Journal of Applied Bioinformatics & Computational Biology

Research Article A SCITECHNOL JOURNAL

Possible Critical Mass (kD) Value Limiting Expression of *Drosophila melanogaster* Pupae under Microgravity Conditions

Andy Mercado[1], Maria Teresa Del Solar-Vela[1], Diego Orihuela[1], Gabriela Sanabria[3], Jhony de la Cruz[2], David Lavan[2] and Juan Jose Augusto Moyano Munoz[4*]

[1]*Department of Biology, Universidad Peruana Cayetano Heredia, Lima, Peru*

[2]*Department of Biomedical Sciences, Ricardo Palma University, Lima-Peru*

[3]*Department of Biology, Universidad Nacional Mayor de San Marcos, Lima, Peru*

[4]*Department of Cardio-pneumology, San Martin de Porres University, Santa Anita 15011, Peru*

[*]**Corresponding author:** Juan Jose Augusto Moyano Munoz, Department of Cardio-pneumology, San Martin de Porres University, Santa Anita 15011, Peru; E-mail: jmoyanom@usmp.pe

Received date: 02 October, 2023, Manuscript No. JABCB-23-115466;

Editor assigned date: 04 Octoberr, 2023, PreQC No. JABCB-23-115466 (PQ);

Reviewed date: 18 October, 2023, QC No. JABCB-23-115466;

Revised date: 17 January, 2024, Manuscript No. JABCB-23-115466 (R);

Published date: 24 January, 2024, DOI: 10.4172/2329-9533.1000293

Abstract

This study aimed to identify microgravity-responsive regulatory genes in *Drosophila melanogaster* using our microarray dataset. Contrary to expectations, we found that known regulators did not respond to microgravity, prompting a reevaluation of how the 807 identified genes might be regulated. We hypothesized that the differential travel time of polymerase affects transcript production, suggesting that smaller RNA strands correlate with higher expression levels. This led us to calculate a critical mass (kD) of 36.6875 ± 1.36 kD, centered at 1000 bp of mRNA, marking a transition from positive to negative gene expression under microgravity conditions.

Of the 807 IDs initially identified, only 741 corresponded to unique genes, while some IDs lacked gene associations or corresponded to multiple genes. Notably, six IDs identified *via* correlation analysis included five genes (*ImpL-3, DnaJ1, CG4726, CG14598, CG5326*) that showed over 78% detection across 10,750 iterations. Validation of our correlation approach involved repeating a 2003 experiment, confirming a consistent negative expression profile in 2010.

We also selected four positively regulating genes (C*G5703, Surf1, CG8885, CG1970*) and one negative regulator (Mef2) to explore their roles. Further investigation of 134 genes that anchor transcription factor binding sites revealed no response to microgravity. Analyzing mRNA lengths, we focused on the isoform region around 1000 bp, calculating mean values and standard deviations, reinforcing that the critical mass affecting expression in *Drosophila* pupae under microgravity is approximately 36.6875 kD.

Keywords: Regulation factors; Responsive genes; Development differentiation; Microgravity

Introduction

The influence of gravitational forces on the development of living organisms has had an impact on evolution by determining the shape and size of organisms, and their presence was indispensable for the skeletal development of organisms. One of the first organisms used to determine how the absence of gravity influences early embryo development was *Drosophila melanogaster* or fruit fly. The results of the first experiments in space showed the normal development of *Drosophila* in space conditions. However, results obtained during the last successful 7-day flight of the challenger shuttle showed that oogenesis and embryonic development of *Drosophila* change in the absence of gravity.

In the 1980's, when the advances in technology with respect to massive gene expression studies and short processing (microarrays) were not yet known, Adams and Ho suggested that environmental conditions interact with the cytoplasm to affect gene expression. An altered genome that responds to a particular environmental condition could be heritable and produce new phenotypes. According to this view, for gravity to have become a force of evolution there had to be a major environmental condition to which a regulatory response of genes became necessary.

The concept and methodology of microarrays were first introduced and illustrated in antibody microarrays (also called antibody arrays). Microarrays have dozens to millions of probes attached to an inert surface, allowing high-throughput analysis of many biological processes simultaneously on the same sample. At that time the fundamental strategy was post-genomics or functional genomics in broadening the scale of biological research from the study of individual genes or proteins to the study of all genes or proteins simultaneously through a systematic approach. As newly developed methods for obtaining genome-wide mRNA expression data, oligonucleotide and DNA microarrays are especially powerful in the context of genome wide sequence knowledge and can provide a global view of changes in gene expression patterns in response to physiological alterations or manipulation of transcriptional regulators. Microarrays with nucleic acid probes were an alternative for the analysis of gene expression in a massive manner from samples of *Drosophila melanogaster* embryos, larvae, and pupae exposed to microgravity.

Thanks to new techniques of massive studies, it has been possible to demonstrate that the variation of the gravitational field in any direction causes numerous changes in the development of organisms, as also the mechanisms of detection of gravitational forces, their influence on organisms and their response to gravitational changes at the cellular and molecular level in plants and animals have been described.

The experiments that provide us with the data to be analyzed in this study were carried out at the International Space Agency (ISS) in 2003 on the cervantes mission. The development of the fruit fly *Drosophila melanogaster* during this mission resulted in generally normal specimens, but at a slower rate. When an embryo or larva develops in

space or in simulated microgravity here on Earth, there is a delay in the development of the embryo, the cytoskeleton is altered, the morphology of the mitochondrial ridge is altered, producing a possible change in its function, the process of oogenesis is altered, possible implications on the immune system, and accelerated aging. The cosmic radiation existing in low orbit, where the ISS was moving, did not generate mutations in embryos, larvae or pupae of *Drosophila melanogaster*. The experiment consisted of placing *Drosophila* samples in a centrifuge at 1 g within the environment of the International space station, thus demonstrating that cosmic radiation did not generate any mutations in the *Drosophila* specimens; these results were in agreement with those published by Vernos I., et al. However, other experiments carried out by the same group of authors using equipment that allows improved oxygenation of the samples, demonstrated the normal quantitative and qualitative development of *Drosophila* in microgravity conditions. In addition, the weightlessness of space orbit increased transcription of metabolic and cuticle component genes and, as expected, decreased transcription of genes involved in morphogenesis, cell differentiation, cytoskeleton organization and plasma membrane-associated genes.

In this particular case we will start from the data already published and presented in Dr. Lavan's thesis. The first hypothesis proposed was: If the transcriptome of *Drosophila melanogaster* changes under microgravity conditions, this implies that the changes in expression levels must be affected by their respective regulators, which must also change under microgravity conditions. If the negative hypothesis is proved, who regulates the transcriptome? Assuming that regulators do not play an important role in microgravity, this result would lead us to think that not only the initial experimental conditions were the same, but a new condition would be added, which would be the expression machinery. The only thing that could be different was the travel time of the polymerase for each transcript. The number of strands produced would depend on the length of the mRNA of each gene, the smaller the number of strands the higher the expression, the larger the number of strands the lower the expression. This new hypothesis led us to think about the existence of a possible critical value of mass (kD) as a boundary value sensitive to microgravity. To do this we will check if the total ID (Af-fimetrix chip) is equal to the number of genes that respond to microgravity, secondly we will search for a small group of genes by clustering according to the degree of correlation between them and subject this set of genes to microgravity but without cold pretreatment in order to analyze the response profile between the two experiments, thirdly, we will perform a search of all known genes that regulate the expression of proteins that are anchored to transcriptional factor binding sites in order to verify if they respond to microgravity, fourthly, we will order the levels of expression or inhibition according to the length of their mRNA of the genes that respond to microgravity (only those of the ISS) and finally we will calculate the possible value of critical mass (kD) that limits gene expression in *Drosophila melanogaster* pupae under microgravity conditions.

Materials and Methods

Microarray data analysis

RNA samples that were extracted from pupae under real as well as simulated microgravity conditions and controls (normal gravity) were hybridized on *Drosophila melanogaster* Affymetrix chips (version 1.0), which contains probes for 14010 genes (Affymetrix *Drosophila*-1), almost the entire *Drosophila* genome, 55% of which are labeled with some molecular markers while 45% are awaiting identification. Once the data from the Affymetrix chips were obtained in cell format, this information was used to integrate two different techniques for gene expression analysis in order to standardize them for efficient use. The most common statistical algorithms used for gene expression analysis are MAS 5.0 and affylmGUI, the latter working on the R platform, developed by the company AFFYMETRIX [1]. Although both methods are different in the processing of the data using different normalization methods, the conclusions regarding the affinity, or trend, of the results are the same. In order to obtain this affinity, traditional methods have to correct the P-value and cut-off values to obtain a characteristic value with which a higher correlation value is obtained. If both methods work with statistically differentiable genes with presence character, we think that we should not condition our results to a certain value of P-value and cut-off; therefore, we propose a new method whose results are not directly conditioned by the variables mentioned above. Before proceeding with this development, we would like to mention that the data analysis that is being presented is not intended to introduce the already known techniques used for the analysis of expression microarray data, we will mention in a very succinct way the guidelines of each of the algorithms (index.affx and LimmaGUI) and of some changes that we considered very important for us.

