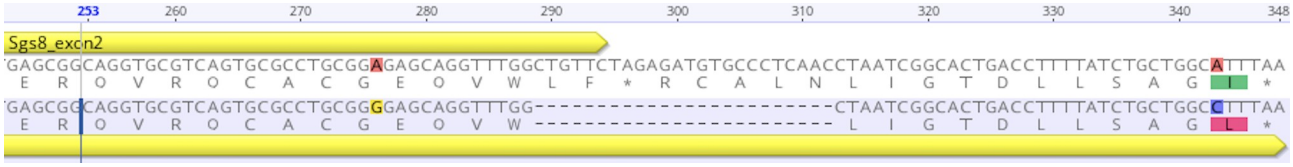

Fig. 3. Alignment of partial sequences of *Sgs8* for strain iso-1 (top) and strain A5 (bottom). The normal amino acid sequence of SGS8 ends with VWLF.

Describe the two candidate mutations and their effects on glue proteins.

3) We designed PCR primers so that they amplify a 162-bp fragment from iso-1 strain and a 138-bp fragment from A5 strain.

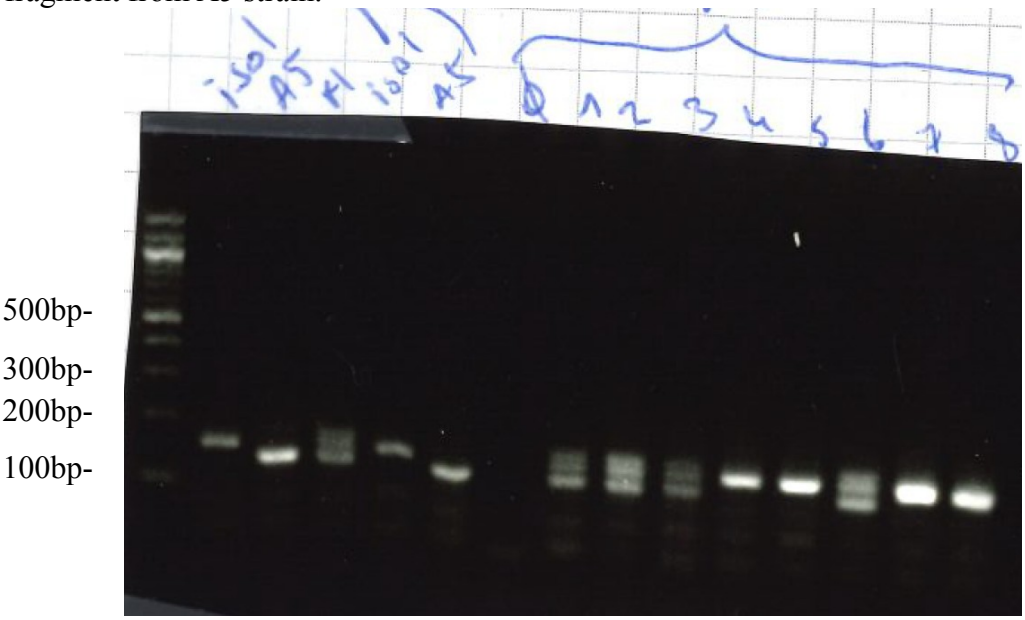

Fig. 4. Electrophoresis gel of the PCR fragments. Lines are in the following order: molecular weight size marker, mix of five iso-1 individuals, mix of five A5 individuals, mix of five F1 hybrid individuals from A5 x iso-1 cross, mix of five other iso-1 individuals, mix of five other A5 individuals, no DNA, individual F2 progeny pupae named from #1 to #8.

Explain the results of this electrophoresis gel. Why was a PCR reaction performed in absence of DNA? (Note: You don't need to explain the presence of a third band in the F1.)

4) We decided to test whether the candidate mutation in *Sgs8* is responsible for the reduction in adhesion. One solution is to make a CRISPR mutation in the iso-1 stock, but this experiment takes several months and costs about 1500 euros. So, in a first step, we decided to cross strain A5 with strain iso-1, collect the F1 hybrid progeny flies and cross them together to obtain F2 individuals. We measured adhesion of several F1 pupae and 150 F2 pupae, extracted their DNA and genotyped each individual for the PCR marker in *Sgs8* described in the question 3 above.