Method MAS 5.0

The Drosophila GeneChip1 consists of 14010 genes each constituting a probe set. Each probe set consists of a series of probe pairs (between 11 and 20). Each probe pair is composed of the complementary (PM) Perfect Match and MM mismatch. The PM consists of 25 oligonucleotides designed to be perfectly complementary to a cRNA sequence to be hybridized. The MM is 25 oligonucleotides designed to be complementary to a cRNA sequence except for one base that has been changed and occupies the middle position (position 13). MM serves as a specific hybridization control when compared to the corresponding PM (the hybridization signal of the MM should always be lower than that of the PM). Each PM or MM forms what is called a "probe set" which is composed of thousands of copies of a given sequence. The different "probe pairs" forming a "probe set" are distributed throughout the array to avoid problems at the time of hybridization, for example, if there is a bubble in an area of the chip, it does not affect the entire "probe set". The multiple probes (oligonucleotides) that have been used represent a gene, but the signals of the different probes for the same genes are not the same, so statistical methods must be applied to find the signal of each probe set.

Once the PM and MM are obtained as a signal of the set of oligos that have been used to detect a gene from both the experimental and control groups, each one from the experimental group will be compared with all those from the control group by constructing a matrix.

Tijk,i'j'k'

$$T = \begin{pmatrix} r_{1jk,1j'k'} & r_{1jk,2j'k'} & r_{1jk,3j'k'} \\ r_{2jk,1j'k'} & r_{2jk,2j'k'} & r_{2jk,3j'k'} \\ r_{3jk,1j'k'} & r_{3jk,2j'k'} & r_{3jk,3j'k'} \end{pmatrix}$$

Where the subscripts i,i' represent the pro-besets of both the experimental and control groups, the subscripts j and j' indicate the number of probe pairs and the subscripts k , k' represent the number of oligos used to identify each of the PM and MM. We extract the minimum

$$Min\left(r_{ijk,i'j'k'}\right)$$

and maximum

$$Max\left(r_{ijk,i'j'k'}\right)$$

value for all

$$i,i'\left(1\rightarrow 3\right)$$

(replicates) from each of the sub matrices and calculate the mean with these two values. A new matrix will be constructed whose values will be given by Sun where the subscripts u, n indicates the means of each of the comparisons. In order to find the signal of each gene, we have to calculate the mean of all matrix elements.

$$S_{un}=\frac{Max\left(r_{1jk,1j'k'}\right)+Min\left(r_{1jk,1j'k'}\right)}{2}.$$

For the ISS experiment the S matrix was 4 elements due to the two replicates. For the RPM experiment with cold treatment, the S,

$$S=\begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix}$$

was 9 elements due to the three replicates. When comparing the experimental results of both ISS and RPM, it is observed that within the 14010 genes there is the presence of noise, either biological or experimental, due to an external factor or due to the image capture of each pro-beset. In order to solve this problem, two filters were proposed. Before starting with the respective filters, the signals had to be separated, maintaining the order of location in each cell. Two new columns were created, the first one identified the genes expressed with value one (1) and zero (0) those that did not meet this condition. In the second column we identified the genes that are inhibited by identifying them with a value of one (1) and zero (0) for those that do not meet this condition. Once separated, we started the respective filters. The first filter takes into account the dispersion of the signals of each sample, either from the control or experimental group, with respect to the mean value; once these values are detected, the MAS5 algorithm selects those samples that fall within the acceptable range according to the standard deviation where it will calculate the probability (Wilcoxon's Signed Rankn Test) of the presence of the gene per replicate. According to the value, the MAS5 algorithm will label each Probeset for each of the replicates with the letters P, M and A where we have designated to these letters the following values, for operational convenience, P(15): Present M(12): Marginal A(1): Absent (Table 1). This information, MAS5, presents it in columns, each column indicating one replicate. For the ISS experiment, two columns were obtained and for the RPM there were three columns. The values of P_value<0.04 are all those genes with Presence character (P), the values of P_value>0.06 are genes have much variability in their signal considering them Absent genes (A), the genes whose value of 0.04<P_value<0.06 are all those Marginal genes (M), the presence of the gene will depend on the number of elements (P_value) that fall in this range [2]. Once this information is obtained, we begin to discard all those genes that meet the following condition: G>16+12(n-2), being ^{n} the number of replicates for each gene in either the control or experimental group. Where G represents the selection of the gene.

Exp. X Cont	Minimum permutation	G: Minimum values
2 x 2	P A	16
3 x 3	P M A	28
4 x 4	P M A M	40
4 x 5	P M A M M	52
6 x 6	P M A M M M	64

Table 1: Minimum values of the permutations between the replicates of each of the experiments.

If combining the respective labeled values for each of the genes results in a value equal to or less than the minimum values of G, this gene will be absent, it will not be taken into account during the whole process of data analysis.

With this information we elaborate two matrices

$$F_{ISS}^{Increased}=\begin{bmatrix} f_{1,1}^{I} \\ f_{2,1}^{I} \\ \\ f_{14010,1}^{I} \end{bmatrix} \text{ and } F_{ISS}^{Decreased}=\begin{bmatrix} f_{1,1}^{D} \\ f_{2,1}^{D} \\ \\ f_{14010,1}^{D} \end{bmatrix}$$

whose matrix elements are zeros (0) and ones (1) for those genes that are expressed and those that are inhibited, respectively. By adding the two matrices

$$F_{ISS}^{Increased}+F_{ISS}^{Decreas}$$

we obtain the matrix

$$P_{ISS}=\begin{bmatrix} f_{1,1}^{I}+f_{1,1}^{D} \\ f_{2,1}^{I}+f_{2,1}^{D} \\ \\ f_{14010,1}^{I}+f_{14010,1}^{D} \end{bmatrix}, \text{ if } f_{i,1}^{I}=1 \text{ the value of } f_{i,1}^{D}=0 \text{ ó or if } f_{i,1}^{I}=0 \text{ the value of } f_{i,1}^{D}=1.$$

This new

$$P_{ISS} = P_{i,1}^{ISS}$$

matrix gave us a list of all genes with presence character (whose matrix elements were 1) from the experiment performed on the ISS. In the same way the

$$P_{RPM} = P_{i,1}^{RPM},$$

matrix was calculated, which lists all genes with presence character from the experiment that was performed in the RPM microgravity simulator. In order to find the genes with common presence character between the experiments performed in the ISS and in the RPM simulator we had to define the

$$P = \left\lfloor P_{i,1}^{ISS} x P_{i,1}^{RPM} \right\rfloor_{14010,1}$$

matrix (for all i=1 to 14010). The second filter compares all those genes present, PISS and PRPM with their respective controls CISS and PRPM whose matrix element were obtained in the same way as the previous cases (for each experiment we have used different controls both exposed to the same conditions) both from the experimental group with the samples from the control group. The second filter analyzes all those genes that come from multiplying the matrix elements

$$F_{ISS}^{1} = \left\lfloor P_{i,1}^{ISS} x C_{i,1}^{ISS} \right\rfloor_{14010,1},$$

if the matrix elements of the s are one (1) it indicates that they are genes with presence character when compared with their respective ground control. In the same way we calculate the matrix for the experiment performed at RPM

$$F_{RPM}^{1} = \left\lfloor P_{j,1}^{RPM} x C_{j,1}^{RPM} \right\rfloor_{14010,1}.$$

For the ISS experiment, two replicates were used for the control and an experimental group, the maximum number of permutations (comparisons between the control and experimental groups) for each gene was 4; for the RPM, three (3) replicates were used, obtaining a maximum of 9 permutations per gene. With this information we elaborated two new matrices, the S_{ISS} and S_{PRM}. MAS5 calculated the P_Value that indicated the probability of signal dispersion of the experimental group with respect to the control group. Once the comparative P_Values were obtained, we classified the genes into two new groups, maintaining their initial location given by the ID. In the first group, those genes with a P_Value<0.5 constitute the group of genes that are expressed as present (Wilcoxon Signed Rank Test); in the second group, those genes whose P_Value>0.5 represent those genes that are inhibited by the control group. With this information, two matrices

$$S_{ISS}^{Increased} = \begin{bmatrix} s_{1,1} & s_{1,2} & s_{1,3} & s_{1,4} \\ s_{2,1} & s_{2,2} & s_{2,3} & s_{2,4} \\ & & & \\ s_{14010,1} & s_{14010,2} & s_{14010,3} & s_{14010,4} \end{bmatrix}$$

and

$$S_{ISS}^{Decreased} = \begin{bmatrix} s_{1,1} & s_{1,2} & s_{1,3} & s_{1,4} \\ s_{2,1} & s_{2,2} & s_{2,3} & s_{2,4} \\ & & & \\ s_{14010,1} & s_{14010,2} & s_{14010,3} & s_{14010,4} \end{bmatrix}$$

whose matrix elements will be ones (1) and zeros (0) are constructed. Now, if P_Value<0.5 we place in the cell the value one (1) or zero (0) obtaining a matrix of 14010x4 elements that represent the genes that are expressed. Following the Affymetrix protocol, we select those P_Value of each of the above mentioned groups in three zones. The first of them will be genes with P_Value<0.002 constituting those genes that are expressed with character of presence, the second zone is constituted by those marginal genes whose P_Value falls in the interval (0.002 to 0.003) being genes that can possibly be taken into account as present, the third zone is constituted by genes that should not be taken into account for the analysis although they are genes with character of presence whose interval is 0.003<P_Value<0.997. For genes with P_Value>0.5 will also have three zones, (the same) which are the complementary values with respect to the unit value [0.998<P_Value<1 (Zone P), 0.997<P_Value<0.998 (Zone M) and 0.5<P_Value <0.997) (Zone A)].

The 75% of the comparative P_Values that identified a single gene fall in the marginal zone, both for genes that are expressed or inhibited were not taken into account even if they were genes present according to the first filter [3]. This analysis was for experiments in which more than three replicates have been taken as those of the RPM; in the case of the ISS no genes, as present, are accepted as falling in the marginal zone. Following this procedure, we produced two new matrices

$$P_{ISS}^{Increased} = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ & & & \\ p_{14010,1} & p_{14010,2} & p_{14010,3} & p_{14010,4} \end{bmatrix}$$

and

$$P_{ISS}^{Decreased} = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ & & & \\ p_{140101} & p_{140102} & p_{140103} & p_{140104} \end{bmatrix}$$

whose matrix elements were zeros (0) and ones (1). The value one (1) if its comparative P_Value falls in the increased or decreased zone and zero (0) because it did not meet this condition. In order to find the list of genes that pass the second filter, we add all the row elements of the matrix

$$P_{ISS}^{increased} \sum_{j=1}^{4} p_j,$$

the maximum value was 4 and the minimum value was zero (0), with these values we elaborate the following matrix

$$F_{ISS}^{2,Increased} = \left[\ \right]_{14010,1},$$

if the sum is equal to 4 this cell will be assigned the value of one (1) or zero (0). The value one (1) means that the gene is expressed as present. The same is done for the decreased group of the ISS

$$F_{ISS}^{2,Decreased} = [\ ]_{14010,1}$$

In order to find the list of genes that are expressed and inhibited according to the two filters we added

$$F_{ISS}^{2,Increased} + F_{ISS}^{2,Decreased}$$

where we obtained a new matrix

$$F_{ISS} = \begin{pmatrix} i_1 \\ i_2 \\ \\ i_{14010} \end{pmatrix}$$

whose matrix elements were zeros (0) and ones (1). The same filter was used for the RPM data; the only difference was that the maximum value of the matrix elements of

$$F_{RPM}^{2,Increased} = [\ ]_{14010,1}$$

and

$$F_{RPM}^{2,Decreased} = [\ ]_{14010,1}$$

was 9. By summing these two matrices, we produced a new whose matrix elements are zeros (0) and ones (1). When the sum of the matrix elements

$$F_{RPM}^{2,Increased} + F_{RPM}^{2,Decreased}$$

was less than 8 the value corresponding to the matrix element

$$F_{RPM} = \begin{pmatrix} r_1 \\ r_2 \\ \\ r_{14010} \end{pmatrix}$$

was zero (0) and one (1) when it was greater than 7. This tells us that 75% of the comparative P_value for each gene have fallen in the marginal zone and should not be taken into account. Finally, by multiplying the matrix elements

$$\begin{pmatrix} i_1 x r_1 \\ i_2 x r_2 \\ \\ i_n x r_n \end{pmatrix}$$

we obtain a new matrix E whose matrix elements are zeros (0) and (1). The value zero (0) tells us that the gene has no presence in both ISS and RPM, the value one (1) tells us that the gene has presence in both ISS and RPM.

RMA method-AffylmGUI

Since it is the same comparison as the previous case, the same PM and MM will be used, both values being invariable with respect to the algorithm with which the results are analyzed. Once the files with CEL extension were obtained (they store all the information of the microarray), two files in txt format were created containing the following information (Tables 2 and 3).

Name	File name	Target
ISS_cold_.1	RMarco_Droso_FG3.CEL	ISS_Cold
ISS_Cold.2	RMarco_Droso_FG4.CEL	ISS_Cold
ISS_Cold_Cont.1	RMarco_Droso_GG4.CEL	Control_ISS_Cold
ISS_Cold_Cont.2	RMarco_Droso_GG5.CEL	Control_ISS_Cold

Table 2: Experimental group of the ISS crossed with their respective controls 1 g.

Name	File name	Target
RPM_Cold.1	RMarco_Droso_5.CEL	RPM_Cold
RPM_Cold.2	RMarco_Droso_7.CEL	RPM_Cold
RPM_Cold_Cont.1	RMarco_Droso_81.CEL	Control_RPM_Cold
RPM_Cold_Cont.2	RMarco_Droso_82.CEL	Control_RPM_Cold

Table 3: Experimental group of the RPM+cold crossed with their respective controls 1 g.

The first column of both Table 2 and Table 3 indicates the name of each replicate, the second column indicates the name of the files containing the entire raw microarray signal in Cel format and the third

column indicates the object of study (control group and experimental group). These files were read by R-AffylmGUI, the RMA (Robust Multiarray Avereging) option was selected to normalize the signal and finally BH was selected to calculate the P_Value of each of the statistically differentiable genes [4]. The R-AffylmGUI program created a file in txt format that stored all the necessary information for its respective analysis.

Correlation method

This new method that we have implemented consists in the fusion of both methods previously mentioned. Using those genes with presence character according to MAS5 and the normalized signal with its respective P-Value according to BH of R-AffylmGUI. With these two variables which we call dynamic variables, we started the search for an algorithm that calculates the respective correlations of the quantified fluorescent signals of a given set of genes. The algorithm we implemented consisted of fixing the first value of the P_Value and then varying the values of the Cut_Off with steps of 0.001 and 0.01 respectively. For each iteration that the algorithm ran, it calculated the number of genes present in the ISS and RPM with their respective correlation for each group of data. Then the program ordered these groups of data according to the correlation value with a step of 0.1 obtaining 10 columns, in each of the cells was counted the number of times or frequency that a certain gene appeared for each iteration performed by the program, finally, the program displays all the genes with a certain correlation between 0 to 1. The minimum number of genes that the program accepted before jumping to the next iteration was 3 genes.

Having a small number of genes with a correlation of one (1) is not a reliable result, so we decided to shorten our number of iterations in order to decrease the program execution time. This does not imply that genes have been eliminated from our analysis, but rather that they are part of some possible correlation group.

The initial estimate of the time taken by the program to analyze the 14010 genes when calculating the correlation matrix and the formation of groups according to the degree of affinity or correlation was approximately six (6) months (Qbase Software). In order to have a faster calculation time, we worked with the genes with presence character, according to MAS5, reducing the execution time to 45 days. Possibly Qbase is not a good program for the analysis of our data; for this reason, it is proposed to use Fortran or Java as programming language to compile our algorithm in order to reduce the execution time.

Design of the Taqman probes

For qRT-PCR the following probes were purchased from AB applied biosystems:

qRT-PCR: Relative quantification by real time PCR was performed on a 7900HT fast real time PCR system (Applied Biosystem). Quantification was performed using FAM-labeled Taqman probes FAM-labeled Taqman probes for the genes in study, designed by applied biosystems (Table 4) and VIC for the endogenous control [5]. The amplification reaction conditions were established by the established by the real-time PCR and sequencing service of the Instituto de biomedical research institute "Alberto Sols" CSIC-UAM. In all cases they were analyzed in triplicate and in triplicate and in at least 3 independent experiments.

Name of gene	Exon	Length	Stock	Sequence
IMPL3	_03 _04	93	Dm01841229_g1	ggcgaacatggcattgacaaggatgt gttcctct cgctgccctgcgttctcaatgccaacg gtgtgacat ccgtggtcaagcagatcctgact
CG4726	_1 _2	64	Dm01800475_m1	gcaaactcgtgcccgcccgctatgtgc tggccctcc tggggtccatcggcatggccattgtgta
CG14598	2	134	Dm02368409_s1	aaaagtggaagaagcgcaaaattaa ttcctaaaa taatatttatcgtaaaggaaattcatac acccttctt agcgaccaatggatagcgttttacctttt ccagagc ttcatcggcatgtttttaattgtagta
CG5326	_03 _04	81	Dm02143291_g1	Tggctcgtgccgtgtggctgtactacat tgccaagat cacggagctgttggacaccgtgttcttt gtgctgcgca agaaac
Dnaj-1	1	101	Dm02362419_s1	aacacggggaggttacatcagttagt agaacacttat agatttatacgccgtaaggacatagta cgttttagtac caattctattatgttataaacaataa

Table 4: Probes from AB applied bio-systems for qRT-PCR.

MOTIS search: Consensus sequences were extracted from the TFSearch database. It was observed that not all of them had an identified gene. When resuming the study to identify the missing genes, it was not possible to access the aforementioned database. For this reason, the EnhancerAtlas 2.0 database was used to complete the

identification of genes. To identify the remaining Motifs, we permuted all possible options for each consensus sequence and searched the Enhancer database. Of the total list of Motifs extracted from TFSearch, only 22 remained unidentified. Once the final list was obtained, we searched for each of these genes in the 807 microgravity-responsive IDs in order to identify the microgravity-responsive regulators.

Once the final list is obtained, we cross the obtained list with the list of the 807 IDs that are sensitive to microgravity (We compare the IDs of each list. For the location of the IDs of the list of regulators found, we use the raw database of microarrays), in this way we identify those regulators that respond to microgravity. For a better visualization of this result, and to verify that there is some relationship between the lists of regulars obtained with the list of 807 IDs, we use the string database [6]. We entered the entire list of regulatory genes and only a sample of 329 genes that respond to microgravity. The value obtained from the sample was calculated with a total of 741 (Table 5), confidence value of 95%, margin of error of 4% and with a probability of occurrence and non-occurrence of 5%. For the selection of the 329 genes, Excel was used, the length of the mRNA was ordered from smallest to largest, then, in a new column, the random() command was used and then ordered from smallest to largest, the first ones were selected, 332 genes.