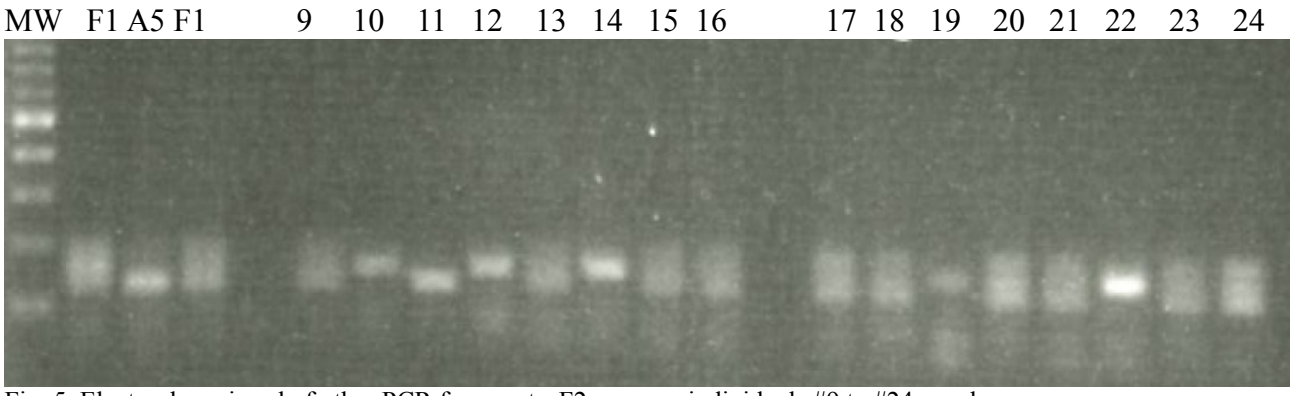

Fig. 5. Electrophoresis gel of other PCR fragments. F2 progeny individuals #9 to #24 are shown.

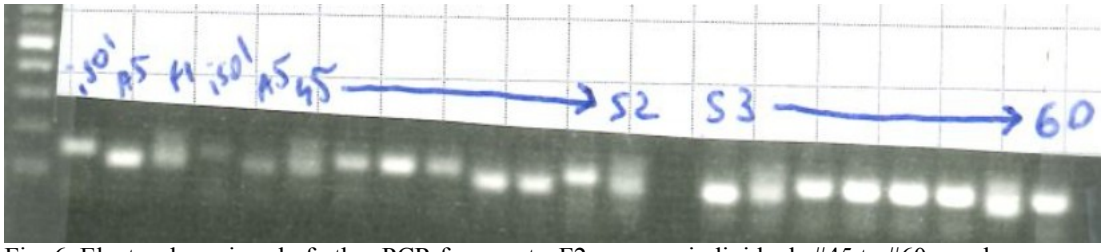

Fig. 6. Electrophoresis gel of other PCR fragments. F2 progeny individuals #45 to #60 are shown

Electrophoresis gels are not always of great quality… (Note: The orientation of the gels is not important here.)
Fill up the attached Table 1 with genotype values. Write "x" if you are unsure of the genotype.

../...

5) Figure 7 presents the adhesion values for four different genotypes.

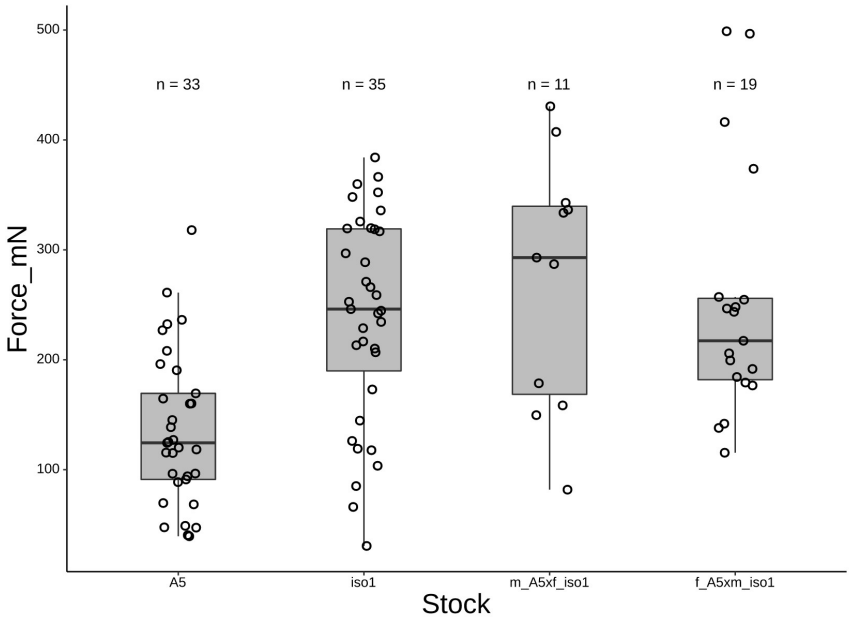

Fig. 7. Adhesion strength of the A5 strain, the iso-1 strain, F1 pupae originating from the following cross: A5 males x iso-1 females, F1 pupae originating from the following cross: A5 females x iso-1 females. Same legend as Fig. 1.

Interpret Figure 7.

6) Use Table 1 to test whether the candidate mutation in *Sgs8* might be responsible for the reduction in adhesion. Please feel free to draw diagrams if necessary.