Isoforms search: Using PubMed we were able to obtain the mRNA length for each isomorph of the microgravity-responsive genes. Once the final list was obtained, the lengths of each isoform were ordered from shortest to longest (the first isoform for each microgravity-responsive gene was selected). They were grouped by each 100 bp of isoform length and the average ISS+Cold signal was obtained. The first previous result (not presented in the manuscript) showed that in a region with a center at 1000 bp there was a change in expression. This previous result led us to propose two new groupings of genes. The first group corresponded to genes with positive expression in the ISS+COLD smaller than 1000 bp and with negative expression larger than 1000 bp. The second group corresponded to genes with negative expression in the ISS+COLD less than 1000 bp and with positive expression greater than 1000 bp. With a center at 1000 bp, a region of +/-20 bp was selected in order to select genes with a specific length of its mRNA according to the selected isoform, with the purpose of detecting a possible critical mass that could regulate the expression of genes that respond to microgravity.

Results

Figure 1 microgravity responsive genes. The MAS 5.0 method was used for this analysis. The blue dots represent 807 ID values of a set of statistically distinguishable genes [7]. Each blue dot represents an ordered pair of values, the first of which (horizontal axis Microgravity Simulator RPM) represents the mRNA expression levels of the pupal stage of *Drosophila melanogaster* exposed for approximately 4 days to simulated microgravity. The second ordered pair (vertical axis International Space Station ISS) represents the mRNA expression levels of the pupal stage of *Drosophila melanogaster* exposed for approximately 4 days to microgravity generated inside the International Space Station ISS. We observed 801 IDs responding in the same manner to microgravity. Of this group of IDs only 338 IDs have positive expression in both the ISS and the RPM microgravity simulator, 467 IDs have negative expression in both the ISS and the microgravity simulator. Only 6 IDs have opposite values (They have positive expression in the RPM and negative expression in the ISS).

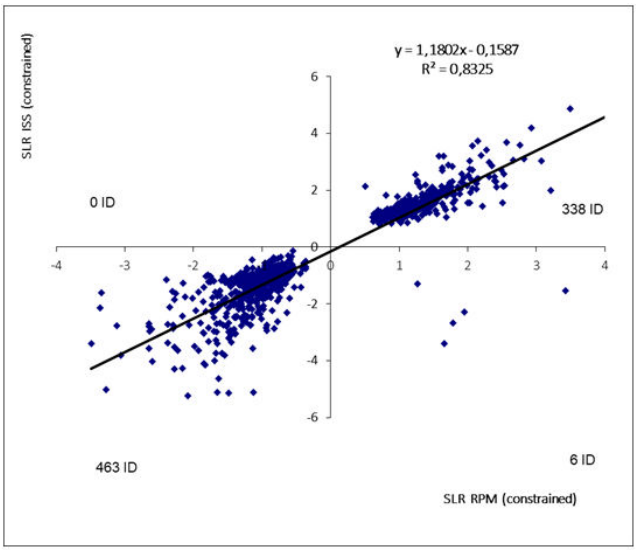

Figure 1: Microgravity-responsive genes (The MAS 5.0 method).

Figure 2 microgravity responsive genes. RMA method-AffylmGUI was used for this analysis. The blue dots represent 806 ID values from a statistically distinguishable set of genes. Each blue dot represents an ordered pair of values, the first of which (horizontal axis microgravity simulator RPM) represents the mRNA expression levels of the pupal stage of *Drosophila melanogaster* exposed for approximately 4 days to simulated microgravity. The second (International Space Station (ISS) vertical axis) represents the mRNA expression levels of the *Drosophila melanogaster* pupal stage exposed for approximately 4 days to microgravity generated within the International Space Station (ISS). We observed 798 IDs responding in the same manner to microgravity. From this group of IDs only 360 IDs have positive expression in both the ISS and the RPM microgravity simulator, 438 IDs have negative expression in both the ISS and the microgravity simulator. Only 8 IDs have opposite values.

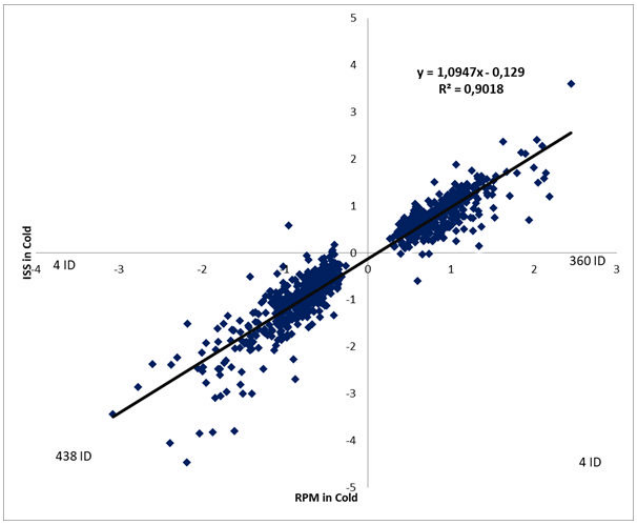

Figure 2: Microgravity-responsive genes (RMA method-AffylmGUI)

Table 5 represents the RMA-The MAS 5.0 method was used for this analysis. Of the 807 statistically distinguishable IDs that respond to microgravity only 30 probes used to detect mRNA sequences share

73 *Drosophila melanogaster* genes, it can also be noted that 15 genes are shared by 33 IDs, 3 IDs have no recognized gene, 4 IDs have no cDNA information, 598 IDs correspond to 598 genes (one to one ratio), 81 IDs have negative signal (set of genes repressed in microgravity) with mRNA length is less than 1000 bp, finally there are 58 IDs with positive signal (set of genes expressed in microgravity) with mRNA length greater than 1000 bp [8]. We cannot identify the exact set of microgravity-responsive genes, but we can state the number of microgravity-responsive IDs. The table shows a total of 807 IDs that can identify 832 possible microgravity-responsive genes of *Drosophila melanogaster*.

	Total Id		**Total genes**
ID shared by different genes	30	Genes	73
Gene shared by different id	33	Genes	15
Id with negative signal and cDNA less than 1000	81	Genes	81
Id with positive signal and cDNA greater than 1000	58	Genes	58
Correct Id and gene	598	Genes	598
Id with unknown gene	3	Genes	3
Id with unknown cDNA	4	Genes	4
Total:	807	Total:	832

Table 5: Number of ID and genes.

RMA method-AffylmGUI was used for this analysis were used for data analysis. Each bifurcation shown in the figure represents the degree of correlation greater than zero and less than one. Each black line represents the IDs that are part of the correlation clusters. The image shows that only six of them share different clusters. 788 IDs have a degree of correlation between [0.9-1] and 806 IDs less than 1 (Figure 3).

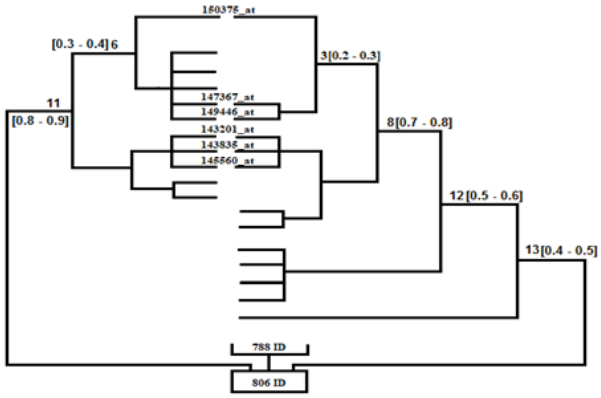

Figure 3: Cluster of correlations.

Frequence of genes responding to real and simulated microgravity

The last step of the algorithm was to calculate the number of times that the IDs appeared for each of the 10750 iterations or cycles performed by the program. In order to have comparative values, we chose to calculate the frequency with respect to 100% of the 10750 iterations, for each of the genes. It was observed that the IDs (red dots in Figure 4) appear in more than 85% of the 10750 iterations and are the same six that appear in the correlation cluster (Figure 3).

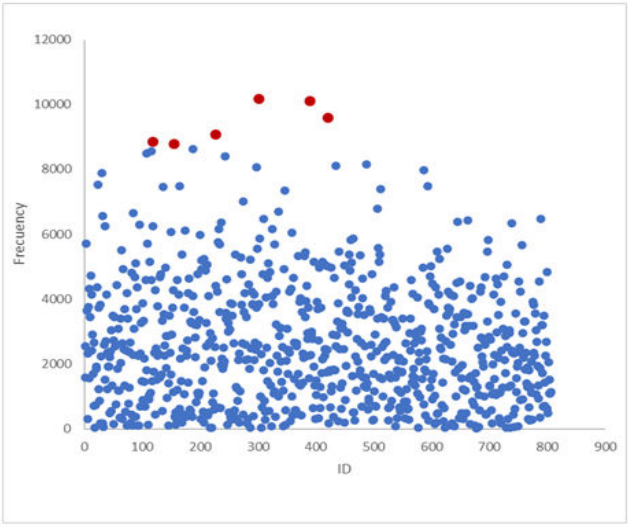

Figure 4: Frequency of genes responding to real and simulated microgravity.

The MAS 5.0 method and the RMA method-AffylmGUI were used for data analysis. The vertical axis indicates the frequency with which the IDs appear for each cycle per-formed by the algorithm (Appendix A1), the horizontal axis indicates the IDs which were ordered from lowest to highest. The blue dots indicate the ID, the red dots indicate the IDs that appeared with a frequency greater than 85%of the 10750 cycles performed by the program (From left to right: 143201_at, 143835_at, 145560_at, 147367_at, 149446_at and 150375_at). Initially it was thought to take all the statistically differentiable IDs in microgravity

both in the ISS and in the RPM, the total number of IDs was 6332, the program execution or compilation time was estimated to exceed 6 months. To overcome this temporal problem, only those IDs that appear in both the experiments carried out in the ISS and in the RPM were used. The execution or compilation time of the program was 45 days. Data processing was performed on an Intel ® Core ™2 Quad CPU Q6600@2.4 GHz, with an installed memory of 8.0 GB, 512 MB RAM with a 64-bit operating system [9].

qRT-PCR studies to validate the correlation algorithm

To validate the results of the gene experiment (cold pretreatment and microgravity); a new experiment was designed where the only variable was microgravity. Therefore, we decided to repeat the gene experiment under simulated microgravity conditions, omitting the treatment of the specimens at 14°C. The main objective of this experiment was to study by qRT-PCR the behavior of the 6 genes identified by bio-informatics analysis of the microarray data obtained from the gene experiment. Only 147367_at (Gen CG18431) was not known to which gene it corresponded The experiment consisted of subjecting pupae to no gravity for 21 hours and 4.5 days, respectively, and then obtaining mRNA from these samples and performing qRT-PCR to analyze the effects of gravity on the expression of the identified genes (Figure 5).

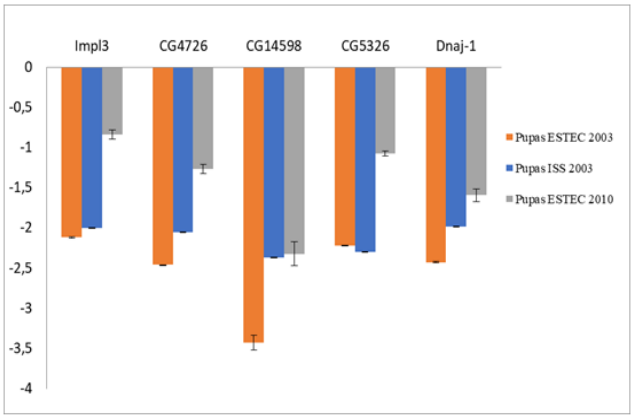

Figure 5: Expression levels of 5 of the six overlapping genes in ISS and in simulated microgravity.

Shown are the expression levels of the five genes measured by RT-PCR with Taq-Man probes. Expression levels of the five genes measured by RT-PCR with Taq-Man probes are shown. The expression levels of the five genes measured by RT-PCR with Taq-Man probes are shown. Expression levels are shown relative to their expression in normal gravity. 18S was used as the endogenous gene. The red columns show the results obtained in simulated microgravity in the initial experiment. The blue columns show the results obtained on the ISS. The green columns show the results obtained in simulated microgravity in 2010.

qRT-PCR studies to validate the experiment

To test the specificity of these results, we decided to study the response to simulated mi-crogravity in genes whose expression a) decreases but does not belong to the identified overlapping group of genes (mef2), b) does not vary (cf2) and c) increases in response to microgravity (CG5703, Surf1, CG8885 and CG1970). Mef2, CG5703, Surf1, CG8885 and CG1970 were identified in the initial analysis as belonging to the microgravity-responsive group (Figure 6).

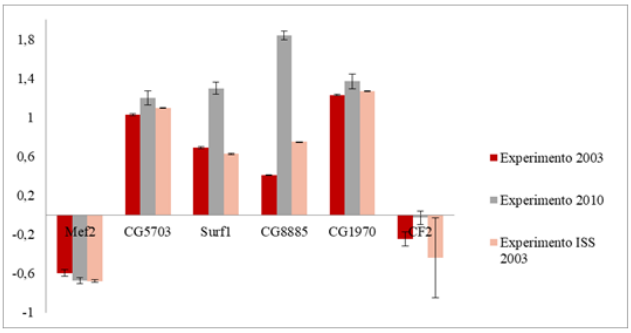

Figure 6: qRT-PCR studies to validate the experiment.

The expression levels measured by RT-PCR of two genes whose expression did not vary substantially in the initial experiment, Mef2 and Cf2, and four genes whose expression increased in real or simulated microgravity are shown. The expression levels relative to their expression in normal gravity are shown [10]. The endogenous gene used was 18S. Red columns show the results obtained in simulated microgravity in the initial experiment. The blue columns show the results obtained on the ISS. The green columns show the results obtained in simulated microgravity in 2010

Regulatory genes that do not respond to microgravity

Consensus sequences were extracted from the TFSearch database. It was noted that not all had a unique sequence that identifies a gene (one-to-one relationship). When the study was resumed to identify the missing genes, it was not possible to access the aforementioned database. For this reason, the Enhancer Atlas 2.0 database was used to complete the identification of genes. To identify the remaining motifs, we permuted all possible options for each consensus sequence and searched the Enhancer database, for example (Gen Br. Consensus Sequence: TTAAWKR. Motifs Sequence: TTAA(AT)(GT)(AG). Consensus Sequence: TTAAAGA). Of the total list of Motifs extracted from TFSearch, only 22 remained unidentified. Cross-referencing the final list with the 807 IDs that were sensitive to microgravity would identify those regulators that are responsive to microgravity. We compared the IDs of each list. For the location of the IDs from the list of regulators found (TFSearch or Enhancer Atlas 2.0), we used the raw data from the microarrays. We first found the gene by searching the raw data and finally selected its corresponding ID. For a better visualization of this result, and to check if there is any relationship between the list of regulars obtained with the list of 807 IDs, we used the string-protein database. The basic principle in string, each protein-protein interaction is annotated with one or more 'scores'. Importantly, these scores do not indicate the strength or the specificity of the interaction. Instead, they are indicators of confidence, *i.e.* how likely string judges an interaction to be true, given the available evidence. We created a new database which consisted of two groups, the first group consisting of 134 IDs (112 genes) corresponding to the regulators that were previously found. The second group consisted of 741 genes that respond to microgravity (Table 5). A total of 853 data (Genes) were entered into string protein, exceeding the limit of data to be analyzed (We include the list of regulatory genes). A sample of 329 of the 741 genes was selected. The value obtained from the sample was calculated from a total of 741 corresponding to microgravity sensitive genes, 95% confidence value, 4% margin of error and with a

probability of occurrence and non-occurrence of 5% [11]. For the selection of the 329 genes excel was used, the mRNA length was sorted from shortest to longest, then in a new column the random command was used and then the random column was sorted from shortest to longest. The first 329 genes were selected. Figure 7 shows 67 genes of the 329 in the sample that have no relationship for the moment not reported in any database, 4 of the 112 regulatory genes that have no report. According to the results obtained in the string-protein database, under normal 1 g (Earth) conditions there is at least one relationship between all genes that have responded to microgravity with the list of regulators. If the regulators do not respond to microgravity how the genes were regulated in microgravity.

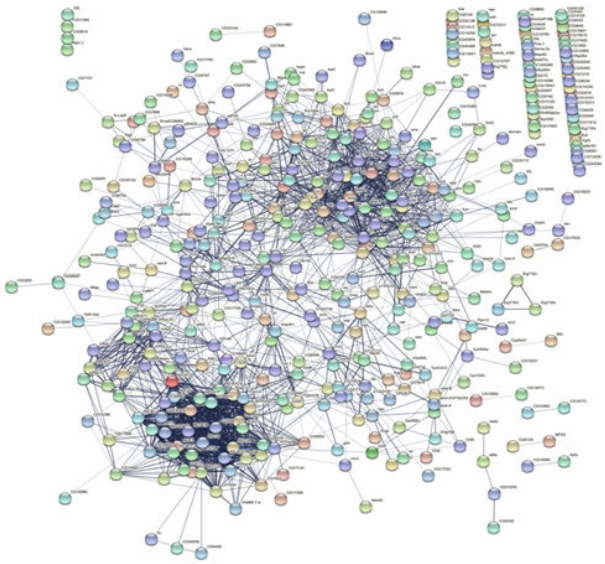

Figure 7: Regulatory genes that do not respond to microgravity. Each of the spheres represents a gene. Each line joining the spheres represents some relationship reported in some database. The column of spheres shown at the top left (4 spheres or genes) corresponding to the 112 regulatory genes that were entered for analysis in the string-protein database. In the image it can be seen that the 4 genes do not have any report with the rest of the genes. On the upper right are 67 spheres distributed in 4 columns of spheres corresponding to the 741 microgravity-responsive genes. The string-protein database indicates that the 67 genes do not have any report in any database that relates them to the rest of the genes.

Expression levels of microgravity-responsive genes according to their cDNA length

To obtain the results we had to make use of the PubMed database, we extracted the mRNA length (bp) of each isomorph of the microgravity-responsive genes (the first isoform was used for the study). Once the final list was obtained, the lengths of each isoform were ordered from shortest to longest. They were grouped by every 100 bp; it could be observed that above 1500 bp there were ranges where the number of signals from the probes did not exceed three, for those cases the range was ex-tended to 1000 bp. For each cluster the mean signal was obtained (for this study we only used the signals coming from the ISS+Cold. The first previous result (not presented in the manuscript) showed the existence of a region with center at 1000 bp where there was a change in expression. This first result led us to propose two new gene clusters. The first group corre-sponded to 600 genes, with positive expression in ISS+COLD smaller than 1000 bp and with negative expression larger than 1000 bp. We also observed a second group formed by a total of 135 genes that had different behavior than the previous case (genes with negative expression in the ISS+COLD lower than 1000 bp and with positive expression higher than 1000 bp).