Note: In Excel/LibreOffice, you can sort values from low to high: (1) select the rows, (2) go to "Data", (3) click on "Sort..", (4) select the column(s) for which you want ascending/descending sorting.

Please don't forget to send both your answers and Table 1!

Corrections

1) Several phenomena can explain the high variability in phenotypic values:
- measurement error (due to the machine or the set up)
- genetic variability (should be low but can still exist)
- environmental variability (size of the individuals, shape of the pupae, position of the pupa relative to the slide, amount of glue produced, humidity)
- noise (not explained by the previous factors, for example due to differences in the number of salivary gland cells due to developmental noise)

2) - a 2bp-insertion, which causes a frameshift, with the last 3 amino acids replaced by five new amino acids.
- a 24-bp deletion, multiple of 3, so no frameshift, the last two amino acids are replaced by a stretch of 11 new amino acids.
Both are coding changes: they affect the protein sequence.

3) iso1: homozygote for the WT allele: 1 band at 162 bp
A5: homozygote for the A5 allele: 1 band at 138 bp
F1: heterozygote, both bands are present
Negative control (to make sure that there was no contaminant DNA in the experimental process).

4) The csv file should be:
name_F2_individual,genotype ,adhesion_force
iso1,2,-
A5,0,-
F1,1,-
1,1,421.33
3,1,358.49
4,2,429.58
5,2,380.23
6,1,269.22
7,2,204.04
8,2,287.25
9,1,270.27
10,2,147.49
11,0,292.11
12,2,419.14
13,1,183.53
14,2,144.48
15,1,201.85
16,1,193.12
17,1,105.38
18,1,285.39
19,2,411.32
20,1,247
21,1,477.65
22,2,211.57
23,1,132.32
24,1,404.6
25,1,223.16

26,1,216.44
27,1,383.38
28,1,418.33
29,2,25.749
30,2,147.53
31,2,507.5
32,1,368.12
33,1,275.33
34,2,164.37
35,1,41.533
36,1,160.88
37,2,256.49
38,0,147.25
39,1,119.78
40,2,268.08
41,2,247.19
42,0,155.16
43,1,450.61
44,2,138.66
45,1,187.16
46,2,306.99
47,2,9.251
48,2,515.75
49,0,78.917
50,0,23.27
51,2,296.4
52,1,288.63
53,0,204.42
54,1,201.27
55,2,139.38
56,2,234.94
57,2,275.18
58,2,396.35
59,1,383.42
60,2,172.66
61,1,198.79
62,2,353.43
63,2,118.02
64,2,180.29
65,1,202.7
66,2,206.66
67,1,118.73
68,1,151.35
69,2,226.93
70,1,28.801
71,0,247.29
72,1,273.99
73,2,403.31
74,2,356.29
75,1,91.028

76,2,420.05
77,1,114.2
78,0,184.77
79,2,221.63
80,2,294.3
81,1,7.296
82,1,30.327
83,1,293.3
84,1,258.59
85,1,106.81
86,2,327.73
87,0,11.969
88,0,11.683
89,2,167.85
90,2,247.67
91,1,321.15
92,1,258.02
93,1,275.18
94,1,287.44
95,2,104.76
96,1,224.92
97,2,69.237
98,0,181.06
99,1,156.4
100,2,113.92
101,2,96.893
102,1,319.39
103,1,162.32
104,2,305.37
105,1,223.21
106,1,456.71
107,2,287.87
108,0,129.89
109,1,424.86
110,2,401.31
111,2,269.89
112,1,376.22
113,2,293.35
114,2,234.03
115,2,385.14
116,2,429.82
117,2,422.43
118,2,306.32
120,1,138.62
121,2,118.06
122,2,333.83
123,1,71.239
124,1,51.212
125,0,117.54
126,2,219.58

127,1,77.629
128,1,204.9
129,1,179.2
130,2,281.91
131,1,137.57
132,1,263.26
133,1,378.75
134,1,256.78
135,2,181.91
136,1,376.8
137,1,208.23
138,1,288.68
139,1,253.15
140,2,398.87
141,1,268.03
142,1,278.76
143,1,318.81
144,1,313.76
145,0,168.71
146,1,226.64
147,2,405.03
148,1,139.09
149,1,387.33
150,1,280.52

5) If the phenotypic variation is due to a single locus:
F1 in both directions display similar forces so the locus is autosomal. This is in agreement with *Sgs8* being located on chromosome 3.
WT allele is dominant because the F1 phenotype is similar to iso-1 phenotype.
We need more data points for the phenotypes of the F1 individuals to confirm the trend.

6) A scheme with the three genotypes (0, 1, 2) on the x-axis and the phenotypic values on the y-axis, with points for all data points, average, statistical test and interpretation of the results. The genetic locus is associated with adhesion variation.