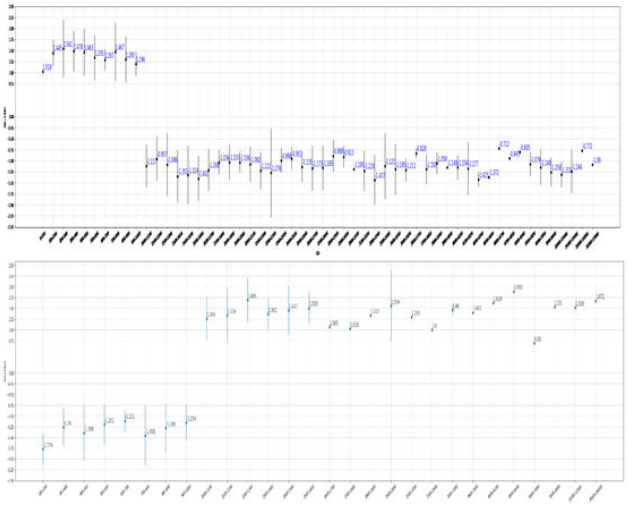

Figure 8: Expression levels of microgravity-responsive genes according to their mRNA length.

The image shows two plans in both panels the horizontal axis indicates the mRNA size in bp. The vertical axis indicates the signal obtained from Affymetrix micro-arrays. Panel "A" shows a set of genes in the positive zone in the range of (0 to 1000) bp. A set of genes in the negative zone in a range of (1000 to 17000) bp [12]. The total number of genes analyzed in panel "A" was 600. Panel "B" shows a group of genes in the negative zone in a range of (0 to 1000) bp. A set of genes in the positive zone in a range of (1000 to 17000) bp. The total number of genes analyzed in panel "B" was 135. Each black line indicates the standard deviation. For the calculation of the standard deviation, all signals within a range of 100 bp were taken. In some cases there were no signals in the established range, so it was decided to increase the range to 1000 bp.

Possible critical mass (kD) value limiting expression

Two panels of tables are shown (Table 6). The first column of each panel represents the names of genes near the 1000 bp cutoff point. The second column of each panel represents the protein size (kDa) of each of the transcripts close to the cutoff point. At the end of the second column (both panels) the average values and the respective standard deviations are presented. Similar values were obtained with standard deviations of less than 7% with respect to the mean value.

Name gene	Masa (kDa) reverse expression at the distance of the cDNA	Name gene	Masa(kDa) direct expression at the distance of the cDNA
CG31743	37.8	*CG7998*	35.3
CG5567	36.4	*CG3597*	37.8
Gapdh1	35.4	*NUDC*	37.8
Dnaj-1	37	*Impl3*	35.5
CG6906	28.3	*CG7860*	34.9
H	37	*Scf*	38
CG4778	38.2	*CG30101*	38.1
Impl1	38.3	*CG3957*	36.1
Mass average	36.05	Mass average	36.6875
Standard deviation	3.276758	Standard deviation	1.366369

Table 6: Mass values (kDa) of transcripts close to the boundary value.

The vertical axis indicates the mass in kDa, the horizontal axis shows the name of each analysis group. The first column represents the mean value of mass (kDa) with inverse expression to mRNA size [13]. The second column represents the mean mass value (kDA) with direct expression to mRNA size. The third column represents the mean value of the mass (kDa) of all transcripts close to 1000 bp. As the mean values are approximately equal and their standard deviations crossed, we proceeded to calculate the mean value using the whole data set presented in Table 6 (Figure 9).

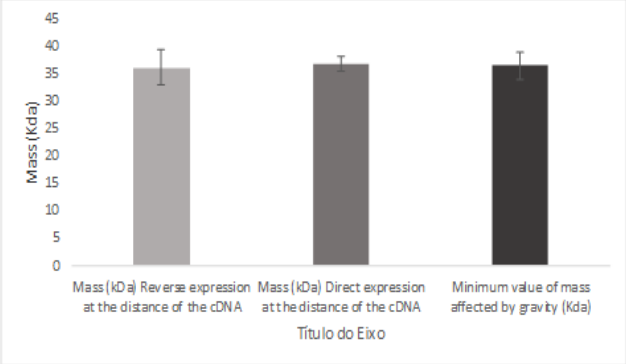

Figure 9: Possible critical mass (kD).

Discussion

Our hypothesis that we initially planned was based on the studies performed, in normal conditions at 1 g, by Arredonde's group, where he explored that the *CF2* gene plays a very important role in muscle formation in *Drosophila melanogaster*. Proposing that *CF2* regulates *Mef2* expression through a Feedforward loop, following this same line, Bagni's group demonstrated that *CF2* expression depends on the myogenic factor MEF2. It is obvious to assume that if we alter the modulus of gravity to values close to zero, muscle formation is altered. A consequence of prolonged exposure to microgravity is muscle atrophy (loss of muscle mass). Since the experiments are carried out in space, it can be hypothesized that cosmic radiation together with microgravity play a synergistic role which would affect the results of the experiments performed under real microgravity conditions. In order to test this hypothesis, Okada's group proved that exposure to artificial 1 g, inside the international space station, prevents muscle atrophy at the molecular level, studies carried out on the basis of gene expression profiles. Okada's result demonstrates that muscle atrophy due to lack of gravity is not due to a consequence of cosmic radiation. To investigate the molecular mechanisms of muscle atrophy under microgravity conditions, Yamakuchi's group examined para-spinal muscles of rats after 14 days of space flight. They found that 42 genes down regulated their expression levels and increased expression of heat shock proteins. Using Northern blotting techniques, Yamakuchi, also demonstrated the down regulation of Myocyte-specific Enhancer binding Factor 2C (MEF2C) and MEF2C-related genes, including aldolase A and muscle ankylatin, observed that MEF2C worked cooperatively with Myostatin (MSTN). MSTN plays an important role in skeletal muscle development and regulates muscle fiber type by modulating *MyoD* and *MEF2C* gene expression in newborn piglets. In the same vein, Li's group, I used biological models such as mice, rats and pigs that have proposed that Mef2C could modulate and restrict myogenesis by activating MSTN. Although the experiments performed by Yamakuchi, Xuan, Li and Okada were in mammals and not in D*rosophila melanogaster*, we performed a search for homologs in *Drosophila melanogaster*, for MSTN it was MYO (ID: 154466_at), for MYOD it was NAU (ID: 143280_at) and for MEF2C it was MEF2 (ID: 153628_at). Just as MYOD is regulated by MEF2C, Sandmann's group showed that NAU, MHC, MBL, MESO18E and Act57B are regulated by MEF2. As MYO is strongly expressed in muscle and glial cells and has been shown to promote neuronal development and remodeling, prevent age-related muscle dysfunction and prolong lifespan in Drosophila, MYO knockdown in muscles using MEF2-GAL4 does not significantly increase muscle size versus the control without GAL4, what he did observe was that MYO knockdown using DA-GAL4 (DA ID:143117_at) results in larger muscles [14]. Enriquez's group observed that NAU mutant embryos show thinner muscle fibers. The results, according to RMA method-AffylmGUI, showed MEF2 values, obtained in the microgravity simulator and in the international space station, which were -0.59512376 and -0.67911535, inhibiting as well as the studies performed by Yamakuchi in mammals. With the MAS 5.0 method it could be observed that the MEF2 gene does not respond to microgravity and the Act57B gene (ID: 153531_at) does not respond to microgravity, a result contrary to that published by Kupriyanova. As for the mammalian homologous genes to *Drosophila* mentioned by Yamakuchi Aldoa (ID: 154814_at) and Ank (ID: 154044_at) none of them respond to microgravity by any of the above mentioned methods. As for heat shock response genes, in *Drosophila melanogaster*

Figure 10), our preliminary results showed, for both methods (MAS 5.0 and RMA Method-AffylmGUI) alone the genes *Gp93* (ID: 154182_at), *HSP83* (ID: 143198_at), the *HSF* (ID: 141526_at) and DNAJ-1 (ID:143835_at) obtained values ((RPM);(ISS)):((-0.9275;-1.0025); (-1.05863942;-0.9317289)), ((-1.29;-1.0225);(-1.77451198;-1.41163629)), ((-2.964418716;-1.703782084); (-2.6925;-1.665)) and ((-2.4389979;-1.98596274), (-2.235;-1.97)) respectively.

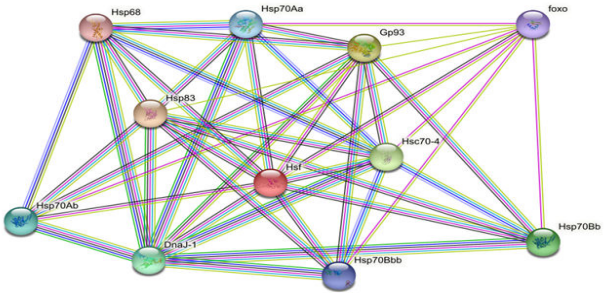

Figure 10: HSF-related shock response genes. Image processed in the string protein database.

These values showed a result contrary to those published by the Yamakuchi group, the rest of the heat shock genes that are directly related to HSF do not respond to real or simulated microgravity (Table 7).

GEN	ID	RMA Method–AffylmGUI		Method MAS 5.0	
		RPM+Cold	ISS+Cold	RPM+Cold	ISS+Cold
Hsp68	143197_at	_____	_____	-3.81	-1.55
Hsp70Aa	149782_at	_____	_____	-5.23	-2.08
Gp93	154182_at	-1.30586	-0.93173	-0.9275	-1.0025
Foxo	149870_at	_____	_____	_____	_____
HSF	141526_at	-2.96442	-1.70378	-2.6925	-1.665
Hsp83	143198_at	-1.77451	-1.41164	-1.29	-1.0225
Hsc70-4	143194_at	-1.05894	-1.0324	_____	_____
Hsp70Bb	141609_at	_____	_____	_____	_____
Hsp70Bb	151036_t_at	_____	_____	-2.31	-0.635
Hsp70Ab	149782_at	_____	_____	-2.08	-5.23
DNAJ-1	143835_at	-2.439	-1.98596	-2.235	-1.97

Table 7: Previous results. Thermal shock response genes in microgravity. The first column shows the gene name, the second column the identifier, the third column shows the signals according to the RMA-AffylmGUI method and the fourth column shows the signals according to the MAS 5.0 method. The black stripes indicate genes that do not respond to microgravity.

As for the MYO, NAU and DA genes none of them respond to the lack of gravity. It is logical to think that if the expression levels of MEF2 decrease the expression levels of NAU, MYO and DA should decrease, in the case of CF2 that does not change in microgravity it would not produce any effect on MEF2, but in microgravity conditions many theories that are the fundamental basis of biology are not fulfilled [15]. Because of the above, we can speculate the existence of some other phenomenon that may be regulating the expression of these genes different from that known in normal conditions on earth, which have not yet been discovered. But science is not speculative; it has to be based on experimental results and the interpretation of these results. Since the appearance of microarrays many researchers have been using this technology in order to be able to find some functional relationship between the thousands of genes studied in a relatively short time, something that could not be done before the 1980's. The works carried out by Tran KN, et al., Vahlensieck C, et al., Thiel CS, et al., Herranz, et al., Frigeri A, et al., Braddock M, et al., and Genchi GG, et al., and Hateley S, et al., used these innovative techniques in different biological models under variable gravity conditions. They all show in their results the number of genes changing under different common conditions which are represented by ensemble plots, they also report the number of genes changing at different gravity modulus conditions, and finally they report studies by functional cluster, to then generate their discussion. On the first point (Affymetrix Drosophila-1) reports that only 55% of the genome is labeled and 45% is pending identification. It is reasonable to assume that chips for some other

species will have the same problem. The results presented in Table 5 show that there are IDs that identify more than one gene as well as genes that are identified by more than one ID. This is the reason why we should not speak of genes that change due to the variation of the gravity modulus but of IDs that change with respect to a control. As for the functional cluster, the results reported by the above-mentioned researchers use information in a database whose experiments have been performed in terrestrial gravity conditions, this point is very important because of the epigenetic variable that was not taken into account, gravity was an evolutionary constant that was part of the biological equilibrium. Accurate, precise and robust regulation of gene expression including comparative epigenetics of closely related and more distant species has given us a better understanding of evolution, being the cornerstone of complex biological life. Such regulation is enhanced by functional enhancers that are composed of concentrated clusters of Transcription Factor (TF) recognition motifs (MOTIFs) that are highly required. As a rule of thumb, a higher number of TF motifs in an enhancer should be positively correlated with higher gene expression. All of this has been occurring in a medium whose gravity value has remained constant in both modulus and direction. It can be concluded that physics is a universal theory while biology depends on the conditions of the environment [16]. The interaction between the molecular machine and the environment generates a synergy that is the key to the biological function of a given number of genes. As it is complicated to understand how genes are responding to microgravity, we set out to study the mRNA lengths of all those genes that respond to microgravity by comparing them with their expression levels given by the array data. We made the following assumption, if the expression machinery (TF) is expressed in the same way both on earth and in microgravity, the expression will depend on the length of the mRNA, the longer the strand the longer it will take the gene to transcribe, therefore the number of copies of the mRNA will be higher the shorter the mRNA, concluding that: The longer the mRNA length the lower the expression. The results show that all the genes reported in the databases that are regulators in *Drosophila melanogaster*, which are in the enhancer region, do not respond to microgravity. Most of these genes did not pass the first filter, which consisted of variability of their respective replicates per experimental group; other genes were eliminated with the second filter by observing too much variability of each experiment with respect to their ground controls. Finally, a small group of genes were eliminated because they were not expressed at the same time (ISS and in the RPM microgravity simulator). This result assured us that possibly the expression of microgravity-responsive genes would depend on the length of their mRNA. The next step was to list all genes that have a single ID and those IDs that only identify a single gene (Table 5). Once the final list was obtained, we searched for the size of each gene's mRNA, using its first isoform. Then, we ordered them from largest to smallest. In the first results (not presented in this manuscript) we observed a bifurcation around 1000 bp. Two groups of data were formed those genes that were expressed at lengths less than 1000 bp and inhibited at lengths greater than 1000 bp (Figure 8a) and a second group formed by those genes that are inhibited at lengths of less than 1000 bp and expressed at mRNA lengths greater than 1000 bp (Figure 8b). This second group which has expression opposite to its mRNA length consisted of a small group of genes (139 genes) (Table 5). The results (Figure 8a) conform to those published by Lyu reporting that proteomic studies of neurospore, yeast, fly, worm and mouse showed protein abundance which correlated negatively with protein length across the genome [17]. Lyu by keeping the same enhancer in their experiments performed on land, this leads us to think that enhancers respond in the same way both on land and in microgravity and that the down- or up-expression will only depend on their mRNA length. In order to accomplish this research we first set out to search for a small group of genes by grouping them according to their degree of correlation. We first checked if the list of genes (807 genes) presented in the manuscript of R. Herranz was correct. According to Table 5 it is reported that 30 IDs can identify 73 genes, that 15 genes can be identified by 33 IDs, this result indicates that more than one ID has been used to identify a gene. There are 3 genes that have no ID, and 4 genes whose mRNA length could not be found in any database. Since we cannot be sure of the genes that respond to microgravity, we report that there are 807 IDs (according to the MAS 5.0 method) that respond to microgravity, obtaining a correlation of 0.8325 (Figure 1). With the RMA method-AffylmGUI, we obtained 798 IDs that respond to microgravity and 8 IDs respond in the opposite direction to microgravity, obtaining a correlation of 0.9018 (Figure 2), surpassing that obtained by R. Herranz.

The next step was to detect the small group of genes, for which we used the correlation method. This method consisted of finding a list of genes for a given correlation value. First we fixed the P-value and then we proceeded to increase the cut-off values with steps of 0.01. In this way we found the matrix elements corresponding to the correlation and the total number of IDs. Finally, the list of genes for each correlation value was reported. This last result we had the computer 45 days processing all the data that were entered. It was observed that some IDs were repeated for different correlations, only 6 (150375_at (CG5326), 147367_at (CG18431), 149446_at (CG14598), 143201_at (Impl3), 143835_at (Dnaj-1) and 145560_at (CG4726)) appeared in almost all groups of correlations (Figure 3) [18]. It was logical to assume that these 6 IDs had to appear more frequently in each of the 10750 iterations performed by the program. In Figure 4 we can see that the red dots shown in the frequency graph are the 6 IDs mentioned above; they are the IDs that appear most frequently. For the validation of the algorithm, the 6 IDs were used. The 2010 experiments were repeated. First instar pupae were subjected to simulated microgravity during their entire pupal stage at an ambient temperature of 22°C. Four experiments were performed in triplicate, each group consisted of 7 pupae. Taqman probes were designed (Table 4). We had problems with ID 147367_at (Gen CG18431) and only analyzed 5 of them. These new experiments that were performed in 2010 did not have a cold pretreatment as did those published by Herranz. The way to validate the algorithm was to make sure that these genes only respond to microgravity. The results presented in Figure 5 show that all 5 genes are inhibited. This result is in agreement with the experiments performed at ESTEC 2003 as well as inside the ISS facilities. Another interesting point that can be rescued from this result (Figure 5), the bars shown in gray (Experiment 2010) are a mirror of the orange bars (Experiment 2003 ESTEC). It is observed that the orange bars are longer (have higher inhibition) than the gray bars. This is due to the cold pretreatment that the pupae were subjected to in their larval stage. A cumulative inhibitory process is observed both by the microgravity and by the previous cold treatment. As for the experiments carried out in the ISS (blue bars), ID 149446_at (CG14598) does not coincide with the mirror analysis with respect to the experiment carried out in 2010. The mirror analysis of the experiments (orange and gray bars), tells us that the genes possibly have the same regulators (continuation of this study) or that the regulators do not respond to microgravity and that their expression depends on another factor such as mRNA length. As the expression of these genes respond in the same way (mirror analysis) to

different experiments (experiment 2003 (cold treatment and microgravity) and experiment 2010 (microgravity)) we will consider them as possible microgravity sensor genes. To be considered as sensor genes, they must comply with the principles of linear algebra: Spatial relationship, temporal relationship and functional independence relationship. This consideration of sensors will not be discussed in this manuscript. The next step was to analyze the expression levels of these sensor genes with respect to their mRNA length, observing an opposite relationship with respect to mRNA length, concluding that the lower the mRNA bp, the higher their expression. In 2011 Marcu published an article in which he reports that the immune system of *Drosophila melanogaster* larvae exposed to microgravity are altered at low mRNA length. The expression of genes involved in cell maturation was reduced. In addition, the constitutive expression level of pattern recognition receptors and opsonins that specifically recognize bacteria, as well as lysozymes, Antimicrobial Peptide (AMP) pathway and immune stress genes, hallmarks of humoral immunity, was also reduced in larvae. We listed all these genes (results not presented in this manuscript) mentioned by Marcus, and correlated them with their respective mRNA bp length, observing the same behavior we obtained (higher bp length lower expression signal) [19]. For the validation of the 2010 experiment we used recycled probes from another research group working with *Drosophila m*. which were located in the same environment. The *MEF2* and *CF2* genes drop their expression levels in microgravity. As for *MEF2*, only significant values were observed with the RMA Method - AffylmGUI in both the RPM simulator and the ISS. *CF2* had a lot of variability both in the experiments performed in 2003 and in the experiment performed in 2010. As for *CG5703*, *Surf1, CG8885* and *CG1970* genes had positive expression and significant values as those obtained in 2003. The results in Figure 6 do not show a mirror analysis like those in Figure 5. This result strengthens that the 5 genes presented in Figure 5 are possible microgravity-responsive genes. Assuming that the expression of microgravity-responsive genes is regulated by mRNA length, we proceeded to search the MOTIS list in different databases (see methods) and concluded that all regulatory genes do not respond to microgravity. To be more confident that the regulators being presented and controlling the expression of the genes reported in this manuscript but under terrestrial gravity conditions, a sampling study was performed (See method). Of the 741 microgravity-responsive genes that are identified with a single ID, only 329 were randomly selected and grouped with the 112 genes corresponding to the TFs. The 441 genes were entered into the Scrig Protein database, which provided us with an image (Figure 7) showing that there is at least one relationship between each of the genes. This assures us that there is at least one relationship which may be functional, co-expression etc. between these genes. This figure indicates that under normal conditions at 1g the genes that respond to microgravity are regulated by at least one gene of the TF family, which play a crucial role in the expression and regulation of genes, this corroborates what was pointed out by. The question we ask, who regulates the genes that were exposed to microgravity. Figure 5 along with the list of genes presented, opened a clue for us, that possibly the 5 genes sensed to the lack of gravity have common regulators (We will raise a new manuscript with this hypothesis) and that their down or up-regulation depends on the mRNA size. This hypothesis that was put forward was corroborated with the results presented in Figure 8. According to comments that gene expression has an inverse relationship than its genome length. Lyu proposed an experiment where he kept the same promoter and changed the length of mRNA for different species. In order for our results to agree with Lyu's, it must be fulfilled that the regulators controlling the expression of microgravity-responsive genes are unchanged so all promoters would possibly have the same role. Therefore, the expression of genes in microgravity has an opposite relationship to the length of their genome. If this statement is correct, the break point observed in Figure 8-(a,b) leads us to think that there may be a critical mass value that is capable of sensing gravity. In other words, the minimum mass value that can react to the effects of gravity. In order to calculate the possible mass, all genes approaching from the left of the 1000 bp cutoff point were selected as well as genes approaching from the right of the cutoff point (Table 6). For each group we calculated the average value and its respective standard deviation (Figure 9), it can be observed that each group is not differentiable, therefore we can group them in a single table and calculate the average value and its standard deviation, which is 36.36875 KD (6,039 172 488 311 9×10^{-23} kilogram (kg)) with a standard deviation of 2.36978869 KD [20]. This result is possibly very controversial because it is the first time in human history that a critical mass value capable of sensing minimum gravity values has been reported. One of the applications that this research can have is the use of this mass value as a sensor to detect gravitational waves, also to understand why some molecular phenomena on earth are different in microgravity conditions, another interesting point of this result is to help to understand why bacteria such as *Pseudomonas aeruginosa* exposed to microgravity respond to space flight conditions by differential regulation, finally, to understand why microgravity affects survival, apoptosis, proliferation, migration and adhesion, as well as the cytoskeleton, extracellular matrix, focal adhesion and growth factors in cancer cells.

Conclusion

In this study, we investigated the gene expression response of *Drosophila melanogaster* pupae under microgravity conditions, identifying a critical mass value of approximately 36.6875 kD that limits expression. Through microarray data analysis and qRT-PCR validation, we confirmed that certain genes respond distinctly to microgravity, with expression levels correlating inversely with mRNA length. The identification of a group of microgravity-responsive genes, particularly five sensor genes, suggests that gene expression under these conditions is regulated differently from Earth-based controls. The calculated critical mass value introduces a novel concept that could offer insights into gravitational sensing mechanisms at the molecular level. Future research could explore the broader implications of this finding, including potential applications in understanding the biological effects of spaceflight and its impact on gene regulation.

Acknowledgement

The research work is in memory of my scientific mentor M.D Dr. Roberto Marco Cuellar. We wish to thank INNOVATEPERU for funding the project N°194-FINCYT-IA-2013, which is named as the following: “Validation of a Microgravity Machine measuring the levels of Gene Expression from a Microarray Assay using algorithms of parallel programming in GRID Computing architecture”. We would like to thank Quality Pharma SAC and BIOTECHCELL SAC for the support we have received during the publication of this manuscript.

References

1. Agrawal P, Heimbruch KE, Rao S (2018) Genome-wide maps of transcription regulatory elements and transcription enhancers in development and disease. Compr Physiol 9: 439.
2. Alahiotis SN, Onoufriou A, Fotaki M, Pelecanos M (1983) *Drosophila* lactate dehydrogenase: Developmental aspects. Biochem Genet 21: 199-211.
3. Bagni C, Bray S, Gogos JA, Kafatos FC, Hsu T (2020) The *Drosophila* zinc finger transcription factor CF2 is a myogenic marker downstream of MEF2 during muscle development. Mech Dev 117: 265-268.
4. Bahrampour S, Gunnar E, Jonsson C, Ekman H, Thor S (2017) Neural lineage progression controlled by a temporal proliferation program. Dev Cell 43: 332-348.
5. Baycin-Hizal D, Tian Y, Akan I, Jacobson E, Clark D, et al. (2011) GlycoFish: A database of zebrafish N-linked glycoproteins identified using SPEG method coupled with LC/MS. Anal Chem 83: 5296-5303.
6. Bergwitz C, Rasmussen MD, Derobertis C, Wee MJ, Sinha S, et al. (2012) Roles of major facilitator superfamily transporters in phosphate response in *Drosophila*. PloS One 7: e31730.
7. Bizzarri M, Masiello MG, Cucina A, Pensotti A (2017) Gravity constraint in cell phenotypic determination. Org J Bio Sci. 1: 61-68.
8. Bowser PR, Tobe SS (2007) Comparative genomic analysis of allatostatin-encoding (Ast) genes in *Drosophila* species and prediction of regulatory elements by phylogenetic footprinting. Peptides 28: 83-93.
9. Braun M, Bohmer M, Hader DP, Hemmersbach R, Palme K (2018) Gravitational biology I: Gravity sensing and graviorientation in microorganisms and plants. Cham, Switzerland: Springer.
10. Chawla G, Deosthale P, Childress S, Wu YC, Sokol NS (2016) A let-7-to-miR-125 microRNA switch regulates neuronal integrity and lifespan in *Drosophila*. PLoS Genet 12: e1006247.
11. Chen L, O'Keefe SL, Hodgetts RB (2002) Control of Dopa decarboxylase gene expression by the Broad-Complex during metamorphosis in *Drosophila*. Mech Dev 119: 145-156.
12. Costantino BF, Bricker DK, Alexandre K, Shen K, Merriam JR, et al. (2008) A novel ecdysone receptor mediates steroid-regulated developmental events during the mid-third instar of *Drosophila*. PLoS Genet 4: e1000102.
13. Diederich RJ, Pattatucci AM, Kaufman TC (1991) Developmental and evolutionary implications of labial, Deformed and engrailed expression in the *Drosophila* head. Development 113: 273-281.
14. Iyer Venky N, Pollard Daniel A, Sackton Timothy B, Larracuente Amanda M, Singh Nadia D, et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203-218.
15. Dubinin NP, Vaulina EN (1976) The evolutionary role of gravity. Life Sci Space Res 14: 47-55.
16. Emery IF, Bedian V, Guild GM (1994) Differential expression of Broad-Complex transcription factors may forecast tissue-specific developmental fates during *Drosophila* metamorphosis. Dev 120: 3275-3287.
17. Fu D, Zhao C, Ma J (2003) Enhancer sequences influence the role of the amino-terminal domain of Bicoid in transcription. Mol Cell Biol 23: 4439-4448.
18. Gajewski KM, Schulz RA (2010) CF2 represses Actin 88F gene expression and maintains filament balance during indirect flight muscle development in *Drosophila*. PloS one 5: e10713.
19. Garcia-Zaragoza E, Mas JA, Vivar J, Arredondo JJ, Cervera M (2008) CF2 activity and enhancer integration are required for proper muscle gene expression in *Drosophila*. Mech Dev 125: 617-630.
20. Gonzy G, Pokholkova GV, Peronnet F, Mugat B, Demakova OV, et al. (2002) Isolation and characterization of novel mutations of the Broad-Complex, a key regulatory gene of ecdysone induction in *Drosophila melanogaster*. Insect Biochem Mol Biol 32: 121-132.